Modeling, fabrication, and characterization of 2D devices for electronic and photonic applications

Ankur Baburao Nipane

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy under the Executive Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021
Abstract

Modeling, fabrication, and characterization of 2D devices for electronic and photonic applications

Ankur Baburao Nipane

Over the last two decades, two-dimensional (2D) materials such as graphene and transition metal dichalcogenides (TMDCs) have invoked tremendous interest of the scientific community due to their unique electronic and optical properties. While TMDCs hold great promise as a potential replacement for silicon for scaling transistors beyond sub-3 nm technology node, graphene holds great potential as transparent electrodes and optical phase-modulators for next-generation photonic devices. In addition to the aforementioned applications, these 2D devices also provide a great platform for studying novel physical phenomena associated with 2D materials such as Moiré interactions, valley-dependent spintronics, and correlated electron physics. In order to realize high-performance 2D material based devices, advancement of three key aspects are imperative - (1) analytical modeling to gauge insights into the electrostatics and current transport in 2D devices, (2) development of efficient techniques for fabricating 2D devices, and (3) understanding the fundamental limitations of the existing characterization techniques and developing better methods.

We started by modeling the unique electrostatics of the 2D lateral pn junctions, wherein we developed analytical expressions for the electric field, electrostatic potential, and depletion width across 2D lateral p-n junctions. We extend these expressions for use in lateral 3D metal-2D semiconductor junctions and lateral 2D heterojunctions. The results show a significantly larger depletion width (∼ 2 to 20×) for 2D junctions compared to conventional 3D junctions. Further, we show that the depletion widths at metal-2D semiconductor junctions can be significantly modulated by the surrounding dielectric environment and semiconductor doping density. Finally, we derived a minimal dielectric thickness for a symmetrically-doped 2D lateral p-n junction, above which the
out-of-plane simulation region boundaries minimally affect the simulation results. After electrostatics, we attempted to understand the current transport in 2D material-based devices. Typically used back-gated field-effect transistors (BGFETs) are often modeled as Schottky barrier (SB)-MOSFETs assuming that the current flow is limited by the source-contact in the OFF state, while the channel limits the current in the ON state. Here, using an analytical model and drift-diffusion simulations, we show that the channel limits the overall current in the OFF state and vice versa, contrary to past studies. For top-contacted BGFETs, we modeled different current paths at a top-contacted metal-2D semiconductor junction and illustrated the unique “corner effect”—where the potential change and current transport is dominated by the metal-2D semiconductor edge and the associated lateral region. We determined that the edge transport supersedes the vertical current injection in monolayer TMDCs and hence, to reduce contact resistance in 2D devices degenerate doping of channel region next to contact regions is of paramount importance.

After developing models to theoretically analyze these devices, we focused on understanding the shortcomings in the existing characterization techniques affecting the extraction of important device parameters such as contact resistance, SBH, and channel mobility. We prove that the transfer length estimated using the standard TLM measurement technique can severely overestimate the true transfer length. We also discuss the large discrepancy in SBH values extracted using the Arrhenius method compared to their theoretical values. Using our analytical modeling, we attribute this to the presence of long channel regions in experimental devices. Furthermore, we highlight that the presence of large contact resistance results in underestimation of channel mobilities which renders Kelvin measurements such as four-probe and Hall-bar measurements imperative for 2D devices.

Finally, we introduced a unique etch and doping method using self-limiting oxidation which allows us to design and fabricate various high-performance 2D devices. We
first used the method to demonstrate a selective, damage-free atomic layer etch (ALE) that enables layer-by-layer removal of monolayer WSe$_2$ without altering the physical, optical, and electronic properties of the underlying layers. Using a comprehensive set of characterization techniques, we show that the quality of our ALE processed layers is comparable to that of pristine layers of similar thickness. Further, using graphene as a testbed, we demonstrate the use of sacrificial monolayer WSe$_2$ layer to protect the channel, which is etched in the final process step in a technique we call Sacrificial WSe$_2$ with ALE Processing (SWAP). Furthermore, the top oxidized layer acts like an atomically-thin degenerate p-type dopant for a large variety of semiconductors such as graphene, carbon nanotubes, and WSe$_2$. We show that the TOS-doped graphene yields low sheet resistance ($118 \, \Omega/\square$) due to high mobility ($2000 \, \text{cm}^2/\text{V} \cdot \text{s}$) at very high hole density ($3.2 \times 10^{13} \, \text{cm}^{-2}$) that remains active even at 1.5 K. We apply this principle to improve the transmittance of graphene (>99%) at telecommunication bandwidth (1.5 to 1.6 $\mu$m), that makes it a suitable replacement for Indium tin oxide (ITO) as a transparent electrode.
# Table of Contents

List of Tables ................................................................. v

List of Figures ................................................................. vi

Acknowledgements ............................................................ xxii

Dedication ................................................................. xxv

Chapter 1: 2D Devices - Motivation and Challenges ......................... 1

1.1 Why 2D materials? ......................................................... 1

1.1.1 TMDCs for electronic applications ................................. 2

1.1.2 Graphene for photonic and optoelectronic applications .......... 6

1.2 Current challenges in making 2D devices ............................. 8

1.2.1 Fabrication challenges ............................................... 9

1.2.2 Lack of appropriate device models .............................. 12

1.3 Thesis Overview .......................................................... 14

Chapter 2: Electrostatic Modeling of Lateral Junctions in 2D Semiconductors .............................. 15

2.1 Electrostatics of 2D lateral p-n junctions ............................... 16

2.1.1 Formulation of the 2D lateral p-n junction as a collection of 1D line charges 17
2.1.2 Derivation of the electric field and electrostatic potential as a function of position using the image charge method ........................................... 22

2.1.3 Analytical formulation of the depletion width ........................................... 30

2.1.4 Extending the analytical model to lateral metal-semiconductor junctions .... 35

2.1.5 Verification of analytical model against experimental and simulation results 38

2.1.6 Summary .................................................................................................. 42

2.2 Role of out-of-plane dielectric thickness in electrostatic simulation of 2D devices ................................................................................................. 43

2.2.1 Effect of boundary conditions ...................................................................... 45

2.2.2 Estimating minimum acceptable simulation window .................................. 47

2.2.3 Summary .................................................................................................. 52

Chapter 3: Modeling of Current Flow in Back-Gated Field-Effect Transistors .... 54

3.1 Where is the current bottleneck in 2D transistors? ...................................... 55

3.1.1 Analytical formalism for source and channel current in long-channel BGFET 55

3.1.2 Load-line analysis for extracting current bottlenecks .................................. 62

3.1.3 Summary .................................................................................................. 66

3.2 Understanding current transport in top-contacted 2D semiconductors .......... 66

3.2.1 Corner effect in a metal/2D-semiconductor junction .................................. 67

3.2.2 Vertical Schottky junctions with global back gates .................................... 71

3.2.3 Current flow under the contact ................................................................. 73

3.2.4 Current flow at the edge of the contact ..................................................... 82

3.2.5 Summary .................................................................................................. 84

Chapter 4: Pitfalls in Electrical Measurements of 2D Transistors ..................... 86
4.1 Contact resistance and Schottky barrier extraction 86
  4.1.1 Transfer line method 87
  4.1.2 Challenges with TLM 90
  4.1.3 Schottky barrier heights and Fermi-level pinning 92
  4.1.4 Challenges with SBH extraction 94
4.2 Mobility extraction 103
  4.2.1 MOSFET mobility 103
  4.2.2 Challenges due to contact resistance 106
  4.2.3 Hall mobility 108
  4.2.4 Challenges of Hall-effect measurements 110
4.3 Summary 112

Chapter 5: Fabrication of High-Mobility 2D Devices Using Atomic Layer Etch 114
  5.1 Introduction to atomic layer etch (ALE) 115
    5.1.1 Process flow 116
    5.1.2 Monolayer precision of ALE process 118
    5.1.3 Repeatability and patternability of ALE technique 122
  5.2 Detailed characterization of ALE-processed flakes 124
    5.2.1 Material characterization 124
    5.2.2 Optical characterization 126
    5.2.3 Electrical characterization 130
  5.3 Graphene devices using sacrificial WSe$_2$ with ALE processing 132
  5.4 Summary 137
List of Tables

2.1 Analytical expressions for various electrostatic parameters for 2D lateral junctions. ......................................................... 42

2.2 Material parameters of the simulated generic 2D semiconductor. ........... 46

2.3 Minimal out-of-plane dielectric thickness ($t_{ox,min}$) and normalized simulation time as a function of $\alpha$, the fraction of the total electric flux that emanates through the top and bottom surfaces of the Gaussian box shown in Fig. 2.12. .......................................................... 52

3.1 Model parameters for a monolayer MoS$_2$ BGFET with a 3 nm SiO$_2$ gate oxide [88]. ................................................................. 64

5.1 Room-temperature hole mobilities in p-FETs fabricated on pristine WSe$_2$ flakes. ................................................................. 132
List of Figures

1.1 **Spectrum of 2D materials.** 2D materials encompass a large range of bandgaps making them useful for applications across the spectral range. The atomic crystal of different 2D materials such as $h$-BN, MoS$_2$, BP, and graphene is provided along with their electrical properties [4].

1.2 **Different applications of 2D materials.** Possible applications of 2D materials along with the complexity associated with 2D material-silicon process flow integration. The position of a circle shows the year when the particular product could potentially be introduced in the industry [11].

1.3 **Unique properties of TMDCs.** (a) Comparison of mobile charge distribution along the thickness of the semiconductor for 3D vs 2D materials. Narrower mobile charge distribution in 2D materials allows better electrostatic control by the gate electrodes [13]. (b) Channel mobility ($\mu$) as a function of channel thickness ($t_{ch}$) for different materials. 2D materials offer much higher mobilities at smaller thickness [11]. (c) Energy-delay benchmarking of TMDC-based TFETs with other technologies showing higher switching speeds at lower switching energies [16].

1.4 **Graphene for optoelectronics.** (a) Band structure of graphene. Inset shows linear dispersion relation and absence of bandgap [19]. (b) Room temperature mobility versus electron density for different material systems. Graphene shows much higher mobilities than other systems [20]. (c) Transmittance as function of wavelength in the visible regime for different materials useful for transparent conductor application [17]. (d) Variation of sheet resistance with thickness. Graphene provides higher transmittance and low sheet resistance at smaller thickness [17].

1.5 **Major challenges associated with 2D materials.** Timeline associated with different challenges associated with 2D materials. Currently, the major challenges include lack of large-scale growth & transfer, fabrication issues associated with doping and contacts, and design principles for better device design [11].
1.6 **Disorder in 2D materials.** Internal and external disorders which affect carrier mobility of 2D materials [25].

1.7 **Scattering mechanisms in 2D materials** (a) Theoretical comparison of different scattering mechanisms affecting mobility of monolayer MoS$_2$ as a function of temperature. At room temperature, the mobility is limited by optical phonon scattering (red line). However, low temperature mobilities is limited by impurity scattering (blue line) [26]. (b) Experimental mobilities on different encapsulation (substrate+passivation) layers. Low-temperature mobility is highest for h-BN encapsulation [25].

1.8 **Issue of contact resistance in 2D materials.** (a) Contact resistance ($R_C$) as a function of carrier density. Compared to other materials, TMDCs show two-order of magnitude higher contact resistance [13]. (b) Cross-sectional transmission electron microscope (TEM) image of metal-semiconductor interface shows fabrication-induced damage [29].

2.1 **Electric field profiles in 2D lateral p-n junctions.** (a) Schematic of a 2D lateral p-n junction showing depletion and quasi-neutral regions with different dielectric media above ($\epsilon_1$) and below ($\epsilon_2$) the plane containing the 2D lateral p-n junction. The out-of-plane electric field is prominent in these junctions due to the large surface to volume ratio. (b) Electric field lines for symmetric and asymmetric doping configurations. For asymmetric doping, a non-zero tangential ($z$) electric field exists at the junction interface, which prevents the use of previous analytical methods for asymmetrically doped junctions.

2.2 **Comparison of bulk vs 2D electrostatics.** (a) Schematic of 2D lateral p-n junction showing the depletion region as a collection of infinitesimally thin line charges. (b) Schematic of 3D junction showing the depletion region as a collection of infinitesimally thin sheet charges. (c) Electric field profile due to single line charge. The electric field is inversely dependent on distance from the line charge. (d) Electric field profile due to single sheet charge. The electric field is independent of distance from the sheet charge.
2.3 Qualitative analysis of the charge distribution, electric field, and electrostatic potential across a 2D lateral p-n junction with symmetric doping. (a) Charge distribution, electric field, and potential profile using the standard depletion approximation. The standard depletion approximation results in infinite electric fields at the edges of depletion region and a non-monotonic potential across the junction. (b) Enforcement of $E_x(-x'_p) = 0$ using the method of image charges. The charge to the right is reflected across the $x = -x_p$ plane. (c) Extension of (b) to enforce $E_x(x'_n) = 0$ by reflecting the charge to the left across the $x = x_n$ plane. However, this leads to a small but non-zero electric field at $x = -x_p$. (d) Extension of the image charge method to an infinite number of reflections to satisfy both $E_x(-x'_p) = 0$ and $E_x(x'_n) = 0$ boundary conditions. This also results in monotonic electrostatic potential throughout the depletion region.

2.4 Approximations of $f$ and $g$ functions for asymmetrically doped junctions. Functions $f$ and $g$ from Equation (2.37a) and Equation (2.37b) plotted for different doping ratios $\left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right)$. For strongly asymmetric junctions, a logarithmic dependence for either $f$ or $g$ is observed.

2.5 Comparison of 2D vs 3D depletion width as a function of sheet charge density. Depletion width calculated for (a) symmetric and (b) asymmetric doping configurations calculated using our 2D method and the conventional 3D method. For symmetric doping configuration, the 2D depletion width is proportional to $\frac{1}{N_{S2D}}$ whereas the 3D depletion width is proportional to $\frac{1}{\sqrt{N_{S2D}}}$ For strongly asymmetric junctions, the 3D depletion width is determined solely by the lowly doped side (e.g. it is independent of $N_{A,2D}^-$ for $N_{A,2D}^- \gg N_{D,2D}^+$), whereas the 2D depletion width shows a weak (logarithmic) dependence on the doping density of the highly doped side.

2.6 Comparison of electric field profiles for 2D lateral p-n and metal-semiconductor junction (a,b) Schematics of a symmetrically doped 2D lateral p-n junction and a 3D metal-2D semiconductor junction. (c) A symmetrically doped 2D lateral p-n junction is simply a mirror charge construct of a 3D metal-2D semiconductor junction. The depletion for a 3D metal-2D semiconductor junction is half of that of a symmetrically doped 2D lateral p-n junction having the same semiconductor doping density.

2.7 Impact of surrounding dielectric media on 2D depletion width. The significant out-of-plane electric field of 2D lateral p-n junctions leads to a linear dependence of the depletion width on the relative dielectric permittivity of the surrounding dielectric media.
2.8 Verification of the analytical model against simulations. (a) Comparison of simulated and calculated space-charge distribution ($\rho$) along a 2D lateral p-n junction. The calculated space charge accurately models the box-like section of space charge profile. The simulated 2D lateral p-n junction contains long depletion tails which decay as $\frac{1}{x}$. (b) Electric field ($E_x$) and (c) electrostatic potential ($V$) profiles for the same structure in (a). Although the space-charge calculated from the analytical model and TCAD simulation show significant differences, the electric field and electrostatic potential profiles show little discrepancy.

2.9 Verification of the analytical model against experimental data. (a) Experimental [38] and calculated depletion width for a 2D lateral p-n junction for the experimentally given range of $\epsilon_{\text{eff}}$. Our 2D method provides excellent fits to the experimental data, whereas the depletion widths calculated using the conventional 3D method are off by an order of magnitude irrespective of $\epsilon_{\text{eff}}$. (b) Comparison of different 2D methods against the same experimental data from part (a) [40, 41, 43, 45, 58]. Assuming $\epsilon_{\text{eff}} = 10.3\epsilon_0$, our 2D method provides best estimate to the experimental data.

2.10 Effect of varying dielectric thickness on 2D depletion width. (a) A thin (2D) lateral p-n junction surrounded by a dielectric with a simulated thickness of $t_{\text{ox}}$. A significant out-of-plane electric field exists in the surrounding dielectric, unlike in a 3D p-n junction. (b) In-plane energy-band diagram of a 2D lateral p-n junction with symmetric doping. The significant variation of the energy band diagram with $t_{\text{ox}}$ highlights the impact of the out-of-plane simulation dimensions on the in-plane junction electrostatics.

2.11 Undesired reflections caused by Neumann boundary conditions. (a) A simulated 2D lateral p-n junction with Neumann boundary conditions on the out-of-plane edges of the surrounding dielectric media. The Neumann boundary conditions mirror the structure such that the simulated structure is equivalent to the structure shown in (c) as opposed to (b). A small simulated $t_{\text{ox}}$ yields closely stacked reflections that significantly affect the simulation results. As the simulated $t_{\text{ox}}$ becomes large, the reflections becoming increasingly far apart, minimally impacting the simulation results.
2.12 **Calculation of minimal out-of-plane dielectric thickness for accurate electrostatic simulations.** A Gaussian box enclosing the right half (n-side) of a symmetrically-doped 2D lateral p-n junction surrounded by an infinitely thick dielectric. The minimal dielectric thickness ($t_{ox,min}$) is derived by determining the Gaussian box thickness ($t_{box}$) for which the flux passing through the top and bottom surfaces ($\Phi_{top}$ and $\Phi_{bottom}$) is some small fraction ($\alpha$) of the total flux. For a 2D lateral p-n junction with a fairly long quasi-neutral region (compared to its depletion region), the electric flux emanating from the right surface is approximately zero ($\Phi_{right} \approx 0$).

2.13 **Variation in simulated 2D depletion width with out-of-plane dielectric thickness.** Impact of the out-of-plane dielectric thickness ($t_{ox}$) on the simulated depletion width of a symmetrically-doped p-n junction, shown in the inset. The simulated depletion width varies significantly with $t_{ox}$, approaching the theoretical bulk (3D) junction depletion width as $t_{ox} \to 0$ and the theoretical 2D depletion width as $t_{ox} \to \infty$. The simulated 2D depletion width is nearly equal to the theoretical 2D depletion for $t_{ox} > t_{ox,min}$ as indicated by the shaded region of the plot.

3.1 **Modeling of long-channel BGFETs.** (a) A typical long-channel n-type monolayer BGFET with global back gating of both channel and contact regions. (b) Energy band diagram along the channel ($L_{CH}$) highlighting the electron SB ($\phi_{Bn}$) at the source-drain contacts. The source SB is reverse-biased while the drain SB is forward-biased by the applied $V_{DS}$. (c) Lumped-circuit model for a long-channel BGFET with the source/drain contacts modeled as ballistic SBFETs and channel region modeled as a conventional MOSFET. Conventionally, the source SBFET is assumed to dominate in the OFF state and channel MOSFET is assumed to dominate in the ON state.

3.2 **$V_{GS}$-dependent current components in long-channel BGFET.** (a) Simplified model for long-channel BGFET [given in Fig. 3.1(a)] as the drain barrier is forward-biased for $V_{DS} > \Phi_{Bn}$. $I_{SB}$ denotes the current allowed by the source SBFET and $I_{CH}$ denotes the current allowed by the channel MOSFET. The band diagram relates $V_{SB}$ to the electron quasi-Fermi level at $x = 5\lambda$. (b) Different components of $I_{SB}$ and their relative magnitude as a function of $V_{GS}$. In the OFF state, $I_{SB}$ is predominantly thermionic in nature with pure thermionic emission for $V_{GS} < V_{FBSB}$, i.e., $\psi_{S} < 0$. The tunneling component increases beyond $V_{GS} > V_{FBSB}$ and becomes dominant in the ON state ($V_{GS} > V_{TH}$). (c,d) Typical $I_{SB}$ (red) and $I_{CH}$ (blue) for a long-channel BGFET as a function of $V_{SB}$ for constant $\psi_{S}$. 
3.3 **Load-line analysis for calculating current and applied potential drop across the source in a long-channel BGFET.** (a) Load-line diagram showing $I_{SB}$ (solid lines) and $I_{CH}$ (dashed lines) as a function of $V_{SB}$ for a given $\psi_s$. The cross-over points, highlighted by square dots, represent the allowed device current and applied potential drop across the source contact. (b) Allowed device current, $I_D$, calculated using the load-line analysis. A comparison with $I_{SB}$ and $I_{CH}$ confirms that the channel region dominates in the OFF state as $I_D \approx I_{CH}$ and contact dominates in the ON state with $I_D \approx I_{SB}$. The left inset shows band diagram in the subthreshold region and the right inset shows the band diagram in the ON state highlighting the presence of electric field in the channel region. (c) $V_{SB}$ as a function of $\psi_s$ showing channel dominated behavior at small $\psi_s$ (low applied gate voltage). (d) Comparison of $I_D$ calculated using the analytical method (solid lines) and extracted from self-consistent drift-diffusion simulation (solid dots) of the same MoS$_2$ BGFETs. The analytically calculated current show good agreement to the simulated data. ........................................ 63

3.4 **A vertical metal/2D-semiconductor Schottky junction.** (a) The two layers are separated by the van der Waals gap, and a depletion region exists in the 2D semiconductor due to the built-in potential of the metal/semiconductor Schottky junction. (b) Charge induced in the metal due to the depletion charge of the semiconductor. (c) Electrostatic potential as a function of lateral position in the 2D semiconductor layer. Nearly all of the potential drop occurs at the right-edge of the overlap region, which we call the *corner effect*. .................................................. 68

3.5 The total electrostatic potential across the structure is equal to the built-in potential of the junction ($\phi_{bi}$), which is divided in the vertical potential drop ($\phi_{\perp}$) and lateral potential drop ($\phi_{lat}$). $\phi_{\perp}$ is approximated as the potential drop between two equal and opposite sheet charges. $\phi_{lat}$ is approximated as the lateral potential drop due to the dipole consisting of a line charge at the edge of the metal and a sheet charge representing the depletion width in the semiconductor. .................................................. 70

3.6 **A 2D semiconductor with a top contact and global back gate.** (top) Zero bias is applied to the back gate with respect to the top contact. The 2D semiconductor is depleted due to the built-in potential of the metal/semiconductor junction. The depletion width extends laterally into the 2D layer. (bottom) A bias of 100 V is applied to the back gate. The back gate accumulates electrons in the n-layer to the right of the top contact, which decreases the extent of the lateral depletion width; however the semiconductor remains depleted directly under the top contact. The strong capacitive coupling of the top contact to the 2D layer severely limits the ability of the back gate to affect the potential (and, therefore, the electron density) directly under the top contact. .................................................. 73
3.7 (a) Schematic cross-sectional diagram showing possible carrier injection process from metal to monolayer TMDC. This is modeled as a series and parallel connection of resistors in a back-gated device. The arrows indicate the direction of electron flow (opposite of current flow). (b) Simplified circuit model of (a) where the vertical carrier injection is modeled by a parameter $R_{C,\bot}$. .......................................................... 74

3.8 Edge ($G_{edge}$) and vertical ($G_{C,\bot}$) conductance as function of carrier density ($n_{2D}$) sufficiently far from the contact with different Schottky barrier height, $\Phi_{Bn}$. Relative dielectric constant at van der Waals gap between metal and TMDCs is determined from 1 (dashed line) to 10 (solid line). The $G_{edge}$ is larger than $G_{C,\bot}$ for monolayer TMDCs for all practical SBH and carrier densities. .......................................................... 84

3.9 Contact resistance ($R_C$) as function of carrier density ($n_{2D}$) shows that degenerate doping is required to lower $R_C$ .......................................................... 85

4.1 **Extraction of $R_C$ and $L_T$.** (a) Top view of the TLM configuration showing different channel lengths (L1, L2,...). The enlarged view shows the distribution of total resistance in terms of $R_C$ and $R_{CH}$. (b) A linear fit of the plot of $R_{Total}$ versus channel length giving rise to $R_C$, $R_{SH}$, and $L_T$. (c) Schematic of a MoS$_2$-based TLM device with back gating through SiO$_2$. (d) Schematic of a bilayer MoS$_2$ device and the TLM plot for the device showing the linear trend of $R_{Total}$ versus channel length as a function of carrier density (gate voltage). The inset shows the extracted $R_C$ values as a function of carrier density, demonstrating contact region gating. ....... 88

4.2 **Schottky barrier height and Fermi level pinning.** (a) Band diagram of a metal-semiconductor junction. (b) SBH versus metal work function showing the Schottky–Mott rule and Fermi level pinning. (c) Schematic image of Fermi level pinning [119]. .......................................................... 92

4.3 The top panel compares the theoretical and experimental values of electron Schottky barrier height (SBH) for different metal contacts to few-layer MoS$_2$. The theoretical values are generally higher than the experimental values indicating a systematic underestimation by the temperature-dependent SBH extraction method. The bottom panel highlights the absolute error (disparity) in the measured SBH values compared to theoretical values determined by comparing the mean values of experimental and theoretical values for different metals, provided in top panel. ............. 94
4.4 **Arrhenius method of Schottky barrier extraction:** (a) Schematic and circuit diagram of the back-gated monolayer MoS$_2$ transistor. The 1L-hBN acts as a tunnel barrier and provides excellent contacts to monolayer MoS$_2$. (b) Temperature dependent transfer characteristics of the device shown in part (a). The expected region ($V_{GS} < V_{FBSB}$) which is dominated by thermionic transport is highlighted in grey. (c) Arrhenius plots corresponding to the transfer characteristics in part (b) at different gate voltages. The voltage-dependent slopes (highlighted in red) were used to extract activation energy ($E_A$) as a function of gate voltage. (d) Activation energy as a function of gate voltage. The true electron barrier ($\Phi_{BN}$) was extracted by identifying the point at which the slope (shown by solid black line) deviates from the linear trend at low gate voltages. 96

4.5 **Self-consistent check for the accuracy of extracted Schottky barrier height:** (a) Flow chart for temperature dependent Schottky barrier extraction method. The first three steps are part of the standard/typical extraction method. A follow-up step involving comparison of theoretical and experimental drain current values is added to the flow chart to make the extraction self-consistent. (b) Comparison of drain current at extracted flat band voltages ($I_D(V_{FBSB})$) for various studies involving few-layer MoS$_2$ based transistors with the drain current predicted by thermionic emission model for extracted barrier heights. The model predicts $\sim 3000 \times$ larger current at the extracted $V_{FBSB}$ for our device. Similarly, the model also predicts significantly higher theoretical drain currents than the experimental values for other prominent contact engineering works. 100

4.6 (a) Variation in the extracted flat-band voltage ($V_{FBSB}^{ext}$) and flat band current ($I_{FBSB}^{ext}$) with $L_{CH}$. (b) Extracted SBH as a function of $L_{CH}$ using current-based and Arrhenius method. The ideal SBH is also indicated. 102

4.7 **Effective mobility and field-effect mobility.** (a) Effective mobility is extracted from the drain conductance near the origin of the output characteristics. For both mobility extraction techniques, both the transfer and output characteristics should show linear behavior around the extraction bias point. (b) Field-effect mobility extracted from the transconductance of a MOSFET, biased in the linear regime [114]. 103

4.8 **Relation between effective mobility and field-effect mobility.** (a) Typical universal $\mu_{\text{eff}}$ curves for Si for different doping densities ($N_{D_1} < N_{D_2} < N_{D_3}$). Coulomb (impurity) ($\mu_c$), phonon ($\mu_{ph}$), and surface scattering ($\mu_{sr}$) limited mobilities are also indicated [139]. (b) Calculated electron mobilities for monolayer MoS$_2$ as a function of carrier density (top panel) and gate voltage (lower panel) for different temperatures [67]. 106
4.9 Hall effect measurements of a bridge-type Hall bar structure. (a) Illustration of the Hall effect of an electron. (b) Circuit configuration of a typical bridge-type Hall bar structure device. (c) $V_H$ versus B-field of a graphene device dependent on $V_{GS}$. (d) SdH effect in graphene showing oscillatory behavior of $\rho_{xx}$ and $\sigma_{xx}$ in the presence of B-fields [141].

5.1 ALE process flow and XPS characterization. (a) The topmost layer of WSe$_2$ is removed by first oxidizing the flake using UV+O$_3$ for 30 minutes at room temperature, forming monolayer tungsten oxyselenide (TOS) that is subsequently removed with a mild 1M KOH dip for 10 s. The etching process selectively removes the oxidized topmost layer without damaging the underlying layers. (b) XPS spectra of the W (4f) peak for a 4L WSe$_2$ flake showing the appearance of W-O peaks after the UV-ozone process. The peaks corresponding to W-O bond formation disappears after the KOH treatment indicating the removal of the TOS layer.

5.2 Monolayer precision of ALE process. (a) Optical images (top), AFM scans (middle), and height profiles (bottom) of an exfoliated WSe$_2$ flake on SiO$_2$ at each stage of the ALE process. The as-exfoliated bilayer region (2L) is etched down to monolayer (1L) and the monolayer region is completely removed after the oxidation and etching step. (b) Raman spectra for the as-exfoliated, oxidized, and etched flake from (a). The Raman spectra were taken near the center of the 2L region of the flake. The removal of $B_{1g}^1$ peak (in the inset) corroborates the conversion from bilayer to a monolayer. (c) PL spectra are taken from the as-exfoliated bilayer region of the flake given in (a) at each stage of the ALE process. The PL spectra shows a clear transition from a broad, low intensity bilayer spectrum to a sharp, high-intensity monolayer spectrum, in agreement with our optical and AFM results.

5.3 AFM height profiles of etched-to 1L WSe$_2$ (a,b) Color-corrected optical images of the pristine and etched-to 1L WSe$_2$ on h-BN. (c) Height profiles along A-A' line cuts indicate similar heights (~0.7 nm) for both the flakes. The pristine 2L used to obtain etched-to 1L is also shown as a reference.

5.4 Application of ALE process to other TMDCs Optical images of MoSe$_2$, MoS$_2$, and WS$_2$ flakes processed using the ALE method. In the case of MoSe$_2$, the 1L and 2L regions are uniformly oxidized and removed by the ALE process leaving behind an etched-to 1L MoSe$_2$ with no apparent damage. However, the oxidation process is not self-limiting for MoS$_2$ and WS$_2$ samples possibly due to different diffusion barrier strengths of sulfur-based TMDCs.
5.5 **Repeatability of the ALE process.** (a) Optical images show clear demonstration of repeatable ALE process. In the first ALE cycle, the (n) layer section becomes (n-1) layer, e.g., 3L region become 2L, 2L region become 1L, and 1L region is completely removed. The same process is repeated twice to completely remove the 3L region with clear change in optical contrast of thick area illustrating layer-by-layer etching of all the regions. No visible damage was observed on the thick or thin parts of the flake during the multiple runs of the ALE process. (b) Normalized Raman spectra of the different monolayers obtained from multiple ALE cycles showing no change in crystalline structure of WSe$_2$ with ALE processing. (c) Normalized photoluminescence (PL) spectra of the same monolayers illustrating slight increase ($\sim$20 meV) in sample inhomogeneity with multiple ALE cycles possibly due to small size of the etched-to flakes, however no defects peaks were observed in the PL spectra.

5.6 **Patternability of the ALE process.** Optical micrographs showing different etch patterns in exfoliated few-layer WSe$_2$. Different patterns (2 $\mu$m $\times$ 15 $\mu$m rectangles, multiple 1 $\mu$m $\times$ 20 $\mu$m lines) were etched out in separate regions of as-exfoliated flakes showing selective patterning of the flake with PMMA mask.

5.7 **Comparison of pristine and etched-to 1L WSe$_2$ flakes.** (a) (i). Optical images of WSe$_2$ flake confirming the monolayer etch. The as-exfoliated bilayer (2L) region is etched down to monolayer (1L) and the monolayer region is completely removed. (ii). AFM characterization confirms the atomic layer etch with no apparent change in shape, size, or topography of the flake. RMS surface roughness ($R_q$) of the etched-to 1L is similar to that of the as-exfoliated monolayer indicating that the ALE process minimally affects the underlying layers. (b) Raman spectrum of the etched-to 1L WSe$_2$ is similar to that of as-exfoliated 1L WSe$_2$. No additional defect peaks were observed in the etched-to 1L WSe$_2$ with negligible change in the peak positions confirming the crystalline nature of the etched layers. (c) STEM images of as-exfoliated and etched-to 1L WSe$_2$ taken at the same scale. The etched flake does not show any noticeable damage or defects when compared to pristine flake. The electron diffraction patterns also indicate a single-crystalline structure of the etched-to 1L flake with similar angle and distance between adjacent planes as that of pristine 1L WSe$_2$, indicated by arrows in the inset.
5.8 **Optical characterization of etched-to 1L WSe$_2$.** (a) Optical (gray-scaled) images of pristine WSe$_2$ flake (before and after etch) show that only one layer is removed per ALE cycle. Corresponding PL spectra confirm this with a clear transition from a broad low-intensity bilayer WSe$_2$ spectrum to a sharp monolayer spectrum for etched-to 1L WSe$_2$ after the ALE process. The etched-to 1L WSe$_2$ flake depicts PL characteristics similar to the pristine 1L WSe$_2$ flake with a 0.55× reduction in the peak PL intensity. (b) Deconvolved room-temperature PL spectra of the pristine and etched-to 1L WSe$_2$ from (a). The PL spectrum for the etched-to 1L WSe$_2$ shows the same number of deconvolved peaks (exciton peak ($X^0$) and hole trion peak ($X^+$)) and similar $X^+/X^0$ ratio. (c) Integrated PL intensity ($I$) for as-exfoliated and etched-to monolayers for different pump laser powers ($P$). Both as-exfoliated and etched-to flakes show similar sub-linear behavior. (d) Time-resolved PL decay also corroborates the similar optical characteristics of pristine and etched-1L WSe$_2$ flakes with comparable PL dynamics and decay time constants ($\tau$).

5.9 **Temperature-dependent PL characterization of etched-to 1L WSe$_2$.** (a) PL spectra of as-exfoliated (pristine) and ALE-processed (etched-to) monolayer WSe$_2$ flakes measured at 77K. The ALE-processed flake has similar PL spectrum as pristine flake with no discernible defect peaks. (b) Excitonic linewidth as a function of temperature show a clear reduction in the linewidth and linewidth difference between the pristine and etched-to monolayers with lowering the temperature, indicating reduced sample inhomogeneity.

5.10 **Electrical properties of ALE-processed WSe$_2$ flakes.** (a) Enhanced optical image of a back-gated 2D device made from an etched-to 3L WSe$_2$ flake with transferred via contacts made from Pt in a Hall-bar pattern. The WSe$_2$ was exfoliated as a 5L flake and etched-to 3L using two cycles of the ALE process. (b) Output characteristics show linear behavior at low $V_{DS}$ for high gate voltages. (c) Transfer characteristics show good p-type conduction with high ON-current, low hysteresis, and large ON/OFF ratio. (d) Temperature-dependent transfer characteristics showing the linear region of device operation for $V_{GS} < -80$ V. (e) Extracted 2-probe and 4-probe field-effect mobility ($\mu_{FE}$) at $V_{GS} = -95$ V at different temperatures. High 4-probe hole mobility of 515 cm$^2$/V·s at room-temperature denotes the high-quality electrical properties of ALE-processed WSe$_2$ flakes.
5.11 **Sacrificial WSe$_2$ with ALE Processing (SWAP) technique for high-quality 2D devices.** (a) Schematic and process flow of a graphene device made with direct PCL transfer (control device) and using SWAP technique. The ALE process removes the sacrificial WSe$_2$ layer along with the polymer residue accumulated in the transfer and lithography steps. (b) Raman measurements show the relative cleanliness of the samples. Raman spectra of the control device show a broadened 2D peak with a small $I_{2D}/I_G$ ratio (2.2) indicating the presence of polymer residue on top of the graphene. Raman spectra for the direct transfer device with WSe$_2$ sacrificial layer has sharper 2D peak, compared to that of control device. The subsequent ALE process removes of the top WSe$_2$ layer along with the polymer residue and leads to a clean graphene surface indicated by the large $I_{2D}/I_G$ ratio suggesting minimal damage to the underlying graphene layer during the ALE process. (c) Comparison of sheet resistance ($R_{sh}$) for graphene devices made using direct transfer and SWAP technique showing better electrical characteristics of the SWAP device with Dirac peak close to 0 V, $2.5\times$ reduction in $R_{sh}$ at corresponding Dirac peaks, and $3\times$ increase in effective mobilities extracted using the Kim’s model (dark grey lines). (d) Room-temperature 4-probe field-effect mobility ($\mu_{FE}$) of the device made using SWAP technique shows minimum $3\times$ improvement at any carrier density ($n_{2D}$) over the devices fabricated with direct PCL transfer. The SWAP enabled device shows much better mobility from an unencapsulated graphene device.[196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 78, 206] .......................... 133

6.1 **Electrical and Raman characterization of the TOS-doped graphene.** (a) Device structure at different points in the measurement process. UV-ozone exposure only oxidizes the top WSe$_2$ layer creating monolayer TOS, which strongly p-dopes the underlying graphene layer. (b) Optical image of the TOS-doped graphene device with Hall-bar geometry used for Hall-effect and four-terminal measurements. (c) Raman characterization of the graphene device shows blueshifts in G and 2D peaks after oxidation, indicative of p-doping effect. The insets show a clear reduction in $I_{2D}/I_G$ and FWHM (G peak) after doping. (d) Comparison of $R_{4p}$ as a function of $V_{GS}$ for the graphene device in part b, before and after doping, depicts the strong reduction in $R_{4p}$ after doping. The Dirac peak shifts to higher $V_{GS}$ (beyond our measurement range) after doping indicating the degenerate p-doping of graphene. (e) Zoomed-in view of $R_{sh}$ shows 118 $\Omega/$ for doped sample at zero gate voltage. .......................... 140
6.2 **Transfer curves of a 1L WSe$_2$ device.** Transfer curves of a 1L WSe$_2$ device before and after UV-ozone at $V_{DS} = 1$ V. The Cr/Au contacts were formed by the edge contact method used for graphene devices. It shows an n-type semiconducting behavior before doping and the current drops to the noise floor of our measurement system (gray line indicates the leakage current, $I_{GS}$) after doping. .......................... 141

6.3 **Material characterization of monolayer TOS.** (a) Optical micrographs of monolayer (1L) WSe$_2$ before and after oxidation. After oxidation, 1L WSe$_2$ is transformed into monolayer tungsten oxyselenide (TOS) which makes it completely indiscernible from the surrounding SiO$_2$ due to higher transparency. (b) AFM images and height maps confirm that the change in optical contrast in part (a) is due to oxidation. Similar to 1L tungsten oxide (WO$_x$), 1L TOS also show increase in height after oxidation [156]. The AFM scans show that the UV-Ozone process results in uniform oxidation of sample without any pits or damage [167]. .......................... 142

6.4 **Self-limiting nature of UV-ozone oxidation.** Optical images highlighting the self-limiting nature of our oxidation process. Only a single layer of WSe$_2$ is removed despite using (5×) longer oxidation process (2.5 hrs) .......................... 143

6.5 **Structural characterization of monolayer TOS.** (a) Raman spectra of the 1L WSe$_2$ flake given in Figure S1 before and after oxidation. The Raman spectrum after oxidation does not show any additional peak near 800 cm$^{-1}$ which indicates amorphous nature of 1L TOS [222]. (b) TEM image of monolayer and few-layer WSe$_2$. (c,d) SAED patterns of 1L WSe$_2$, before and after UV-ozone process. Single-crystal diffraction patterns with zone axis [0 0 0 1] are clearly visible before UV-ozone and are completely removed after UV-ozone process, confirming the amorphous nature of the TOS layer. (e,f) Corresponding SAED patterns for few-layer WSe$_2$ before and after UV-ozone treatment. Unlike monolayer WSe$_2$, few-layers still show single crystalline patterns even after UV-ozone treatment, indicating the underlying layers are protected by self-limited TOS layer. .......................... 144

6.6 **Stoichiometry of monolayer TOS.** (a) The table shows the obtained atomic percentage of C, W, Se, O atoms, and Se/W and O/W ratio for monolayer WSe$_2$ from EDS measurements. The atomic ratio of Se/W for monolayer is significantly reduced (although not completely removed) while O/W ratio increases by the oxidation process, indicating that mostly the Se atoms are replaced by oxygen atoms. (b,c) XPS characterization of 1L WSe$_2$ and 1L-TOS indicating the W (4f) and Se (3d) peaks, respectively. A comparison of W (4f) peaks show a creation of new oxidation states in the TOS layer confirming the formation of W-O bonds and substoichiometric nature of the oxide. Se (3d) peaks also show a formation of Se-O bonds after oxidation with the appearance of broad peaks at higher binding energies. .......................... 145
6.7 Tunable p-type doping and hole mobilities in graphene with WSe$_2$ interlayers. (a) Hole density, (b) corresponding Fermi-level of graphene with respect to its $E_{CNP}$, and (c) sheet resistance as a function of back-gate bias, extracted from Hall-effect measurements. Increasing the number of WSe$_2$ interlayers between the TOS doping and graphene reduces the hole density in the graphene. High hole doping densities push the Fermi-level deep into the valence band as shown in the inset. (d,e) Hole mobility as a function of hole density for TOS-doped graphene with WSe$_2$ interlayers at room temperature. LA phonon-limited mobility is achieved at high hole densities in doped graphene with 3L and 4L WSe$_2$ interlayers, suggesting phonon-limited scattering due to the TOS doping. These mobilities are significantly higher than other chemical doping techniques with similar doping densities, highlighted in the bottom gray zone. This also extends the ideal hole mobility, previously achieved in h-BN-encapsulated graphene (black points), to higher hole densities. At extremely high hole densities for TOS-doped graphene without and with 1L WSe$_2$ interlayer, our work shows a significant mobility improvement compared to electrolyte gating.

6.8 Charge transfer mechanism for TOS-based doping. (a) Schematic for TOS-Gr system indicating the origin of charge transfer due to work function mismatch. In equilibrium, the high work function of TOS layer results in hole doping of the graphene layer. (b) Fits (gray line) using Equation (6.4) to the hole density extracted using Hall measurements (black squares). A TOS work function of 5.6 eV was extracted from the fits. The thickness of the films is indicated in parentheses. Our results for the TOS-doped graphene are highlighted in purple and red, showing the highest transmittance and comparable sheet resistance at remarkably small thickness (<5 nm). (c) Simulated sheet charge density in different layers in the TOS-WSe$_2$-Gr stacks for different WSe$_2$ interlayers. Most of the charge is mirrored in the Gr and TOS layers. (d) Raman map of 2D vs. G peak frequencies for pristine graphene (gray), graphene on 1L-TOS (red), and 1L-TOS on graphene (dark red). The gray and orange dotted lines indicate a peak shift of G and 2D peaks with different carrier density and strain, respectively [248]. A clear blueshift obtained from the G and 2D peaks in graphene along the $\epsilon = -0.2\%$ line can be seen after the TOS doping, irrespective of the TOS location. The shifts are in good agreements with previous work on p-type carrier modulation in graphene devices by Das et al.(shown in sky blue) [217].

6.9 Universal p-type doping of various semiconductors with TOS. Transfer characteristics for (a) SWCNT, and (b) 4L WSe$_2$, with and without 1L-TOS. A significant increase in the hole current is observed in all the devices with TOS layer. The insets show the structure of each device.
6.10 Transmittance of TOS-doped graphene. (a) Transmittance of CVD-grown 1L-WSe$_2$ on graphene before and after the UV-ozone oxidation. Shaded area indicate the standard deviation. The dashed line indicates the transmittance of intrinsic graphene (97.7%). Before UV-ozone, the transmittance remains around graphene’s intrinsic absorption for photon energies less than 1.4 eV. An excitonic band gap peak of WSe$_2$ is also seen at 1.67 eV. After UV-ozone treatment, the peak is reduced significantly along with an increment in the transmittance from 97.2 to 99.2% at telecommunication wavelength (1550 nm). The insets show different CVD stacks on quartz substrate to compare transparency in visible regime. (b) Optical transmission at 1550 nm as a function of sheet resistance for widely-used transparent conducting films [261, 262, 263, 264, 265]. The thickness of the films is indicated in parentheses. Our results for the TOS-doped graphene are highlighted in purple and red, showing the highest transmittance and comparable sheet resistance at remarkably small thickness (<5 nm).

A.1 Analyzing a symmetrically-doped 2D lateral p-n junction surrounded by an infinitely thick dielectric. The depletion region is divided into infinitely many line charges, which extend into the width of the plane containing the p-n junction. The vertical components of the electric field at the junction interface for two symmetrically placed line charges across the junction interface ($n$–line and $p$–line) cancel each other; resulting in only the horizontal component along the x-direction.

A.2 Application of method of image charges to model symmetrically-doped p-n junction. A symmetrically doped 2D lateral p-n junction with a depletion width of $x_{D,inf}$. The image charges enforce zero x-directed electric field at the depletion region edges; resulting in accurate calculation of $x_{D,inf}$. The blocks of charges were chosen in an order that enables easy analytical formulation of the x-directed electric field at the junction interface.

C.1 Schematic of the MoS$_2$ BGFET used for self-consistent simulations.

C.2 Simulated total current ($I_D$) for different $N_{IT}$ (thick solid line). The dotted line denotes the hopping current, and the thin solid line denotes the band (drift-diffusion) current. Note that the hopping current for $1 \times 10^{12}$ cm$^{-2}$ and the band current for $1 \times 10^{13}$ cm$^{-2}$ and $6 \times 10^{12}$ cm$^{-2}$ are smaller than $1 \times 10^{-12}$ A/µm. The dominant current component changes from band to hopping transport for $N_{IT}$ between $1 \times 10^{12}$ cm$^{-2}$ and $6 \times 10^{12}$ cm$^{-2}$.

C.3 Verification of the load-line model using self-consistent simulations for different trap densities. The simplified $\mu_{eff}$ model given in Eq. (6) provides nearly perfect fits to the simulation results.
D.1 Normalized lateral potential profile for different sheet densities ($N_{D,2D}$) for fixed $t_{gap}$. The x-axis is scaled with respect to lateral depletion width ($x_D$).
Acknowledgements

As the curtains fall on a wonderful five-year period, I can definitely say that embarking on this journey was one of the best decisions of my life. On the cusp of graduation, I believe that I have added many important skills such as critical thinking, first-principle analysis, and systematic problem-solving to my quiver. However, none of this would have been possible without the ample support that I received during this tenure from the many talented, inspirational, and empathetic people around me. Their continuous support and encouragement made this incredible journey possible.

First and foremost, please allow me to convey my sincere gratitude to my advisor, Prof. James Teherani, who played a pivotal role in enabling me to overcome the multitudes of challenges I faced during this process. I feel tremendously lucky to learn first-hand about scientific research from one of the absolute best. His pedagogy of identifying the fundamental bottleneck of a complex problem, dissecting and effectively brainstorming it helped me grow leaps and bounds as a researcher. He created an environment that made thinking outside the box and questioning existing theories conducive. By challenging me, rattling my complacency, actively participating in never-ending discussions and honing my potential, he graciously and rigorously pushed me towards my finish line. Outside of academics, his kind gestures like group lunches for publishing papers, a pat on the back for the smallest of ideas, and an invitation to thanksgiving dinner with his family shows how deeply he cared for his students. I will always be grateful to him for letting me be a part of his group.

I would like to express my profound gratitude to Prof. James Hone and Prof. P. James Schuck for their constant support and guidance. I feel privileged for getting an
opportunity to work with Prof. Hone who is one of the pioneers in the field of 2D materials. With his vision, subject knowledge, and exceptional experimental skills, his counsel was indispensable in tackling any problem statement. He always inspired me to challenge myself and will continue to do so. Similarly, my interactions with Prof. Schuck were always enthralling and his immense knowledge of optical characterization helped me immensely in my research. Prof. Schuck’s unending optimism and constant strive for path-breaking research encouraged me to push my own envelope. I would also like to thank my thesis committee members, Prof. John Kymissis and Prof. Dion Khodagholy, whose keen insights and constructive suggestions have been of great value in improving the overall quality of my research work.

A special shout-out to Dr. Minsup Choi, Dr. Younghun Jung, Dr. Bumho Kim, and Dr. Kaiyuan Yao with whom I worked closely during my stay at Columbia. Being embodiment themselves, they taught me the value of academic rigor, persistence, and perseverance in experimental research. Their never-say-never attitude motivated me to continuously work towards overcoming some of the toughest challenges I faced during my thesis. I would also like to thank Dr. Richard Gyuchull Han and Dr. Akiko Ueda who collaborated with me on the theoretical projects. Their intellect and years of experience helped me hone my modeling and simulation skills, which tremendously improved my research productivity. I would also like to acknowledge the technical and administrative support received from the staff at Columbia Nanoscience Initiative (CNI) as well as CUNY Advanced Science Research Center (ASRC), and the Columbia EE department. I would like to express my gratitude and appreciation to Elsa and Dennis from the EE department for taking care of all the administrative work and helping me whenever I reached out to them.

In addition, I would like to thank my colleagues in the Teherani, Hone, Dean, Schuck, Pasupathy, and Lipson groups for helping me during this journey. It has been an absolute pleasure to work beside these wonderful people. I am grateful to these excep-
tional researchers and my dear friends: Dr. En-Min Shih, Dr. Alexander Kerelsky, Dr. Sanghoon Chae, Dr. Brian S. Kim, Dr. Nate Finney, Punnu Jose Sebastian, Ipshita Datta, Emanuil Yanev, Baichang Li, Abhinandan Antony, and Kevin Kwock.

Finally, I would like to express my gratitude to my beloved friends and family. The morass of a long and winding PhD life often bogs one down. Apoorv Jindal, Abhinandan Borah, Peter Wei, Stephen Xia, Mark Barbet, and Lei Lei were always there to help me overcome these low phases and celebrate my little successes. My weekend escapades with them to several eateries in NYC and trekking trips are some of my most cherished memories. Most importantly, I would like to acknowledge the immense love and support I received from my parents and in-laws throughout this period. Thank you Sanket and Archana for bringing fun and laughter in an otherwise busy life of mine. Last but not the least, I dedicate this work to the one person who never ever gave up on me –my wife, Anjaly. Thank you for being my best cheerleader and worst critique. Now that you are trudging the same path of PhD, I look forward to being everything that you were to me and more.

Ankur Baburao Nipane
New York, NY
To my love and best friend, Anjaly

&

my family.
Chapter 1: 2D Devices - Motivation and Challenges

1.1 Why 2D materials?

Ever since graphene’s discovery in 2004, two-dimensional (2D) materials have morphed into a promising material platform to study novel electronic and optical properties arising due to quantum confinement [1, 2, 3]. Since then, many classes of 2D materials like transition metal dichalcogenides (TMDCs), black phosphorus, and hexagonal boron nitride (h-BN) have been identified as shown in Figure 1.1. 2D materials consist of metals, semimetals, semiconductors, and insulators, making them truly ubiquitous. These 2D materials offer wide range of electronic and optical properties that renders them useful for devices including field-effect transistors (FETs), photodetectors (visible/infrared), light-emitting diodes (LEDs), solar cells, and chemical sensors. Additionally, the fact that 2D materials are compatible with silicon technology and could be integrated into silicon CMOS process flow has aggressively propelled its investigation for beyond-Moore applications.

In this system, atomically-thin layers are held together by weak van der Waals forces. The lack of interfacial dangling bonds and presence of van der Waals bonding in out-of-plane direction makes these material surfaces agnostic and which enables these crystals to be ripped layer-by-layer and to be re-stacked in a configuration of ones choice, forming heterostructures [5]. This also allows easy integration with other material platforms such as silicon and GaN [6, 7]. The interactions between layers in a heterostructure, in effect, enable band structure engineering and thus allow exploration of further exotic phenomenon such as observation of Moiré patterns [8]. These materials also exhibit thickness-dependent properties which offer a range of band alignments and bandgaps.
Figure 1.1: **Spectrum of 2D materials.** 2D materials encompass a large range of bandgaps making them useful for applications across the spectral range. The atomic crystal of different 2D materials such as h-BN, MoS$_2$, BP, and graphene is provided along with their electrical properties [4].

for engineering different optical and electronic devices such as tunnel field-effect transistors (TFETs) and lasers [9, 10]. Furthermore, owing to the aforementioned properties, these material systems are actively being pursued other applications like chemical and biological sensors, image detectors, piezoelectric devices, optical modulators, memory devices and many more, as shown in Figure 1.2.

Next, we dive deeper into understanding why certain 2D materials such as TMDCs and graphene have created huge interest especially for next-generation electronic and photonic applications, respectively.

1.1.1 **TMDCs for electronic applications**

Arguably, one of the most important applications of TMDCs is in logic and memory devices. International Technology Roadmap for Semiconductors (ITRS) has long provided futuristic technological requirements for next-generation of logic and mem-
Figure 1.2: Different applications of 2D materials. Possible applications of 2D materials along with the complexity associated with 2D material-silicon process flow integration. The position of a circle shows the year when the particular product could potentially be introduced in the industry [11].

ory devices to continue scaling according to the famed Moore’s law — an observation that the number of transistors in an integrated circuit doubles every two years. For long, the semiconductor industry has managed to keep scaling alive and reached higher transistor density along with lower cost per chip following the specifications laid out by the ITRS roadmap. However, in recent years, this has visibly hit roadblocks due to both engineering and fundamental physics-based limitations. Till now, problems of patterning transistors with extremely small feature sizes (beyond the diffraction limit of conventional immersion photolithography) have been circumvented by smart engineering solutions such as self-aligned double and quadrature patterning (SADP/QADP). Although it is getting difficult to push the boundaries of conventional lithography techniques, advanced lithographic techniques such as using extreme ultraviolet lithography (EUV) still promises to push the limits of transistor scaling. This has come at the cost of
an increased R&D budget and therefore many companies have subsequently opted out of this race. Nevertheless, an even more pertaining challenge is the fundamental limitations to scaling set by quantum-mechanical effects that become prominent at such small dimensions. These quantum-mechanical effects often manifest themselves in enormous reliability challenges such as leakage current, dielectric breakdown, and device-to-device threshold voltage variation. The only path to overcome these challenges is to look for potential materials which could outperform Si at such small dimensions, and that’s the appeal for TMDCs such as MoS$_2$ and WSe$_2$.

Unlike graphene, TMDCs offer a wide range of bandgaps from 1~3 eV and thus are useful for both low-power and high-frequency applications. The presence of bandgap also results in large transconductance which is required by both logic and analog applications. Along with bandgap, one of the crucial advantages of using 2D materials is its atomically thin dimensions in out-of-plane direction ($t_{ch}$) which can go down to <1 nm for monolayer TMDCs. To understand the importance of small $t_{ch}$, let’s look at the electrostatic screening length ($\lambda$) that determines the minimal channel length below which short-channel effects overcome the device performance resulting in large leakage current and potential loss of switching characteristics [12]. Although several analytical formulations of $\lambda$ are available in literature, one of the simplest form which captures the essential physics can be given as $\lambda = \sqrt{\frac{\epsilon_{ch}t_{ox}t_{ch}}{\eta\epsilon_{ox}}}$, where $\epsilon_{ch}$ is the dielectric constant of the semiconductor channel, $\epsilon_{ox}$ is the dielectric constant of the gate oxide, $t_{ox}$ is the thickness of gate oxide, and $\eta$ varies depending upon number of gates [13]. Clearly, a lower $t_{ch}$ results in a smaller value of $\lambda$ and thus helps to avoid short channel effects even at sub-5 nm channel lengths [14]. Figure 1.3(a) show another manifestation of the same thing. Due to quantum confinement caused by the nanometer-scale thickness of TMDCs, the gates can more efficiently control the carrier density in the channel and thus leads to better gate electrostatics. Thus, the presence of bandgap and atomically-thin thickness leads to better OFF state/subthreshold characteristics of 2D transistors built from
Figure 1.3: **Unique properties of TMDCs.** (a) Comparison of mobile charge distribution along the thickness of the semiconductor for 3D vs 2D materials. Narrower mobile charge distribution in 2D materials allows better electrostatic control by the gate electrodes [13]. (b) Channel mobility ($\mu$) as a function of channel thickness ($t_{ch}$) for different materials. 2D materials offer much higher mobilities at smaller thickness [11]. (c) Energy-delay benchmarking of TMDC-based TFETs with other technologies showing higher switching speeds at lower switching energies [16].

TMDCs [15].

To highlight another significant advantage of TMDCs, we first ask a question — what is the ultimate limit of ultra-thin body silicon on insulator (UTB-SOI) technology? Figure 1.3(b) shows the effective channel mobility for sub-5 nm SOI technology. Even though Si has been shown to achieve high mobilities of the order of $\sim$500 cm$^2$/V·s at $t_{ch}$ of 3 nm, the mobilities drop sharply in the sub-3 nm regime. Moreover, producing Si at such small thickness also results in large interface trap density and lattice-mismatch induced strain during the epitaxial process. Thus, for all practical purpose, traditional SOI technology has reached a standstill with no clear path to achieve sub-3 nm thickness. Contrary to Si, TMDCs such as MoS$_2$ and WSe$_2$ maintains their high mobilities even at sub-1 nm thickness due to their layered nature. TMDCs have a chemical formula of MX$_2$, where M is a transition metal element (Mo, W, etc.) and X is a chalcogen element (S, Se, or Te, etc.). A single layer forms an X-M-X structure bound by covalent bonding and one layer to another layer is bound by weak van der Waals force. The weak van der Waals
force allows thinning down of TMDCs to monolayer limits without any dangling bonds or process-induced strain. Moreover, the absence of dangling bonds leads to a pristine surface which results in minimal surface-scattering, hence, giving record high mobilities. Theoretically, monolayer TMDCs can achieve mobilities of $\sim 1000$-$3000 \text{ cm}^2/\text{V} \cdot \text{s}$ at room temperature with a very small disorder of $<10^{11} \#/\text{cm}^2$.

Figure 1.3(c) shows an energy-delay benchmarking of TMDC-based tunnel-FETs compared to other competitive technologies. The out-of-plane van der Waals also allows easy integration of dissimilar TMDCs with low interface traps leading to high-quality tunnel FETs. TMDCs allow scaling of planar technology which allows TMDC-based FETs and TFETs to switch at faster speed with lower switching energy due to a significant decrease of parasitic capacitance associated with 3D FinFETs. Thus, the unique material properties of TMDCs put them at the forefront to evaluate their usability at scaled nodes.

1.1.2 Graphene for photonic and optoelectronic applications

Although the absence of bandgap in graphene has limited its applicability for transistor applications, its tremendous advantages in terms of mobilities and optical properties make it apt for optoelectronic and photonic applications [17]. The data communication industry is growing at an alarming rate with the addition of billions of new mobile phones, laptops, and IoT devices. Silicon photonics and optoelectronic devices have been single-handedly driving this industry forward. Currently, III-V materials are at the core of these devices which are then integrated with silicon photonics platform as photodetectors and optical modulators. However, the silicon photonics industry is inherently very expensive owing to the use of silicon-on-insulator (SOI) substrates and the need for epitaxial growth of III-V materials limits the ubiquity of photonic devices. The enormous growth of data traffic requires better photodetectors, modulators, and switches that support ultra-high bandwidth and have higher performance indexes [18]. In this light, graphene is the one-shot solution to all the problems with current photonic
integration.

Figure 1.4: **Graphene for optoelectronics.** (a) Band structure of graphene. Inset shows linear dispersion relation and absence of bandgap [19]. (b) Room temperature mobility versus electron density for different material systems. Graphene shows much higher mobilities than other systems [20]. (c) Transmittance as function of wavelength in the visible regime for different materials useful for transparent conductor application [17]. (d) Variation of sheet resistance with thickness. Graphene provides higher transmittance and low sheet resistance at smaller thickness [17].

Graphene has a unique linear dispersion relation with zero bandgap as illustrated in Figure 1.4(a). This results in ultra-high electron mobility (upwards of \( \sim10000 \) cm\(^2\)/V·s), as shown in Figure 1.4(b) at room temperatures which is significantly higher than any other currently known material system. Moreover, the absence of bandgap in graphene allows the fabrication of ultrafast broadband photodetectors that absorb incoming photons of any wavelengths and can work at telecommunication wavelengths — a significant advantage over Si photodetectors. From the transmitter side, the excellent electro-absorption [21] and electro-refraction [22] properties of graphene makes them ideal for efficient modulators. These modulators are used to embed information into a carrier signal using either amplitude modulation or by modulating the phase of the carrier sig-
nals. Graphene also exhibits strong odd-order nonlinearity owing to its centrosymmetry and linear dispersion [23]. This enables its use in ultrafast lasers using the mode-locking technique and in optical frequency converters using four-wave mixing (FWM), both essential components in transceivers. A major technological advantage of graphene over III-V materials is that graphene does not require epitaxial growth and thus can be directly integrated with silicon photonics.

In addition to these active devices, the high mobility and low transmittance of graphene allow its usability as a transparent electrode in optoelectronic and photonic applications. Indium tin oxide (ITO) is currently the most dominant material for such applications. Even though ITO has higher performance in the visible range, its low transmittance at telecommunication wavelengths leaves it unfit for higher-frequency operations. Moreover, ITO is very brittle and cracks when used in flexible devices. On the contrary, graphene offers a highly flexible platform with way higher transmittance (> 97%) and very low sheet resistance (< 1000 Ω/□) compared to ITO across the frequency range as illustrated in Figure 1.4(c) and (d), making it an ideal material for transparent conductors. Furthermore, the unique electro-refractive and electro-absorption properties of graphene along with its atomically-thin dimensions make it highly useful for on-chip silicon photonic integration to make ultra-low-loss transparent electrodes and modulators [24].

1.2 Current challenges in making 2D devices

Having understood the enormous advantages of 2D materials, we further investigate the current challenges faced in making high-performance 2D devices. A recent review paper from Akinwande et al. highlighted the major challenges in the integration of 2D materials specifically for electronic applications, as shown in Figure 1.5 [11]. We can combine these challenges into three major categories: (1) device fabrication issues, (2) lack of device models, and (3) insufficient wafer-scale growth & transfer techniques.
Here, we delve deeper into the first two categories which are pertaining to this thesis.

Figure 1.5: **Major challenges associated with 2D materials.** Timeline associated with different challenges associated with 2D materials. Currently, the major challenges include lack of large-scale growth & transfer, fabrication issues associated with doping and contacts, and design principles for better device design [11].

### 1.2.1 Fabrication challenges

The atomically-thin dimensions of 2D materials which gives rise to all the interesting properties also cause significant fabrication challenges when working with these materials. This lack of thickness in 2D materials renders the conventional bulk doping, patterning, and etching techniques ineffective which leads to significant issues involving contact resistance, irregular channel geometries, and lack of thickness control. In the absence of wide availability of high-quality wafer-scale 2D materials, mechanical exfoliation is the most employed method which researchers employ to make 2D materials-based devices.
Figure 1.6: **Disorder in 2D materials.** Internal and external disorders which affect carrier mobility of 2D materials [25].

[5]. These layers are often exfoliated directly on dielectric substrates such as SiO$_2$ and Al$_2$O$_3$. Even though the exfoliated flakes do not have dangling bonds, the large surface-to-volume ratio of 2D materials leads to significant interaction between the 2D material and the dielectric environment. Thus, in practice, the observed room-temperature mobility in 2D materials is way lower than expected due to surface roughness and charged impurities as shown in Figure 1.7(a). Furthermore, typical methods of growing 2D crystals do not produce high-quality crystals and this generally has large defect densities ($>10^{12}$ #/cm$^2$). Thus, these two factors are typically considered to be disorders in 2D materials where the intrinsic disorder arising due to defects in 2D material crystal and exterior arising due to the environment, as shown in Figure 1.6.

The only way to avoid internal disorders is innovative and careful crystal growth techniques like the flux growth technique [25] and chemical vapor deposition (CVD) technique for minimizing aforementioned defects. The obvious way to minimize external disorder is to isolate the 2D material from the substrate with charge impurities and
Figure 1.7: **Scattering mechanisms in 2D materials**

(a) Theoretical comparison of different scattering mechanisms affecting mobility of monolayer MoS$_2$ as a function of temperature. At room temperature, the mobility is limited by optical phonon scattering (red line). However, low temperature mobilities is limited by impurity scattering (blue line) [26].

(b) Experimental mobilities on different encapsulation (substrate+passivation) layers. Low-temperature mobility is highest for h-BN encapsulation [25].

ambient conditions. Figure 1.7(b) shows the impact of various encapsulations on the extracted mobilities in MoS$_2$. The best improvement in mobility was achieved by encapsulating the 2D material with h-BN on top and bottom, shielding it from all external disorders. h-BN, being a van der Waals material itself, aided in fabricating h-BN-2D material-h-BN heterostructures, resulting in record-high mobilities. However, h-BN has a lower dielectric constant and thus further improvement in fabrication techniques for open-surface devices is required to allow integration of high-$k$ dielectrics. These devices are also useful in making sensing applications where a top h-BN layer can hamper the sensitivity of the active 2D layer.

Another challenge with the current fabrication methodology is the lack of a suitable dopants. The development of precisely tunable and damage-free doping methods is key
to fabricating advanced electronic and photonic devices. For instance, precise control of doping in silicon is used to minimize contact resistance, suppress leakage currents, and control the threshold voltage in ultra-scaled MOSFETs. Dopants in bulk semiconductors are substitutional in nature, whereby a dopant atom takes the place of a host atom in a semiconductor crystal by a process such as dopant diffusion or ion implantation, followed by a high-temperature annealing step that activates the dopants in the lattice. However, a plethora of issues arise when extending them to other 2D materials such as the requirement of higher activation energy due to large bandgap (MoS$_2$, WSe$_2$, etc.), low thermal budget, and implantation-induced damage [27, 28]. In the absence of an area-selective doping method, 2D devices often suffer from large contact resistance which hampers the device performance, as shown in Figure 1.8(a). This effect worsens due to fabrication-induced damage introduced by e-beam lithography and direct metal-evaporation, shown in Figure 1.8(b). Thus, suitable doping and metallization techniques are required to reduce contact resistance in 2D materials.

1.2.2 Lack of appropriate device models

Over the past decade, numerous experimental and theoretical works have primarily focused on understanding the fundamental electrical and optical properties of 2D materials. However, the development of analytical models to understand the role of key device parameters on the overall performance of 2D devices has been missing. In bulk devices, analytical models for depletion width, contact resistance, threshold voltage, and I-V characteristics provide great insights into the working of pn-junction diodes and MOSFETs. The lack of such models for 2D devices inhibits the understanding of different limiting factors to achieve the maximum potential of 2D devices. Simply extending models developed for bulk semiconductors to 2D materials can result in erroneous interpretation of experimental results. For example, the presence of large contact resistance in 2D materials results in the underestimation of channel mobilities. Thus, it becomes
Figure 1.8: **Issue of contact resistance in 2D materials.** (a) Contact resistance \( (R_C) \) as a function of carrier density. Compared to other materials, TMDCs show two-order of magnitude higher contact resistance [13]. (b) Cross-sectional transmission electron microscope (TEM) image of metal-semiconductor interface shows fabrication-induced damage [29].
imperative to understand the role of factors such as lateral electrostatics, contact resistance, and global back gating to extract important device parameters that are critical for understanding and benchmarking 2D devices.

1.3 Thesis Overview

In this thesis, we first provide analytical models for the electric field, electrostatic potential, and depletion width in 2D p-n and metal-2D semiconductor in chapter 2 to understand the effect of doping density and dielectric environment on lateral electrostatics. In chapter 3, we studied the role of contact and channel region in different operating regimes of 2D devices where it was established that opposite to general understanding, the ON current is limited by contact resistance which can be reduced by degenerate doping next to the metal-semiconductor junction. Endowed with this knowledge, we then highlighted the key limitations of current experimental techniques to extract important device parameters in chapter 4. Subsequently, chapter 5 introduces a novel etching technique that enables easy fabrication of high-performance 2D devices built from graphene and WSe$_2$. While WSe$_2$ devices fabricated using the ALE method demonstrate the highest mobilities ever reported in the literature, chapter 6 introduces TOS-doped graphene which signifies a major step towards establishing graphene as a potential substitute of ITO as a transparent electrode in next-generation photonic devices.
Chapter 2: Electrostatic Modeling of Lateral Junctions in 2D Semiconductors

P-n and metal-semiconductor junctions act as building blocks of any 2D semiconductor based devices [30]. Several important electronic devices, such as field effect transistors (FETs), light-emitting diodes (LEDs) and solar cells have p-n junctions at the heart of their operation. An in-depth understanding of 2D lateral p-n junction electrostatics is critical for designing high-performance devices using 2D materials. But unlike the case of bulk semiconductors, these materials often form 2D lateral junctions with non-zero out-of-plane electric field [31]. A proper understanding of these electric fields and related electrostatic potential profiles is critical for design of high-performance electrical and photonic devices with 2D materials. Although significant research is on-going to attain high-quality 2D lateral junctions where experimental techniques involving selective area growth [32] and selective material functionalization [33] have been employed to realize lateral p-n junctions, simple analytical models are still missing. This becomes even more critical in reducing contact resistance by optimizing edge and top contacts in 2D transistors as the tunneling current at the metal-semiconductor interface is strongly affected by the electrostatic field profile which limits the ON current of 2D transistors, as discussed in Chapter 3. In this chapter, we develop an analytical model for determining the electrostatics of 2D lateral p-n and metal-semiconductor junctions.

1This chapter is based on [34, 35] and reproduced with the permission from AIP Publishing.
2.1 Electrostatics of 2D lateral p-n junctions

Although charge in 2D materials is tightly confined to the plane of these materials, a significant out-of-plane electric field is typically present due to their large surface-to-volume ratio, shown in Figure 2.1(a) [36, 37]. This leads to weaker screening of charge carriers in these 2D lateral p-n junctions. The weaker screening in 2D lateral p-n junctions leads to larger depletion widths compared to bulk (3D) junctions – an effect that has been experimentally determined by optical characterization [38] and Kelvin Probe Force Microscopy (KPFM) [39]. \textit{The non-zero out-of-plane electric field also makes analytical calculation of the electric field, electrostatic potential, and depletion width challenging since the conventional Poisson formulation ($\nabla \cdot \vec{E} = \frac{\rho}{\varepsilon}$) within the depletion approximation can no longer be simplified to a one-dimensional derivative as in the case of bulk p-n junctions.}

For the 2D lateral p-n junction shown in Figure 2.1(a), $\nabla \cdot \vec{E} \rightarrow \frac{2D}{\varepsilon} \frac{\partial E_z}{\partial x} + \frac{\partial E_z}{\partial z}$.

Several interesting methods for analyzing the electrostatics of 2D lateral p-n junctions have been previously proposed to overcome the difficulties associated with a non-zero out-of-plane electric field. Gurugubelli et al. developed an analytical formulation of the electrostatic potential and depletion width by solving the Poisson equation in two-dimensions (2D Poisson equation) for symmetrically doped thin nano-films [40]. Their methodology requires determination of exact boundary conditions at the depletion edges, junction interface, and top and bottom surfaces of the 2D lateral p-n junction; however, these boundary conditions are not straightforward to determine for asymmetric doping configurations, especially along the junction interface (right image of Figure 2.1(b)). Similar issues are faced for determining boundary conditions for the top and bottom surfaces as they neither satisfy Neumann nor Dirichlet boundary conditions. This limits the applicability of their methodology to symmetric junctions. In a similar method, Achoyan et al. solved the 2D Poisson equation by using the method of conformal mapping, which maps the out-of-plane electric field to an planar equivalent
and results in relatively easy boundary conditions for solving the 2D Poisson equation [41, 42]. However, just like the previous method, conformal mapping is also limited to symmetric junctions and, hence, has a limited applicability. To avoid the complexity associated with the determination of exact boundary conditions for solving the 2D Poisson equation, Gharekhanlou et al. modeled the depletion charge as a series of infinitesimally thin line charges [43, 44]. Although their method is intuitive, it results in incorrect electric field and electrostatic potential profiles as the method does not honor the zero-electric-field boundary condition for the quasi-neutral regions. Their expressions yield non-zero electric fields in the quasi-neutral regions along with non-monotonic potential profiles within the depletion region, which are non-physical and do not match simulation and experimental results [45, 46]. Hesameddin et al. improved the line-charge method by considering the screening induced by mobile carriers in the quasi-neutral regions [47]. They enforced the zero-electric-field boundary condition in the quasi-neutral regions by using the method of image charges. Their study, however, was limited to symmetric junctions and approximated the image charge by considering only a single reflection of the depletion charge on each side of the 2D lateral p-n junction, which results in small, but non-zero electric fields at the depletion edges (discussed in further detail in Section 2.1.2 and Figure 2.3).

2.1.1 Formulation of the 2D lateral p-n junction as a collection of 1D line charges

Given the limitations of previous techniques, we present a comprehensive study of the electrostatics of symmetric and asymmetric 2D lateral p-n junctions. We derive analytical expressions for the electric field, electrostatic potential, and depletion width for abrupt 2D lateral p-n junction formed using 2D semiconductors with uniform (but not necessarily equal) doping on each side of the junction. Although abrupt, uniform doping is an idealization of real p-n junctions, this assumption enables analytical solutions for various electrostatic parameters that provide important insights into device opera-
Figure 2.1: **Electric field profiles in 2D lateral p-n junctions.** (a) Schematic of a 2D lateral p-n junction showing depletion and quasi-neutral regions with different dielectric media above ($\epsilon_1$) and below ($\epsilon_2$) the plane containing the 2D lateral p-n junction. The out-of-plane electric field is prominent in these junctions due to the large surface to volume ratio. (b) Electric field lines for symmetric and asymmetric doping configurations. For asymmetric doping, a non-zero tangential ($z$) electric field exists at the junction interface, which prevents the use of previous analytical methods for asymmetrically doped junctions.

Asymmetric doping, just as is typically done in analyzing bulk (3D) junctions. In addition, our analytical model assumes a continuum model of doping (as opposed to considering discrete dopants). Interestingly, the total number of dopants within the depletion region of a 2D lateral p-n junction is weakly (logarithmically) dependent on the doping density, as the total number of dopants in the depletion region depends only on the built-in potential of the junction, $\phi_{bi}$. For instance, the number of dopants in the depletion region of a 1
\( \mu \text{m} \) wide 2D p-n junction with SiO\(_2\) and air as surrounding dielectric media and having a doping density of \(10^{12} \, \text{#/cm}^2\) is \(\sim 350\), which implies that a slight variation in the dopant number and spatial distribution does not affect the overall junction electrostatics. Although the total number of dopants in the 2D depletion region is weakly dependent on doping density, higher doping densities result in extremely small depletion widths on the order of the average distance between dopants. Hence, the continuum model breaks down for narrow, highly doped, and strongly forward-biased 2D lateral p-n junctions and therefore, cannot be accurately modeled using our method.

For wide sheets of 2D semiconductors in which the thickness of the sheet \( (T_{\text{sheet}}) \) is much smaller than the depletion width \( (x_D) \), i.e. \( T_{\text{sheet}} \ll x_D \) and the width of the sheet \( (W_{\text{sheet}}) \) is much larger than the depletion width, i.e. \( W_{\text{sheet}} \gg x_D \), we can model the depletion region as a collection of infinite line charges extending into the width of the 2D plane as shown in Figure 2.2(a). A similar methodology was used by Hesameddin et al. to study the electrostatics of bulk semiconductor junctions by dividing them into many infinitesimally thin sheet charges as shown in Figure 2.2(b) [47]. Figure 2.2(c) and Figure 2.2(d) highlight the difference in the electric field profiles for a single infinite line charge and a single infinite sheet charge. The electric field for an infinite sheet charge is independent of distance from the sheet charge, which simplifies the analysis of bulk junction electrostatics, whereas the electric field for an infinite line charge is inversely dependent on the radial distance from the line, which makes analysis of 2D-junction electrostatics more complicated. In the following subsections, we first find the electric field profile due to a combination of these line charges representing the space charge in depletion region of a 2D lateral p-n junction. We then calculate the electrostatic potential profile by integrating the electric field across the depletion region of the 2D lateral p-n junction. Finally, the depletion width is determined by demanding that the total potential difference across the junction is equal to the built-in potential of the junction.
The electric field for a single infinite line charge is given by

\[ \vec{E}(\vec{r}) \mid_{\text{line}} = \frac{qN_{1D}}{2\pi\varepsilon_{\text{eff}}||r||} \hat{r}, \tag{2.1} \]

where \( \vec{r} \) is the radial vector from the line, \( qN_{1D} \) is the line charge density (in C/length), and \( \varepsilon_{\text{eff}} \) is the effective dielectric constant. For 2D materials, the effective dielectric constant in Equation (2.1) is highly affected by the surrounding dielectrics [48, 49]. Sarma
et al. showed that for the limiting case of negligible thickness of 2D semiconductors ($T_{\text{sheet}} \ll W_{\text{sheet}}, L_{\text{sheet}}$), $\epsilon_{\text{eff}}$ can be estimated as the arithmetic mean of dielectric constants of the surrounding media above ($\epsilon_1$) and below ($\epsilon_2$) the 2D semiconductor [50] and can be expressed as

$$\epsilon_{\text{eff}} = \frac{(\epsilon_1 + \epsilon_2)}{2}. \quad (2.2)$$

where $\epsilon_1$ and $\epsilon_2$ are dielectric constants of the media on either side of 2D semiconductor. Since the effective dielectric constant for 2D materials is completely determined by the surrounding dielectric media, our method also works for 2D lateral heterojunctions. The results in the manuscript are derived assuming that the dielectric constant of the surrounding dielectric materials is constant, i.e. the dielectric materials are isotropic and extend to infinity. This assumption models 2D p-n junctions with thick dielectric media above (such as air) and below (such as SiO$_2$) the 2D layer. A detailed analysis for symmetrically doped junctions, discussed in Section 2.2, suggests that dielectric thicknesses greater than $\sim 1.5 \times$ the 2D depletion width ($x_{2D,\text{sym}}$) do not impact the electrostatics of the 2D lateral p-n junctions. For dielectric thicknesses less than $\sim 1.5x_{2D,\text{sym}}$, the layers above the top dielectric (and below the bottom dielectric) impact the junction electrostatics, and numerical simulation is required.

For a single infinite line charge lying along the $y$-axis and positioned at ($x = x_0, z = 0$), the $x$-directed electric field along the $z = 0$ plane is given by

$$E_x(x, x_0)_{\text{line}} = \frac{qN_{1D}}{2\pi \epsilon_{\text{eff}}(x - x_0)}. \quad (2.3)$$

If we divide a block of sheet charge representing the space charge region of a 2D lateral p-n junction that extends from $x_1$ to $x_2$ into infinitely many line charges, the line charge density for a single line is given by

$$qN_{1D} = \lim_{n \to \infty} qN_{2D} \frac{(x_2 - x_1)}{n} = qN_{2D} dx_0, \quad (2.4)$$
where $q_{N2D}$ is the sheet charge density of the block (in C/area). The $x$-directed electric field due to one of these infinitely many line charges is given by

$$dE_x(x, x_0) = \frac{q_{N2D}dx_0}{2\pi\varepsilon_{eff}(x - x_0)},$$

(2.5)

such that the $x$-directed electric field for the entire block of charge is equal to

$$E_x(x)|_{block} = \int dE_x(x, x_0) = \int_{x_1}^{x_2} \frac{q_{N2D}dx_0}{2\pi\varepsilon_{eff}(x - x_0)} = \lambda \ln \left| \frac{x - x_1}{x - x_2} \right|,$$

(2.6)

where $\lambda = \frac{q_{N2D}}{2\pi\varepsilon_{eff}}$.

For the 2D lateral p-n junction shown in Figure 2.2(a), the combined electric field due to the p and n block of depletion charge is given by

$$E_x(x)|_{depletion \ charge} = -\lambda_N^{-} \ln \left| \frac{x + x_p}{x} \right| + \lambda_N^{+} \ln \left| \frac{x}{x - x_n} \right|$$

(2.7)

where $\lambda_N^{-} = \frac{q_{N_{A,2D}}}{2\pi\varepsilon_{eff}}$ and $\lambda_N^{+} = \frac{q_{N_{D,2D}}}{2\pi\varepsilon_{eff}}$. $N_{A,2D}$ and $N_{D,2D}$ are the sheet charge densities of ionized acceptors and donors of the p and n regions, respectively, and the negative sign in front of the $\lambda_N^{-}$ term, accounts for the negative charge of ionized acceptors. $x_p$ and $x_n$ are depletion widths for the p and n sides of the junction.

2.1.2 Derivation of the electric field and electrostatic potential as a function of position using the image charge method

Although Equation (2.7) represents the standard depletion approximation, it incorrectly yields infinite electric fields at the edges of the depletion region, non-zero electric field in quasi-neutral regions, and a non-monotonic electrostatic potential profile within depletion region as shown in Figure 2.3(a). This is in sharp contrast with experimental findings and, hence, illustrates the limitation of the standard depletion approximation for analyzing 2D lateral p-n junctions [38, 51]. To enforce the $E_x = 0$ boundary condition at the edges of depletion region, we use the method of image charges proposed by
Hesameddin et al. [47]. The method of image charges captures electric field screening by mobile carriers in the quasi-neutral regions. The image charges, however, do not represent a real charge distribution in the quasi-neutral regions, and the electric field and potential calculated using the method of image charges is only valid in the depletion region. To enforce $E_x(-x_p) = 0$ boundary condition, the image charge method mirrors the charge distribution across the $x = -x_p$ plane, as shown in Figure 2.3(b). To enforce the second boundary condition, i.e. $E_x(x_n) = 0$, we reflect the charge distribution across the $x = x_n$ plane including the mirror charge due to the first boundary condition as shown in Figure 2.3(c). While this leads to zero electric field at $x = x_n$, the electric field at $x = -x_p$ is small, but non-zero and the first boundary condition is no longer satisfied. The presence of the two $E_x = 0$ boundary conditions leads to an infinite number of reflections and, therefore, an infinite number of repeating blocks of image charges as depicted in Figure 2.3(d). The image charge method yields a monotonic potential profile across the depletion region and zero electric field at both the depletion region edges. It is worth repeating that these image charges are not real charges but used only to establish the desired boundary conditions.

Analytical formulation of electric field

From Equation (2.6), the $x$-directed electric field due to all the $p$-type blocks of charge given in Figure 2.3(d) can be written as

$$E_x(x)|_{p \text{ blocks}} = -\lambda_{N_A^-} \sum_{k=-\infty}^{\infty} \ln \left[ \frac{x - 2k(x_p + x_n) + 2x_p}{x - 2k(x_p + x_n)} \right], \text{for } -x_p < x < x_n. \quad (2.8)$$

Here, $\lambda_{N_A^-}$ is a constant for all $p$-type charge blocks as it only depends on the constant $p$-side sheet charge density ($N_{A,2D}^-$). After some algebraic manipulations, the above expression can be written as

$$E_x(x)|_{p \text{ blocks}} = -\lambda_{N_A^-} \left[ \ln \left[ \frac{x + 2x_p}{x} \right] + \sum_{k=1}^{\infty} \ln \left[ \frac{(\frac{x+x_p}{2x_D})^2 - k^2}{\left(\frac{x}{2x_D}\right)^2 - k^2} \right] \right], \text{for } -x_p < x < x_n. \quad (2.9)$$
Figure 2.3: Qualitative analysis of the charge distribution, electric field, and electrostatic potential across a 2D lateral p-n junction with symmetric doping. (a) Charge distribution, electric field, and potential profile using the standard depletion approximation. The standard depletion approximation results in infinite electric fields at the edges of depletion region and a non-monotonic potential across the junction. (b) Enforcement of $E_x(-x_p) = 0$ using the method of image charges. The charge to the right is reflected across the $x = -x_p$ plane. (c) Extension of (b) to enforce $E_x(x_n) = 0$ by reflecting the charge to the left across the $x = x_n$ plane. However, this leads to a small but non-zero electric field at $x = -x_p$. (d) Extension of the image charge method to an infinite number of reflections to satisfy both $E_x(-x_p') = 0$ and $E_x(x_n') = 0$ boundary conditions. This also results in monotonic electrostatic potential throughout the depletion region.

Using the Euler infinite product representation of the sine function [52], i.e., $\frac{\sin(x)}{x} =$
\[
\prod_{k=1}^{\infty} \left(1 - \frac{x^2}{k^2 \pi^2}\right)
\] along with the logarithmic property \(\sum \ln(f(x)) = \ln(\prod f(x))\), we get

\[
E_x(x)|_{\text{p blocks}} = -\lambda_{Np}' \ln \left| \frac{x + 2x_p}{x} \right| - \lambda_{Np}' \ln \left| \frac{\sin \left( \frac{\pi(x+2x_p)}{2x_D} \right)}{\sin \left( \frac{\pi x}{2x_D} \right)} \right|
\]

(2.10)

\[
E_x(x)|_{\text{n blocks}} = -\lambda_{Nn}' \ln \left| \frac{\sin \left( \frac{\pi(x+2x_n')}{2x_D} \right)}{\sin \left( \frac{\pi x}{2x_D} \right)} \right|
\]

(2.11)

where the total depletion length \((x_D)\) can be given as

\[
x_D = x_p + x_n.
\]

Similarly, the \(x\)-directed electric field due to the n-type blocks of charge can be expressed as

\[
E_x(x)|_{\text{n blocks}} = \lambda_{Nd}' \ln \left| \frac{\sin \left( \frac{\pi x}{2x_D} \right)}{\sin \left( \frac{\pi(x-2x_n')}{2x_D} \right)} \right|
\]

(2.12)

Rewriting these equations for \(n\) and \(p\)-type blocks in a scaled coordinate system where \(x' = \frac{x}{x_D}, x'_p = \frac{x_p}{x_D}\), and \(x'_n = \frac{x_n}{x_D}\) yields

\[
E_x(x')|_{\text{p blocks}} = -\lambda_{Np}' \ln \left| \frac{\sin \left( \frac{\pi}{2}(x' + 2x'_p) \right)}{\sin \left( \frac{\pi}{2} x' \right)} \right|
\]

(2.13)

\[
E_x(x')|_{\text{n blocks}} = \lambda_{Nd}' \ln \left| \frac{\sin \left( \frac{\pi}{2} x' \right)}{\sin \left( \frac{\pi}{2}(x' - 2x'_n) \right)} \right|
\]

(2.14)
and the total $x$-directed electric field is equal to

$$E_x(x') = E_x(x')|_{p \text{ blocks}} + E_x(x')|_{n \text{ blocks}}$$

$$= -\lambda_{N_A}^- \ln \left| \frac{\sin\left(\frac{\pi}{2} (x' + 2x_p')\right)}{\sin\left(\frac{\pi}{2} x'\right)} \right| + \lambda_{N_D}^+ \ln \left| \frac{\sin\left(\frac{\pi}{2} x'\right)}{\sin\left(\frac{\pi}{2} (x' - 2x_n')\right)} \right|, \text{ for } -x_p' < x' < x_n'.$$

(2.15)

By rewriting Eq. (2.11) as $x_p' + x_n' = 1$ and using the fact that $\sin\left(\frac{\pi}{2} (x' - 2x_n')\right) = \sin\left(\frac{\pi}{2} (x' + 2x_p')\right)$, the total $x$-directed field can be simplified to

$$E_x(x') = (\lambda_{N_A}^- + \lambda_{N_D}^+) \ln \left| \frac{\sin\left(\frac{\pi}{2} x'\right)}{\sin\left(\frac{\pi}{2} (x' + 2x_p')\right)} \right|, \text{ for } -x_p' < x' < x_n'.$$  

(2.16)

The depletion widths of the $p$ and $n$ regions can be expressed solely in terms of acceptor and donor doping densities by making use of charge neutrality of the overall structure (i.e. $qN_{A,2D}^-x_p = qN_{D,2D}^+x_n$) in combination with the scaled version of Equation (2.11):

$$x_p' = \frac{N_{D,2D}^+}{N_{A,2D}^- + N_{D,2D}^+} \text{ and } x_n' = \frac{N_{A,2D}^-}{N_{A,2D}^- + N_{D,2D}^+}.$$  

(2.17)

Hence, the $x$-directed electric field $E_x(x')$ for arbitrary doping configurations can be expressed as

$$E_x(x') = (\lambda_{N_A}^- + \lambda_{N_D}^+) \ln \left| \frac{\sin\left(\frac{\pi}{2} x'\right)}{\sin\left(\frac{\pi}{2} \left( x' + \frac{2N_{D,2D}^+}{N_{A,2D}^- + N_{D,2D}^+} \right) \right)} \right|, \text{ for } -x_p' < x' < x_n'.$$  

(2.18)

Finally, the $x$-directed electric field $E_x(x)$ can be expressed in the non-scaled coordinate
system by using Equation (2.18) and the scaled version of Equation (2.11):

\[ E_x(x) = \frac{q(N_{A,2D}^+ + N_{D,2D}^+)}{2\pi \epsilon_{eff}} \ln \left| \frac{\sin \left( \frac{\pi x}{2x_D} \right)}{\sin \left( \frac{x}{2} + \frac{2N_{D,2D}^+}{N_{A,2D}^+ + N_{D,2D}^+} \right)} \right|, \text{ for } -x_p < x < x_n. \] (2.19)

### Analytical formulation of electrostatic potential

Next, we derive an analytical expression for the electrostatic potential across the depletion region. The electric field can be found by integrating the \(x\)-directed electric field

\[ V(x) = -\int E_x(x)dx, \text{ for } -x_p < x < x_n. \] (2.20)

By transforming this expression to the scaled coordinate system where \(x' = \frac{x}{x_D}\), the electrostatic potential can be expressed as

\[ V(x') = -x_D \int E_x(x')dx', \text{ for } -x'_p < x' < x'_n. \] (2.21)

Substituting the electric field expression from Equation (2.16) into Equation (2.21) gives

\[ V(x') = -x_D \left( \lambda_{N_A^+} + \lambda_{N_D^+} \right) \int \ln \left| \frac{\sin \left( \frac{\pi x'}{2} \right)}{\sin \left( \frac{x' + 2x'_p}{2} \right)} \right| dx', \text{ for } -x'_p < x' < x'_n. \] (2.22)

Given that \( \ln(x) = \ln(|x|e^{i\theta}) = \ln |x| + i\theta \), then \( \ln |x| = \text{Re}\{\ln(x)\} \). Therefore, we can rewrite Equation (2.22) as

\[ V(x') = -x_D \left( \lambda_{N_A^+} + \lambda_{N_D^+} \right) \text{Re}\left\{ \int \ln \left( \sin \left( \frac{\pi x'}{2} \right) \right) dx' - \int \ln \left( \sin \left( \frac{\pi (x' + 2x'_p)}{2} \right) \right) dx' \right\}, \text{ for } -x'_p < x' < x'_n. \] (2.23)
Using integration by parts, the two integrals can be expressed as

\[
\int \ln \left( \frac{\sin(\pi x')}{2} \right) \, dx' = -x' \ln(2) + i \frac{\pi}{2} x' - i \frac{\pi}{4} (x')^2 + i \frac{\text{Li}_2(e^{i\pi x'})}{\pi} \tag{2.24a}
\]

\[
\int \ln \left( \frac{\pi (x' + 2x_p')}{2} \right) \, dx' = -(x' + 2x_p') \ln(2) + i \frac{\pi}{2} (x' + 2x_p') - i \frac{\pi}{4} (x' + 2x_p')^2 + i \frac{\text{Li}_2(e^{i\pi(x'+2x_p')})}{\pi} \tag{2.24b}
\]

where \( \text{Li}_2(z) \) is the second-order Polylog function given by

\[
\text{Li}_n(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n}. \tag{2.25}
\]

Substituting Equation (2.24a) and Equation (2.24b) into Equation (2.23) and obtaining the real part results in

\[
V(x') = -x_D(\lambda_{N^-} + \lambda_{N^+}) \left[ 2x_p' \ln(2) + \frac{\text{Re} \left( i \text{Li}_2(e^{i\pi x'}) - i \text{Li}_2(e^{i\pi(x'+2x_p')}) \right)}{\pi} \right] + C,
\]

for \(-x_p' < x' < x_p'\).

Next, we simplify the Polylog terms by expanding them using Equation (2.25), which gives

\[
\text{Re} \left\{ i \text{Li}_2(e^{i\pi x'}) - i \text{Li}_2(e^{i\pi(x'+2x_p')}) \right\} = \text{Re} \left\{ \sum_{k=1}^{\infty} \frac{(e^{i\pi/2})(e^{i\pi x'})^k}{k^2} - \sum_{k=1}^{\infty} \frac{(e^{i\pi/2})(e^{i\pi(x'+2x_p')})^k}{k^2} \right\}. \tag{2.27}
\]
Taking the real part of the individual components

\[
\text{Re} \left\{ \sum_{k=1}^{\infty} \left( \frac{e^{i\pi/2}}{k^2} \right) \left( e^{i\pi x'} k \right) - \sum_{k=1}^{\infty} \left( \frac{e^{i\pi/2}}{k^2} \right) \left( e^{i\pi (x' + 2x_p')} k \right) \right\}
\]

\[
= \sum_{k=1}^{\infty} \frac{\cos\left( \frac{\pi}{2} + \pi x' k \right) - \cos\left( \frac{\pi}{2} + \pi (x' + 2x_p') k \right)}{k^2}
\]

\[
= \sum_{k=1}^{\infty} \frac{\sin(\pi x' k) - \sin(\pi x' k)}{k^2},
\]

and using the trigonometric identity \( \sin(A) - \sin(B) = 2 \cos\left( \frac{A+B}{2} \right) \sin\left( \frac{A-B}{2} \right) \), we can express Equation (2.26) as

\[
V(x') = -x_D (\lambda_{N_A}^- + \lambda_{N_D}^+) \left[ 2x_p' \ln(2) + \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{\cos(\pi (x' + x_p') k) \sin(\pi x_p' k)}{k^2} \right] + C,
\]

for \(-x_p' < x' < x_n'\). (2.29)

Hence, we arrive at a simplified expression for the electrostatic potential by setting the potential at \(x' = -x_p'\) equal to 0, i.e. \(V(-x_p') = 0\). Hence, Equation (2.29) can be expressed as

\[
V(x') = \frac{2x_D (\lambda_{N_A}^- + \lambda_{N_D}^+)}{\pi} \left[ \sum_{k=1}^{\infty} \frac{\sin(\pi x_p' k)}{k^2} - \sum_{k=1}^{\infty} \frac{\cos(\pi (x' + x_p') k) \sin(\pi x_p' k)}{k^2} \right],
\]

for \(-x_p' < x' < x_n'\). (2.30)

Finally, the electrostatic potential \(V(x)\) can be expressed in the non-scaled coordinate system by using Equation (2.30) and and the scaled version of Equation (2.11):

\[
V(x) = \frac{q(N_{A,2D}^- + N_{D,2D}^+)}{\pi^2 \epsilon_{eff}} x_D \left[ \sum_{k=1}^{\infty} \frac{\sin(\pi x_p' k)}{x_D k^2} - \sum_{k=1}^{\infty} \frac{\cos(\pi (x + x_p') k) \sin(\pi x_p' k)}{x_D k^2} \right],
\]

for \(-x_p < x < x_n\). (2.31)
2.1.3 Analytical formulation of the depletion width

Next, we derive an expression for the depletion width by imposing the condition that

\[ V(x'_n) - V(-x'_p) = \phi_{bi}. \]  \tag{2.32}

Since, we have chosen \( V(-x'_p) = 0 \) as the reference potential, \( V(x'_n) \) must be equal to \( \phi_{bi} \) to satisfy Equation (2.32). This yields

\[
\phi_{bi} = \frac{2x_D(\lambda_{N_A^-} + \lambda_{N_D^+})}{\pi} \left[ \sum_{k=1}^{\infty} \frac{\sin(\pi x'_p k)}{k^2} - \sum_{k=1}^{\infty} \frac{\cos(\pi x'_n + x'_p k) \sin(\pi x'_p k)}{k^2} \right]. \tag{2.33}
\]

Simplifying this expression using \( x'_n + x'_p = 1 \) and rearranging Equation (2.33) to give the total depletion width yields

\[
x_D = \frac{\pi \phi_{bi}}{2(\lambda_{N_A^-} + \lambda_{N_D^+}) \sum_{k=1}^{\infty} \frac{[1-(-1)^k] \sin(\pi x'_p k)}{k^2}}. \tag{2.34}
\]

Furthermore, by substituting \( \lambda_{N_A^-} = \frac{qN_{A,2D}^-}{2\pi\epsilon_{eff}} \), \( \lambda_{N_D^+} = \frac{qN_{D,2D}^+}{2\pi\epsilon_{eff}} \), and \( x'_p = \frac{N_{D,2D}^+}{N_{A,2D}^- + N_{D,2D}^+} \) into Equation (2.34) and generalizing the 2D depletion width expression to include the impact of external potential applied across the junction \((V_a)\) leads to

\[
x_D = \frac{\pi^2\epsilon_{eff}(\phi_{bi} - V_a)}{qN_{D,2D}^+ f \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right)} = \frac{\pi^2\epsilon_{eff}(\phi_{bi} - V_a)}{qN_{A,2D}^- f \left( \frac{N_{A,2D}^+}{N_{D,2D}^-} \right)}, \tag{2.35}
\]

where

\[
f \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) = \sum_{k=1}^{\infty} \frac{[1-(-1)^k](1 + \frac{N_{A,2D}^-}{N_{D,2D}^+}) \sin \left( \frac{\pi k}{1 + \frac{N_{A,2D}^-}{N_{D,2D}^+}} \right)}{k^2}. \tag{2.36a}
\]
\[ g \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) = \sum_{k=1}^{\infty} \frac{[1 - (-1)^k](1 + \frac{N_{A,2D}^+}{N_{D,2D}^+}) \sin \left( \frac{\pi k}{1 + \frac{N_{A,2D}^+}{N_{D,2D}^+}} \right)}{k^2}, \]  

(2.36b)

and \( V_a \) is defined to be positive when the 2D lateral p-n junction is forward biased.

Figure 2.4 shows \( f \) and \( g \) as functions of the doping ratio \( \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) \). For strongly asymmetric junctions, a logarithmic dependence on the doping ratio is observed:

\[ N_{A,2D}^- \gg N_{D,2D}^+ : f \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) \approx f_{fit} \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) = 7.12 \log_{10} \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) + 2.01 \]  

(2.37a)

\[ N_{D,2D}^+ \gg N_{A,2D}^- : g \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) \approx g_{fit} \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) = -7.12 \log_{10} \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) + 2.01, \]  

(2.37b)

Next, we discuss the impact of 2D vs 3D electrostatics on screening of charge carriers within the depletion region. Two-dimensional lateral p-n junctions with different doping configurations, both symmetric and asymmetric, are studied. Here, we use the full (non-approximate) expression for 2D depletion width given by Equation (2.35) to study the 2D lateral p-n junction electrostatics. For symmetric doping, Eq. (2.36a) can be simplified to

\[ f \left( \frac{N_{A,2D}^-}{N_{D,2D}^+} \right) = 4G, \]  

(2.38)

where \( G \) is the Catalan’s constant (\( G \approx 0.915 \)). Replacing \( f \) in Equation (2.35) provides a simplified expression for 2D depletion width for symmetric doping

\[ x_{D,2D,\text{sym}} = \frac{\pi^2 \epsilon_{\text{eff}} \left( \phi_{bi} - V_a \right)}{4GqN_{S,2D}}, \]  

(2.39)
Figure 2.4: Approxi\textbf{mations of }f \textbf{and }g \textbf{ functions for asymmetrically doped junctions.} Functions \( f \) and \( g \) from Equation (2.37a) and Equation (2.37b) plotted for different doping ratios \( \left( \frac{N_{\text{A},2D}}{N_{\text{D},2D}} \right) \). For strongly asymmetric junctions, a logarithmic dependence for either \( f \) or \( g \) is observed.

where \( N_{\text{S},2D} \) is the sheet charge density of the symmetric junction where \( N_{\text{S},2D} = N_{\text{A},2D}^- = N_{\text{D},2D}^+ \). We then calculate the 3D depletion width using the expression

\[
x_{D|3D,\text{sym}} = \sqrt{\frac{4e_{\text{eff}}(\phi_{\text{bi}} - V_a)}{qN_{\text{S},3D}}},
\]

where \( qN_{\text{S},3D} \) is the doping concentration of the symmetric junction (in C/volume) [53].
We can express Equation (2.40) in terms of $N_{S,2D}$ as

$$x_{D\mid 3D,sym} = \sqrt{\frac{4\varepsilon_{eff}(\phi_{bi} - V_a)T_{sheet}}{qN_{S,2D}}}$$ (2.41)

through the relationship that $N_{S,2D} = N_{S,3D}T_{sheet}$ assuming a uniformly doped semiconductor.

Figure 2.5(a) compares the 2D and 3D depletion widths calculated using Equation (2.39) and Equation (2.40), respectively. For this analysis, we have assumed $\varepsilon_{eff} = 2.45\varepsilon_0$, which reflects the case where the 2D lateral p-n junction has SiO$_2$ ($\varepsilon_1 = 3.9\varepsilon_0$) and air ($\varepsilon_2 = \varepsilon_0$) as its surrounding dielectric media. Further, we consider $T_{sheet} = 0.7$ nm, which is the approximate thickness for monolayer TMDCs. For symmetric junctions, the 2D depletion width is proportional to $\frac{1}{N_{S,2D}}$, whereas the 3D-depletion width is proportional to $\frac{1}{\sqrt{N_{S,2D}}}$. At low sheet charge densities, the 2D depletion width can be significantly larger than that calculated from the conventional 3D method. The longer depletion width for 2D lateral p-n junctions can be attributed to the weaker screening of in-plane charge carriers due to the large out-of-plane electric field [54]. We further analyzed the 2D and 3D depletion widths for asymmetric doping configurations by varying the sheet charge density on the p-side while keeping constant n-side doping ($N_{D,2D}^+ = 10^{11}$ cm$^{-2}$). We calculate the 2D depletion width from Equation (2.35) and the 3D depletion width from the conventional expression given by

$$x_{D\mid 3D,asy} = \sqrt{\frac{2\varepsilon_{eff}(\phi_{bi} - V_a)T_{sheet}}{q}} \left[ \frac{1}{N_{A,2D}^-} + \frac{1}{N_{D,2D}^+} \right].$$ (2.42)

As shown in Figure 2.5(b), the depletion width dependence of asymmetric junctions varies for 2D and 3D geometries. For a $p^+-n$ junction, i.e., $N_{A,2D}^- \gg N_{D,2D}^+$, the 3D depletion width becomes independent of doping density of the highly doped $p$-side. The depletion width for a 3D $p^+-n$ junction can be obtained by taking the limit of Equa-
Figure 2.5: Comparison of 2D vs 3D depletion width as a function of sheet charge density. Depletion width calculated for (a) symmetric and (b) asymmetric doping configurations calculated using our 2D method and the conventional 3D method. For symmetric doping configuration, the 2D depletion width is proportional to \( \frac{1}{N_{S,2D}} \) whereas the 3D depletion width is proportional to \( \frac{1}{\sqrt{N_{S,2D}}} \). For strongly asymmetric junctions, the 3D depletion width is determined solely by the lowly doped side (e.g. it is independent of \( N_{A,2D}^- \) for \( N_{A,2D}^- \gg N_{D,2D}^+ \)), whereas the 2D depletion width shows a weak (logarithmic) dependence on the doping density of the highly doped side.
tion (2.42):

$$3D \text{ p}^+\text{-n junction: } \lim_{N_{A2D} \gg N_{D2D}^+} x_D |_{3D,\text{asym}} \propto \sqrt{\frac{1}{N_{D2D}^+}} \left( \frac{N_{A2D}}{N_{D2D}^+} \right)^{\frac{1}{2}} \right)$$

(2.43)

Hence, for asymmetrically doped 3D junctions, the 3D depletion width is completely determined by the doping density of the lowly doped side. However, as shown in Figure 2.5(b), the contribution of the highly doped side cannot be neglected for 2D junctions. Following a similar analysis as before, we take the limit of Equation (2.35) (using Equation (2.37a)), which leads to

$$2D \text{ p}^+\text{-n junction: } \lim_{N_{A2D} \gg N_{D2D}^+} x_D |_{2D,\text{asym}} \propto \frac{1}{N_{D2D}^+ \log_{10} \left( \frac{N_{A2D}}{N_{D2D}^+} \right)}$$

(2.44)

This equation shows the logarithmic dependence of the 2D depletion width on the doping density of the highly doped side.

2.1.4 Extending the analytical model to lateral metal-semiconductor junctions

In this section, we extended our model to understand the electrostatics of 2D metal-semiconductor junctions given their critical role in 2D transistors. Estimating depletion width at metal-semiconductor junctions is also crucial for optoelectronic devices such as Schottky photodiodes, tunnel field-effect transistors, and solar cells [55, 56, 57]. For bulk semiconductors, the analytical expression for 3D depletion region can be easily modified to study metal-semiconductor junctions since the depletion region depends only upon the doping density of lowly-doped side. Since the 2D depletion width for asymmetric doping configurations still depends on the doping density of highly doped side, Equation (2.35) cannot be used to calculate the depletion width for 2D metal-2D semiconductor junctions by simply considering an infinite carrier density in the metal. An
Symmetrically doped 2D lateral p-n junction as mirror charge construction of 3D metal-2D semiconductor junction.

Figure 2.6: Comparison of electric field profiles for 2D lateral p-n and metal-semiconductor junction (a,b) Schematics of a symmetrically doped 2D lateral p-n junction and a 3D metal-2D semiconductor junction. (c) A symmetrically doped 2D lateral p-n junction is simply a mirror charge construct of a 3D metal-2D semiconductor junction. The depletion for a 3D metal-2D semiconductor junction is half of that of a symmetrically doped 2D lateral p-n junction having the same semiconductor doping density.

extension of Equation (2.35); however, can be derived for calculating depletion widths for 3D metal-2D semiconductor junctions by exploiting the similarities of symmetrically doped 2D lateral p-n junctions (Figure 2.6(a)) and 3D metal-2D semiconductor junctions (Figure 2.6(b)). As shown in Figure 2.6(c), a symmetrically doped 2D lateral p-n junction can be equally represented as a mirror charge arrangement of a 3D metal-2D semiconductor junction since both configurations yield the same $E_z|_{interface} = 0$ boundary condition. Since no depletion occurs in an ideal metal, the depletion width for a 3D
metal-2D semiconductor junction can be obtained by dividing Eq. (2.39) by 2 since only one side of the junction is depleted:

\[ x_D |_{3D_{metal-2D_{semi}}} = \frac{\pi^2 \epsilon_{e f f} (\phi_{bi} - V_a)}{8 G q N_{S,2D}}. \]  

(2.45)

The built-in potential \( \phi_{bi} \) used in Equation (2.45) corresponds to \( 2 \phi_{MS} \) (twice the built-in potential of the metal-semiconductor junction). If recast in terms of \( \phi_{MS} \), we can rewrite the expression as

\[ x_D |_{3D_{metal-2D_{semi}}} = \frac{\pi^2 \epsilon_{e f f} (\phi_{MS} - V_a)}{4 G q N_{S,2D}}. \]  

(2.46)

Thus, the depletion width for metal-semiconductor junctions in 2D materials is strongly dependent on the sheet doping density on the semiconductor side.

Interestingly, the depletion width also strongly depends upon the surrounding dielectric media. The atomically thin nature of 2D materials allows tuning of the depletion width by changing the surrounding dielectrics because the out-of-plane electric field is highly affected by the surrounding media. Hence, the surrounding dielectric media can be engineered to reduce or increase the depletion width. A smaller depletion width is desired for low resistance contacts whereas a larger depletion width is desired for solar cells and photo-detectors. Figure 2.7 shows the dependence of the depletion width on the effective dielectric constant \( \epsilon_{e f f} \) for symmetric p and n-side doping densities. Here, we assume that the dielectric-semiconductor interface remains pristine for all dielectrics, i.e., there is negligible interface charge density compared to the intrinsic sheet charge densities of semiconductor. The impact of interface charge density can also be included in our analysis by adding the interface charge density to the intrinsic sheet charge density. Clearly, the impact of the surrounding dielectric environment must be considered in the design of 2D lateral p-n junctions.
Figure 2.7: **Impact of surrounding dielectric media on 2D depletion width.** The significant out-of-plane electric field of 2D lateral p-n junctions leads to a linear dependence of the depletion width on the relative dielectric permittivity of the surrounding dielectric media.

### 2.1.5 Verification of analytical model against experimental and simulation results

Finally, we discuss the extent of depletion tails in 2D lateral p-n junctions and their impact on the validity of the depletion approximation. In general, the boundaries of the depletion region are not abrupt due to the diffusive nature of the mobile charge on each side of the junction. Similar to 3D (bulk) junctions, the 2D depletion tails show similar dependencies on various structural/material parameters, such as the effective dielectric constant and sheet doping density since both the depletion width and depletion tails result from the screening behavior of charge carriers which only depends on the junction dimensionality [40]. To better understand the impact of depletion tails on the electrostatics of 2D lateral p-n junctions, we modeled a quasi-2D 2-nm-thick Si p-n junction with symmetric doping ($N_{D,2D}^+ = N_{A,2D}^- = 10^{12}$ cm$^{-2}$) and SiO$_2$ as the surrounding dielectric...
Figure 2.8: Verification of the analytical model against simulations. (a) Comparison of simulated and calculated space-charge distribution ($\rho$) along a 2D lateral p-n junction. The calculated space charge accurately models the box-like section of space charge profile. The simulated 2D lateral p-n junction contains long depletion tails which decay as $1/x$. (b) Electric field ($E_x$) and (c) electrostatic potential ($V$) profiles for the same structure in (a). Although the space-charge calculated from the analytical model and TCAD simulation show significant differences, the electric field and electrostatic potential profiles show little discrepancy.
media using *Sentaurus Device* TCAD. Figure 2.8 (a) compares the profiles of the simulated and calculated space charge ($\rho$) along the 2D lateral p-n junction. As expected, the box-like profile of the depletion charge is accurately predicted by our analytical method; however, the simulated profile of the space charge also contains a long depletion tail that decays as $\frac{1}{x}$ [36, 40]. Unlike 3D (bulk) junctions, these long depletion tails in 2D lateral p-n junctions contain a significant amount of the total depletion charge; however, as shown in Figure 2.8(b,c), the depletion tails do not significantly impact the electric field and electrostatic potential profiles, and our 2D analytical method provides a great fit to the simulated data, supporting the use of the depletion approximation in modeling 2D lateral p-n junctions.

To further validate our model, we compare the 2D depletion width calculated using our model with the experimental data for a 2D lateral $p-n$ junction ($T_{\text{sheet}} = 10$ nm) fabricated from III-V materials and measured using optical beam induced current (OBIC) measurements as shown in Figure 2.9(a) [38]. Our model provides a good match to the experimental data for the given experimental parameters ($N_{D,2D}^+ = 5 \times 10^{10}$ cm$^{-2}$, $N_{A,2D}^- = 4 \times 10^{11}$ cm$^{-2}$). Here, $\epsilon_{\text{eff}}$ is considered as a fitting parameter that varies between $7.05\epsilon_0$ (when the top dielectric is air and the bottom dielectric is a semiconductor) and $13.1\epsilon_0$ (when both top and bottom dielectrics are semiconductors). Compared to the 2D depletion width, the depletion width for conventional 3D method calculated using Equation (2.42) is off by an order of magnitude irrespective of $\epsilon_{\text{eff}}$. In Figure 2.9(b), we compare the depletion width calculated from our method and other 2D methods with the same experimental data from Figure 2.6(a). Since the other 2D methods are only formulated for symmetric doping configurations, we calculate the depletion width for these models by considering the doping density of only the lowly doped $n$-side — the doping of the highly doped $p$-side is ignored. Our method includes both $n$- and $p$-side doping and calculates the depletion width using Equation (2.35). The effective dielectric constant is taken to be $\epsilon_{\text{eff}} = 10.3\epsilon_0$, in accordance with the value extracted from
Figure 2.9: Verification of the analytical model against experimental data. (a) Experimental [38] and calculated depletion width for a 2D lateral p-n junction for the experimentally given range of $\varepsilon_{eff}$. Our 2D method provides excellent fits to the experimental data, whereas the depletion widths calculated using the conventional 3D method are off by an order of magnitude irrespective of $\varepsilon_{eff}$. (b) Comparison of different 2D methods against the same experimental data from part (a) [40, 41, 43, 45, 58]. Assuming $\varepsilon_{eff} = 10.3\varepsilon_0$, our 2D method provides best estimate to the experimental data.

Figure 2.9(a). This comparison shows the importance of the generalized 2D depletion width expression derived in this work that accounts for both $n$ and $p$-side doping since no other analytical 2D methods are applicable for asymmetric doping configurations.
Table 2.1: Analytical expressions for various electrostatic parameters for 2D lateral junctions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric Field</td>
<td>( E_x(x) = \frac{q(N_{D,2D}^+ + N_{A,2D}^-)}{2\pi \epsilon_{\text{eff}}} \ln \left</td>
</tr>
<tr>
<td>Electrostatic Potential</td>
<td>( V(x) = \frac{q(N_{D,2D}^+ + N_{A,2D}^-)}{\pi^2 \epsilon_{\text{eff}}} x_D \sum_{k=1}^{\infty} \left[ \sin \left( \frac{\pi x p k}{x_D} \right) \left( 1 - \cos \left( \frac{\pi (x + x_p) k}{x_D} \right) \right) / k^2 \right] )</td>
</tr>
<tr>
<td>Depletion Width</td>
<td>General case: ( x_D = \frac{\pi^2 \epsilon_{\text{eff}} (\phi_{bi} - V_a)}{q(N_{D,2D}^+ + N_{A,2D}^-) \sum_{k=1}^{\infty} [1 - (-1)^k] \sin \left( \frac{\pi x p k}{x_D} \right) / k^2} )</td>
</tr>
<tr>
<td></td>
<td>Symmetric doping: ( x_D</td>
</tr>
<tr>
<td></td>
<td>P⁺-n junction: ( x_D</td>
</tr>
<tr>
<td></td>
<td>N⁺-p junction: ( x_D</td>
</tr>
<tr>
<td></td>
<td>3D metal-2D semiconductor: ( x_D</td>
</tr>
</tbody>
</table>

2.1.6 Summary

To summarize, we developed analytical expressions to calculate the electric field, electrostatic potential, and depletion width for symmetric and asymmetric doping configurations for 2D lateral p-n junctions. Table 2.2 provides the key analytical expressions derived for various electrostatic parameters of 2D lateral p-n junctions. We used the method of image charges to achieve zero electric field at the edges of the depletion region. Compared to bulk semiconductors, 2D lateral p-n junctions demonstrate weaker
screening of charge carriers resulting in larger depletion widths compared to 3D junctions of similar doping densities. Further, we validated our method by comparing the experimental and calculated depletion widths for a 2D lateral p-n junction where our method provides best match to the experimental data when compared to conventional 3D method and other 2D methods. We also investigated the impact of the surrounding dielectric media and sheet doping density on the electrostatics of 2D lateral metal-semiconductor junctions. Thus, increasing the doping density on the semiconductor side along with the use of lower dielectric media is essential for making Ohmic contacts with low contact resistance.

2.2 Role of out-of-plane dielectric thickness in electrostatic simulation of 2D devices

In this section, we focus on some subtle aspects of electrostatic simulations of 2D devices which is often ignored in 2D device community. Accurate modeling and simulation of 2D lateral p-n junctions are needed to complement the ongoing experimental studies in order to develop high-performance 2D-material devices [59, 60, 61]. Consider the structure shown in Figure 2.10(a), in which a thin semiconductor layer is surrounded by a thick dielectric. Given that there are no gates in the structure, it is tempting to assume that the dielectric thickness does not matter; however, this reasoning is incorrect. Unlike bulk (3D) p-n junctions, the presence of a large surface-to-volume ratio in the thin (2D) lateral junctions leads to a significant out-of-plane electric field, indicated by the electric field lines in Figure 2.10(a) [62]. The impact of these out-of-plane electric fields have been discussed in the previous section.

Here, a practical issue for accurate modeling of 2D lateral junctions is determining an appropriate position of the simulation boundaries in the out-of-plane direction, especially when the desired dielectric thickness is exceedingly large. The simulated dielectric thickness ($t_{ox}$) surrounding a 2D lateral junction significantly affects the simulation results by modifying the out-of-plane electric field due to the boundary conditions im-
Figure 2.10: **Effect of varying dielectric thickness on 2D depletion width.** (a) A thin (2D) lateral p-n junction surrounded by a dielectric with a simulated thickness of $t_{ox}$. A significant out-of-plane electric field exists in the surrounding dielectric, unlike in a 3D p-n junction. (b) In-plane energy-band diagram of a 2D lateral p-n junction with symmetric doping. The significant variation of the energy band diagram with $t_{ox}$ highlights the impact of the out-of-plane simulation dimensions on the in-plane junction electrostatics.

posed on the edges of the simulation region. Unfortunately, a large simulated $t_{ox}$ leads to a large number of grid points that increase the computational burden and complicate the grid-meshing process. Conversely, a small simulated $t_{ox}$ affects the results due to the imposed boundary conditions. Though the effect of the simulated dielectric thickness on simulation results has been previously recognized [63, 64], no clear guidelines exist
to determine the extent to which $t_{ox}$ affects the simulated junction properties including the electric field, potential, and depletion width. In this work, we derive the relationship between the simulated $t_{ox}$ and depletion width for symmetrically-doped 2D lateral p-n junctions and provide guidelines for a minimal value of the simulated $t_{ox}$, above which the impact of the out-of-plane simulation domain boundaries on the in-plane junction electrostatics is minimal.

2.2.1 Effect of boundary conditions

To highlight the impact of the out-of-plane simulation dimensions on the electrostatics of 2D lateral p-n junctions, we simulated a symmetrically-doped 2D lateral p-n junction (shown in Figure 2.10(a)) using nextnano3, a research-focused quantum-mechanical device simulation tool [65]. The use of quantum models in the semiconductor region minimally affects the simulation results due to the absence of carriers in the depletion region, which dominates the electrostatics properties of the junction. We chose to simulate a generic 2D semiconductor (with material properties provided in Table 2.2) since our focus is the simulation methodology and not a specific semiconductor material. Neumann boundary conditions were imposed on the simulation domain boundaries. Zero-biased ohmic contacts were placed at the left and right edges of the semiconductor sheet. Figure 2.10(b) illustrates the impact of the simulated dielectric thickness ($t_{ox}$) on the in-plane energy-band diagram of the 2D lateral p-n junction. A small simulated $t_{ox}$ leads to significant changes in the in-plane junction electrostatics and the erroneous calculation of various electrostatic parameters, including the electric field, electrostatic potential, and depletion width.

Figure 2.11 provides further insight into the simulation results of 2D lateral p-n junctions. Commercial TCAD device simulators, including nextnano3, Sentaurus Device, and Silvaco Atlas, generally apply Neumann boundary conditions at the boundaries of the simulation region unless otherwise specified. Neumann boundary conditions for the
Table 2.2: Material parameters of the simulated generic 2D semiconductor.

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>band gap (with quantization), $E_g$</td>
<td>2 eV</td>
</tr>
<tr>
<td>electron effective mass, $m_e$</td>
<td>$0.5m_0$</td>
</tr>
<tr>
<td>hole effective mass, $m_h$</td>
<td>$0.5m_0$</td>
</tr>
<tr>
<td>dielectric constant, $\epsilon_s$</td>
<td>$5\epsilon_0$</td>
</tr>
<tr>
<td>sheet charge density, $N_{S,2D}$</td>
<td>$10^{11}$ cm$^{-2}$</td>
</tr>
<tr>
<td>sheet thickness, $T_{sheet}$</td>
<td>1 nm</td>
</tr>
</tbody>
</table>

Figure 2.11: Undesired reflections caused by Neumann boundary conditions. (a) A simulated 2D lateral p-n junction with Neumann boundary conditions on the out-of-plane edges of the surrounding dielectric media. The Neumann boundary conditions mirror the structure such that the simulated structure is equivalent to the structure shown in (c) as opposed to (b). A small simulated $t_{ox}$ yields closely stacked reflections that significantly affect the simulation results. As the simulated $t_{ox}$ becomes large, the reflections becoming increasingly far apart, minimally impacting the simulation results.
electrostatic potential enforce $\partial V / \partial \hat{n} = 0$, where $\hat{n}$ is the direction normal to the boundary. Neumann boundary conditions at the outer surfaces of the surrounding dielectric media reflect the problem across the boundary. Just as two mirrors facing each other create an infinite number of reflections, Neumann boundary conditions imposed on the top and bottom surfaces of the structure in Fig. 2.11(a) result in infinite reflections. These reflections transform the simulated problem from a single-layer structure to a multi-layer structure depicted in Fig. 2.11(c), though the intent may have been to simulate the structure shown in Fig. 2.11(b). A small simulated $t_{ox}$ significantly alters the junction electrostatics due to the closely-packed reflections. As $t_{ox} \to 0$, the junction electrostatics become that of a 3D ‘bulk’ junction.

2.2.2 Estimating minimum acceptable simulation window

Next, we determine a minimal simulated dielectric thickness ($t_{ox, \text{min}}$) above which the simulation results are minimally affected. To derive $t_{ox, \text{min}}$, we first consider a 2D lateral p-n junction with symmetric doping and an infinite dielectric thickness above and below the semiconductor layer. The 2D lateral p-n junction is assumed to be sufficiently thin and wide such that it can be treated as an ideal 2D plane. For such a structure, we have previously derived analytical expressions for the in-plane electric field, electrostatic potential, and depletion width, given in Section 2.1.

Imagine a Gaussian box that encloses the right-half of the structure, as shown in Figure 2.12. Gauss’s law states that the charge enclosed in the box is proportional to the total electric flux ($\Phi_{\text{total}}$) emanating from the box. For the case shown in Figure 2.12, this gives

$$Q_{D}|_{\text{n-side}} = \epsilon_{ox} \Phi_{\text{total}},$$

where

$$\Phi_{\text{total}} = \Phi_{\text{left}} + \Phi_{\text{right}} + \Phi_{\text{top}} + \Phi_{\text{bottom}},$$
Figure 2.12: Calculation of minimal out-of-plane dielectric thickness for accurate electrostatic simulations. A Gaussian box enclosing the right half (n-side) of a symmetrically-doped 2D lateral p-n junction surrounded by an infinitely thick dielectric. The minimal dielectric thickness \( t_{\text{ox, min}} \) is derived by determining the Gaussian box thickness \( t_{\text{box}} \) for which the flux passing through the top and bottom surfaces \( \Phi_{\text{top}} \) and \( \Phi_{\text{bottom}} \) is some small fraction \( \alpha \) of the total flux. For a 2D lateral p-n junction with a fairly long quasi-neutral region (compared to its depletion region), the electric flux emanating from the right surface is approximately zero \( \Phi_{\text{right}} \approx 0 \).

\( Q_D|_{\text{n-side}} \) is the depletion charge on the n-side of the junction, and \( \varepsilon_{\text{ox}} \) is the permittivity of the dielectric. For a structure that is much longer than its depletion width, the electric field at the right surface of the box is approximately zero, and, therefore, \( \Phi_{\text{right}} \approx 0 \). When the top and bottom surfaces of the box are placed at \( z = +\infty \) and \( -\infty \), respectively, the electric field and flux at these surfaces are also zero. As the top and bottom surfaces of the Gaussian box are brought closer to the semiconductor layer, the electric flux through these surfaces is no longer exactly zero, as seen in Figure 2.12.

If the top and bottom surfaces of the Gaussian box are imagined to be boundaries for a simulation structure with a finite dielectric thickness, Neumann boundary conditions demand that \( \partial V / \partial \hat{n} = 0 \), and, therefore, \( \Phi_{\text{top}} = \Phi_{\text{bottom}} = 0 \). Clearly, these boundary conditions conflict with the analytical solution for a structure with infinite dielectric thickness since the electric field is zero only infinitely far away from the junction. But as long as the simulated dielectric is thick enough (i.e., \( t_{\text{ox}} > t_{\text{ox, min}} \)), the electric flux through the top and bottom surfaces will be negligible compared to the total electric flux.
flux through the Gaussian box, and the electrostatics will be minimally affected by the imposed boundary conditions.

The methodology for determining $t_{ox,\text{min}}$ is outlined as follows:

1. Neumann boundary conditions for the top and bottom surfaces are nearly satisfied if

   $$\Phi_{\text{top}} + \Phi_{\text{bottom}} \approx 0. \quad (2.49)$$

2. We quantify Equation (2.49) by demanding that the electric flux passing through the top and bottom surfaces is less than or equal to a small fraction ($\alpha$) of the total electric flux emanating from the Gaussian box shown in Figure 2.12:

   $$\Phi_{\text{top}} + \Phi_{\text{bottom}} \leq \alpha \cdot \Phi_{\text{total}}. \quad (2.50)$$

   Using Equation (2.48), we rewrite Equation (2.50) as

   $$\Phi_{\text{total}} - \Phi_{\text{left}} \leq \alpha \cdot \Phi_{\text{total}}, \quad (2.51)$$

   under the assumption that $\Phi_{\text{right}} = 0$ for a long device, as shown in Figure 2.12. Using Equation (2.47), this expression can be written as

   $$\Phi_{\text{left}} \geq (1 - \alpha) \frac{Q_{D|n\text{-side}}}{\varepsilon_{ox}}. \quad (2.52)$$

3. The electric flux emanating through the left surface of a Gaussian box with a height of $t_{\text{box}}$ above and below the semiconductor is defined as

   $$\Phi_{\text{left}} \equiv \int_{-t_{\text{box}}}^{t_{\text{box}}} -E_x(z) \, dz, \quad (2.53)$$

   where $E_x(z)$ is the $x$-directed electric field as a function of vertical position at the
Figure 2.13: Variation in simulated 2D depletion width with out-of-plane dielectric thickness. Impact of the out-of-plane dielectric thickness ($t_{ox}$) on the simulated depletion width of a symmetrically-doped p-n junction, shown in the inset. The simulated depletion width varies significantly with $t_{ox}$, approaching the theoretical bulk (3D) junction depletion width as $t_{ox} \to 0$ and the theoretical 2D depletion width as $t_{ox} \to \infty$. The simulated 2D depletion width is nearly equal to the theoretical 2D depletion for $t_{ox} > t_{ox,\text{min}}$ as indicated by the shaded region of the plot.
left surface (which coincides with the plane containing junction interface) for a structure with infinite dielectric thickness (see Figure 2.12). This allows Equation (2.52) to be rewritten as

$$\int_{-t_{\text{box}}}^{t_{\text{box}}} -E_x(z)dz \geq (1 - \alpha) \frac{Q_{D|n-\text{side}}}{\varepsilon_{\text{ox}}},$$  \hspace{1cm} (2.54)

where

$$Q_{D|n-\text{side}} = q N_{S,2D} x_{D,\inf}^2. \hspace{1cm} (2.55)$$

Here, $N_{S,2D}$ is the sheet charge density, and $x_{D,\inf}$ is the theoretical depletion width for a symmetrically-doped 2D lateral p-n junction with an infinite dielectric thickness above and below the semiconductor layer, derived in the previous section as

$$x_{D,\inf} = \frac{\pi^2 \varepsilon_{\text{ox}} \phi_{bi}}{4GqN_{S,2D}}, \hspace{1cm} (2.56)$$

where $\phi_{bi}$ is the built-in potential, and $G$ is Catalan’s constant ($G \approx 0.916$).

An analytical expression for $E_x(z)$, derived in the Appendix 1, yields

$$E_x(z) = \frac{-q N_{S,2D}}{\pi \varepsilon_{\text{ox}}} \ln \left[ \coth \left( \frac{\pi z}{2x_{D,\inf}} \right) \right]. \hspace{1cm} (2.57)$$

4. Finally, $t_{\text{ox,min}}$ is numerically calculated (in terms of $x_{D,\inf}$) by finding the value of $t_{\text{box}}$ for which the left and right sides of Eq. (2.54) are equal:

$$\int_{-t_{\text{box}}}^{t_{\text{box}}} -\frac{1}{\pi} \ln \left[ \coth \left( \frac{\pi z}{2x_{D,\inf}} \right) \right] dz = (1 - \alpha) \frac{x_{D,\inf}}{2}. \hspace{1cm} (2.58)$$

Values for $t_{\text{ox,min}}$ and the normalized simulation time for different $\alpha$ are provided in Table 2.3.

To validate our approach of calculating $t_{\text{ox,min}}$, we extract the simulated depletion
Table 2.3: Minimal out-of-plane dielectric thickness \( t_{ox,min} \) and normalized simulation time as a function of \( \alpha \), the fraction of the total electric flux that emanates through the top and bottom surfaces of the Gaussian box shown in Fig. 2.12.

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>( t_{ox,min} )</th>
<th>Normalized simulation time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.67( x_{D,inf} )</td>
<td>1.0( \times ) (reference case)</td>
</tr>
<tr>
<td>0.05</td>
<td>0.89( x_{D,inf} )</td>
<td>1.5( \times )</td>
</tr>
<tr>
<td>0.01</td>
<td>1.40( x_{D,inf} )</td>
<td>2.5( \times )</td>
</tr>
<tr>
<td>0.001</td>
<td>2.13( x_{D,inf} )</td>
<td>4.5( \times )</td>
</tr>
<tr>
<td>0.0001</td>
<td>2.86( x_{D,inf} )</td>
<td>7.0( \times )</td>
</tr>
</tbody>
</table>

width associated with different simulated \( t_{ox} \) for the 2D lateral p-n junction simulated in Figure 2.10(a). The presence of long depletion tails in 2D lateral p-n junctions makes it difficult to extract sharp depletion region edges from the space-charge plots [66]. Hence, we extracted the 2D depletion width as the distance over which 90% or more of the majority carriers are depleted from their nominal value. Figure 2.13 clearly illustrates the significant modulation of the simulated depletion width as \( t_{ox} \) is varied. The simulated depletion width approaches the theoretical 3D depletion width as \( t_{ox} \rightarrow 0 \) and the theoretical 2D depletion width as \( t_{ox} \rightarrow \infty \). For \( t_{ox} < t_{ox,min} \), the depletion width is strongly affected by the dielectric thickness due to the significant modulation of the out-of-plane electric field by the Neumann boundary conditions imposed by the simulator. But for \( t_{ox} > t_{ox,min} \) (shaded in gray), the depletion width is nearly equal to the theoretical 2D depletion width for a structure with infinite dielectric thickness, validating our approach.

2.2.3 Summary

In conclusion, we showed that Neumann boundary conditions imposed at the edges of the out-of-plane simulation region significantly affect the simulated out-of-plane electric field, electrostatic potential, and depletion width of atomically thin 2D lateral junctions. The Neumann boundary conditions reflect the problem across the boundary, thereby changing a single-layer device simulation to a multi-layer one. As the sim-
ulated dielectric thickness becomes smaller ($t_{ox} \ll t_{ox,\text{min}}$), the junction electrostatics become significantly affected due to the closely-packed reflections. We developed an analytical method to calculate a minimal dielectric thickness ($t_{ox,\text{min}}$) for simulating a symmetrically-doped 2D lateral p-n junction, above which the out-of-plane boundary conditions minimally affect the simulated junction electrostatics. Overall, the results presented in this work provide guidelines for the accurate simulation of 2D devices built with atomically thin junctions.
Chapter 3: Modeling of Current Flow in Back-Gated Field-Effect Transistors

Two-dimensional semiconducting materials such as transition metal dichalcogenides (including MoS$_2$ and WSe$_2$) have shown great promise as post-silicon channel materials for low-power transistor and optoelectronic applications due to their atomically-thin dimensions and lack of surface dangling bonds, yielding excellent gate electronics even for sub-10 nm channel length [13, 14]. To highlight the use of 2D semiconductors as channel material, long-channel back-gated field-effect transistors (BGFETs) are widely used due to the relative ease in fabrication and device analysis [15, 60]. The presence of long channel helps to avoid short-channel effects and thus long-channel BGFETs could be modeled as well-tempered devices. More importantly, long-channel BGFETs also enable easy estimation of several device parameters such as field-effect mobility ($\mu_{FE}$), carrier density ($n_{2D}$), Schottky barrier height ($\phi_{Bn}$), and contact resistance ($R_C$) [67, 68, 60, 69, 70, 71, 72].

Figure 3.1(a) shows the schematic of an n-type, monolayer (1L) TMDC-based long-channel BGFET consisting of top-contacted metallic source-drain junctions and global back-gate (covering both channel and contact regions). Large band gap in 2D semiconductors and Fermi level pinning at metal-semiconductor junctions leads to substantial electron ($\phi_{Bn}$)/hole ($\phi_{Bp}$) Schottky barriers (SBs) at the source-drain junctions as shown in Figure 3.1(b) [71, 29]. Unlike bulk semiconductors, these source-drain SBs result in the formation of Schottky contacts at the two junctions due to absence of efficient doping techniques leading to large contact resistance ($R_C$). These Schottky contacts often

---

1This chapter is based on [73] and reproduced with the permission of IOP Publishing.
compete with the channel region to determine the overall device current. Moreover, the presence of a global back-gate results in simultaneous gating of the channel and source-drain regions that further complicates analysis of experimental BGFETs [74].

3.1 Where is the current bottleneck in 2D transistors?

To examine their transport characteristics, BGFETs are often modeled as Schottky barrier-MOSFETs (SB-MOSFETs) where the device is treated as a fixed SB device at the source with simultaneous gating of the source and channel regions. The model is often used in experimental studies to extract important device parameters: (1) $\mu_{FE}$ from the ON state (beyond threshold) by modeling the device as a conventional long-channel MOSFET, and (2) SBH from the OFF state (below threshold) where the device is treated as a ballistic SBFET. *This demands an assumption that the current flow is limited by the source SB in OFF state whereas the channel is assumed to dominate the current flow in ON state* [75, 76]. In this chapter, we verify these assumptions for TMDC-based BGFET with metallic top/edge-contacts using transmission theory based analytical current models, load-line analysis of lumped-circuit model depicted in Figure 3.1(c), and self-consistent drift-diffusion simulations.

3.1.1 Analytical formalism for source and channel current in long-channel BGFET

The lumped-circuit model allows independent analysis of current flow in the source, channel, and drain regions [77]. Here, the source and drain junctions are modeled as ballistic SBFETs with source (drain) being the reverse-biased (RB) (forward-biased (FB)) and the channel region is modelled as a MOSFET. Our approach of calculating the current bottleneck in a long-channel BGFET is based on calculating the allowed device current ($I_D$) and applied potential drop across the source SB contact ($V_{SB}$) using analytical current models for the source SBFET and channel MOSFET. We then use the calculated currents to perform load-line (current continuity) analysis on the simplified lumped-circuit...
Figure 3.1: Modeling of long-channel BGFETs. (a) A typical long-channel n-type monolayer BGFET with global back gating of both channel and contact regions. (b) Energy band diagram along the channel ($L_{CH}$) highlighting the electron SB ($\phi_{Bn}$) at the source-drain contacts. The source SB is reverse-biased while the drain SB is forward-biased by the applied $V_{DS}$. (c) Lumped-circuit model for a long-channel BGFET with the source/drain contacts modeled as ballistic SBFETs and channel region modeled as a conventional MOSFET. Conventionally, the source SBFET is assumed to dominate in the OFF state and channel MOSFET is assumed to dominate in the ON state.
model, depicted in Figure 3.2(a), considering only the source SBFET and channel MOSFET. For any long-channel BGFET, the drain region can be ignored for large source-drain voltages ($qV_{DS} > \phi_{Bn}$ often true for experimental devices) since the forward-biased drain junction becomes highly conductive. Since we are calculating drop across the source contact, we can also consider the forward-biased drain junction as part of the channel MOSFET. $V_{SB}$ is determined by identifying the drop in electron quasi-Fermi level ($E_{Fn}$) at $x = 5\lambda$ from the source contact, relative to the source junction ($E_{Fn,S} = 0$), as shown in the associated band diagram. Here, $x = 5\lambda$ can be considered as the position of the top of the barrier for the channel region as the effects of the source SB fade away exponentially with a characteristic scaling length $\lambda$, given by

$$\frac{1}{\epsilon_{ox}} \tan \left( \frac{2t_{ox}}{\lambda} \right) + \frac{1}{\epsilon_{ch}} \tan \left( \frac{2t_{ch}}{\lambda} \right) = 0,$$

(3.1)

where, $\epsilon_{ox}$ & $\epsilon_{ch}$ are the dielectric constants, and $t_{ox}$ & $t_{ch}$ are the thicknesses of the gate oxide and channel, respectively. Note that we use this same scaling length to determine whether a BGFET is considered to be a long- or short-channel device.

For the load-line analysis, we require analytical models to determine source SB ($I_{SB}$) and channel ($I_{CH}$) current in different operating regimes determined by the applied gate voltage ($V_{GS}$). For an n-type long-channel BGFET, $I_{SB}$ per unit width can be calculated using

$$I_{SB} = \frac{qG_sG_v}{h^2} \sqrt{2\pi m^* k_B T} \int_{E_{C,x=5\lambda}}^{\infty} T_{SB}(E_x) \left( F_{-1/2} \left( \frac{E_{Fn,S} - E_x}{k_B T} \right) - F_{-1/2} \left( \frac{E_{Fn,x=5\lambda} - E_x}{k_B T} \right) \right) \, dE_x,$$

(3.2)

where $F_{-1/2}$ denotes the Fermi-Dirac integral of order $-1/2$, $T_{SB}(E_x)$ is the $x$-directed transmission coefficient in the transport direction across the source SB, $E_x$ is the $x$-directed longitudinal energy component, $E_{C,x=5\lambda}$ is the conduction band edge at $x = 5\lambda$, and
Figure 3.2: $V_{GS}$-dependent current components in long-channel BGFET. (a) Simplified model for long-channel BGFET [given in Fig. 3.1(a)] as the drain barrier is forward-biased for $V_{DS} > \Phi_{Bn}$. $I_{SB}$ denotes the current allowed by the source SBFET and $I_{CH}$ denotes the current allowed by the channel MOSFET. The band diagram relates $V_{SB}$ to the electron quasi-Fermi level at $x = 5\lambda$. (b) Different components of $I_{SB}$ and their relative magnitude as a function of $V_{GS}$. In the OFF state, $I_{SB}$ is predominantly thermionic in nature with pure thermionic emission for $V_{GS} < V_{FBSB}$, i.e., $\psi_S < 0$. The tunneling component increases beyond $V_{GS} > V_{FBSB}$ and becomes dominant in the ON state ($V_{GS} > V_{TH}$). (c,d) Typical $I_{SB}$ (red) and $I_{CH}$ (blue) for a long-channel BGFET as a function of $V_{SB}$ for constant $\psi_S$. 

\[ \begin{align*}
\text{OFF state:} & \quad \psi_S < 0 \\
& \quad (V_{GS} < V_{FBSB}) \\
\text{ON state:} & \quad 0 < \psi_S < \psi_S(V_{TH}) \\
& \quad (V_{FBSB} < V_{GS} < V_{TH}) \\
& \quad \psi_S \geq \psi_S(V_{TH}) \\
& \quad (V_{GS} \geq V_{TH})
\end{align*} \]
$E_{F_{n,S}}$ is the electron quasi-Fermi level at the source terminal, and $q$, $g_s$, $g_v$, $h$, $k_B$, $m_e^*$, and $T$ are the elementary electron charge, spin-degeneracy factor, valley-degeneracy factor, Planck’s constant, Boltzmann’s constant, electron effective mass, and temperature, respectively. Note that, Equation (3.2) is valid for a variety of contact architectures such as edge contacts and top contacts with low work function contacts where the semiconductor underneath the metal is completely hybridized due to metal-induced gap states [78, 79, 80]. The detailed derivation of Equation (3.2) is presented in Appendix B. $E_{C,x=5\lambda}$ can be analytically calculated using the expression

$$E_{C,x} = q\psi_S(V_{GS}) - q(\psi_S(V_{GS}) - \phi_{Bn}) \exp\left(-\frac{x}{\lambda}\right),$$  \hspace{1cm} (3.3)

where, $\psi_s$ can be expressed as

$$\psi_S = \frac{V_{GS} - V_{FBSB}}{m(V_{GS})}; \quad m(V_{GS}) = \left(1 + \frac{C_S(V_{GS}) + C_{IT}(V_{GS})}{C_{ox}}\right).$$  \hspace{1cm} (3.4)

Here, $\psi_S(V_{GS})$ can be interpreted as the surface potential at $x = 5\lambda$ (as shown in the band diagram provided in Figure 3.2(a)), $V_{FBSB}$ is the source-body flatband voltage that refers to the gate voltage at which $E_{C,x} = q\phi_{Bn}$ (also ambiguously referred to as $V_{FB}$ in some literature), $m(V_{GS})$ denotes the inverse gate-efficiency or band-movement factor, $C_S$ is the semiconductor capacitance, and $C_{IT}$ is the gate-voltage dependent interface trap capacitance.

In bulk semiconductors, $C_S$ consists of both depletion ($C_{S,D}$) and quantum capacitance ($C_{S,Q}$). Even though $C_{S,D}$ is nearly zero in 2D materials due to their atomically thin dimensions, $C_{S,Q}$ can play a significant role. $C_{S,Q}$ is negligible in the subthreshold region since there are very few carriers in the channel however, it increases exponentially with $V_{GS}$ before saturating to its degenerate limit $C_{S,Q} = q^2DOS$, where $DOS$ is the 2D band-edge density of states [67]. $C_{IT}$ reflects the effect of interface traps/defects on gate tunability of semiconductor bands where the presence of finite interface trap density
\((N_{IT})\) leads to large \(m\) and therefore slower band movement, especially in the subthreshold regime where \(C_{IT}\) can be directly expressed in terms of \(N_{IT}\) as \(C_{IT} = qN_{IT}\), where \(N_{IT}\) is given in \(\text{cm}^{-2}\). Note that in practical 2D devices, several factors such as layer thickness, substrate material, trap distribution, and fabrication methodology can significantly affect \(N_{IT}\) and thus expressing the gate voltage modulation directly in terms of \(\psi_S\) helps to analyze a wide range of back-gated 2D devices without determining the actual band-movement factor.

Subsequently, \(I_{SB}\) can be calculated by determining the analytical form of \(T_{SB}(E_x)\) which depends upon the type of current transport at the SB contact. Figure 3.2(b) shows the different current components that dominate the current transport at the source contact in different operating regimes (determined by source-to-channel band alignment). For \(V_{GS} \leq V_{FBSB}\), pure thermionic current, i.e. current above the SB, dominates the current flow and is purely determined by the source SBH and operating temperature. For thermionic transport, \(T_{SB}(E_x) = 1\) since the electron injected from the source contact does not encounter any barrier. For \(V_{FBSB} \leq V_{GS} \leq V_{TH}\) (\(V_{TH}\) denotes the threshold voltage), additional tunneling current can flow through the SB where \(T_{SB}(E_x)\) can be calculated using the WKB approximation for a triangular barrier which can be given as

\[
T_{SB}(E_x) = \exp \left( - \int_{x=0}^{x=3\lambda} \kappa(E_x) \, dx \right),
\]

where, \(\kappa\) is the electron wave-vector in the \(x\)-direction defined as

\[
\kappa(E_x) = \frac{1}{\hbar} \sqrt{2m^*_e (E_x - E_C(x))}.
\]

Note that the approximation of \(T_{SB}\) using triangular barrier actually underestimates the actual tunneling current (which occurs through an exponentially-decaying barrier), so we use \(x = 3\lambda\) instead of \(x = 5\lambda\) to estimate the tunneling current. Finally in the ON state, i.e. for \(V_{GS} \geq V_{TH}\), the tunneling current completely dominates the current flow.
at the source contact. Figure 3.2(c) shows typical $I_{SB}$ for a long-channel BGFET as a function of $V_{SB}$ at a constant $\psi_S$ considering both thermionic and tunneling components.

Further, we calculate $I_{CH}$ per unit width (shown in Figure 3.2(d)) by modifying Equation (3.2) to model a long channel 2D MOSFET following the seminal work from Datta et al. [81],

$$I_{CH} = \frac{q^2 g_s g_v}{\hbar^2} \sqrt{2\pi m^*_e k_B T} \int_{E_{C,x}=5\lambda}^{\infty} \left[ T_{CH}(E_x) \left( F_{-1/2} \left( \frac{E_{Fn,x}=5\lambda - E_x}{k_B T} \right) - F_{-1/2} \left( \frac{E_{Fn,D} - E_x}{k_B T} \right) \right) \right] dE_x,$$

(3.7)

where, $E_{Fn,D}$ is the electron quasi-Fermi level at the drain terminal and $T_{CH}(E_x)$ is the transmission coefficient for the channel region which can be expressed as $T_{CH}(E_x) = \frac{\lambda_{mfp}(E_x)}{\lambda_{mfp}(E_x) + L_{CH,eff}}$. Here, $\lambda_{mfp}(E_x)$ denotes the energy-dependent mean free path for back-scattering events in the channel and $L_{CH,eff}$ is the effective channel length over which the conduction band is nearly flat, i.e. in near-equilibrium with no significant electric field.

For a long-channel BGFET, $L_{CH,eff}$ can be approximated to $L_{CH}$. Although $\lambda_{mfp}(E_x)$ is predominantly used in compact modeling, we used a simplified form for our analysis by assuming that the mean-free path is independent of energy, i.e., $\lambda_{mfp}(E_x) \approx \lambda_{mfp}$ [82]. This approximation allows us to relate the mean-free path to the effective channel mobility ($\mu_{eff}$)—the mobility in band transport regime— using the relation $\lambda_{mfp} = 2 \frac{k_B T}{q} \frac{\mu_{eff}}{v_{ther}}$, where $v_{ther}$ is the $x$-directed thermal velocity which can be given as $v_{ther} = \sqrt{\frac{2k_B T}{\pi m^*_e}}$ [83].

Previously, Ma et al. studied the provided an analytical formulation of $\mu_{eff}$ at room temperature given by [84]

$$\mu_{eff} = \frac{3500}{N_{IT}/10^{11} \text{ cm}^{-2}} \left[ A(\varepsilon_e) + \left( \frac{n_s}{10^{13} \text{ cm}^{-2}} \right)^{1.2} \right] \text{cm}^2/\text{V} \cdot \text{s} .$$

(3.8)

Here, $A(\varepsilon_e)$ represents a fitting factor that depends on the dielectric surrounding and $n_s$
indicates the free electron density. Since the channel region dominates mainly in the OFF state, we simplify Eq. (3.8) to only consider the effect of \( N_{IT} \) (with no \( n_s \) dependence) and thus can be rewritten as

\[
\mu_{eff} = \frac{3500 A(\varepsilon_e)}{N_{IT}/10^{11} \text{ cm}^{-2}} \text{ cm}^2/V \cdot \text{s}.
\]

(3.9)

3.1.2 Load-line analysis for extracting current bottlenecks

To check our analytical model, we analyzed a monolayer MoS\(_2\)-based BGFET with back gating applied through a 3 nm SiO\(_2\) and verified our results by comparing the results with self-consistent drift-diffusion simulations. Although, monolayer channel was used here to illustrate the key issues, the model is no way limited to monolayer devices and can be easily extended to study multilayer channel as long as the multilayer channel can be analyzed using 2D density of states. In this regime, the multilayer channel can be modeled as a channel with increased effective mobility (\( \mu_{eff} \)) extracted for various experimental devices due to the reduction in substrate-induced carrier scattering [85, 86]. The presence of multilayer channel also results in an additional current at the source junction due to vertical injection. The source SB contact is characterized by an electron Schottky barrier height (\( \phi_{Bn} \)) of 0.3 eV and the channel region has a length (\( L_{CH} \)) of 100 nm. Additional material and device parameters are provided in Table 3.1. Using simplified form (Eq. 6), we obtain an effective mobility of \( \sim 100 \text{ cm}^2/V \cdot \text{s} \) which is typical of experimental MoS\(_2\)-based BGFETs [87]. Figure 3.3(a) shows the \( I_{SB} \) (solid lines) and \( I_{CH} \) (dotted lines) for different \( \psi_S \); thereby emulating the impact of gate biasing. The cross-over points (shown by the square dots) represent the load-line operating points at a given \( \psi_S \) indicating the actual device current (\( I_D \)) and applied potential drop across the source SB (\( V_{SB} \)).

Figure 3.3(b) shows the extracted device current (\( I_D \) — black) as a function of \( \psi_S \). For comparison, current calculated assuming ballistic source SBFET (\( I_{SB} \) — red) and ideal
Figure 3.3: **Load-line analysis for calculating current and applied potential drop across the source in a long-channel BGFET.** (a) Load-line diagram showing $I_{SB}$ (solid lines) and $I_{CH}$ (dashed lines) as a function of $V_{SB}$ for a given $\psi_s$. The cross-over points, highlighted by square dots, represent the allowed device current and applied potential drop across the source contact. (b) Allowed device current, $I_D$, calculated using the load-line analysis. A comparison with $I_{SB}$ and $I_{CH}$ confirms that the channel region dominates in the OFF state as $I_D \approx I_{CH}$ and contact dominates in the ON state with $I_D \approx I_{SB}$. The left inset shows band diagram in the subthreshold region and the right inset shows the band diagram in the ON state highlighting the presence of electric field in the channel region. (c) $V_{SB}$ as a function of $\psi_s$ showing channel dominated behavior at small $\psi_s$ (low applied gate voltage). (d) Comparison of $I_D$ calculated using the analytical method (solid lines) and extracted from self-consistent drift-diffusion simulation (solid dots) of the same MoS2 BGFETs. The analytically calculated current show good agreement to the simulated data.
Table 3.1: Model parameters for a monolayer MoS$_2$ BGFET with a 3 nm SiO$_2$ gate oxide [88].

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>electron effective mass, $m_e$</td>
<td>$0.45m_0$</td>
</tr>
<tr>
<td>channel thickness, $T_{\text{sheet}}$</td>
<td>0.65 nm</td>
</tr>
<tr>
<td>channel dielectric constant, $\epsilon_{ch}$</td>
<td>5.21$\epsilon_0$</td>
</tr>
<tr>
<td>gate oxide dielectric constant, $\epsilon_{ox}$</td>
<td>3.9$\epsilon_0$</td>
</tr>
<tr>
<td>interface trap density, $N_{IT}$</td>
<td>$1.25 \times 10^{11}$ cm$^{-2}$</td>
</tr>
<tr>
<td>fitting factor, $A(\epsilon_e)$</td>
<td>0.036</td>
</tr>
<tr>
<td>temperature, $T$</td>
<td>300 K</td>
</tr>
<tr>
<td>source-drain voltage, $V_{DS}$</td>
<td>1 V</td>
</tr>
</tbody>
</table>

channel MOSFET ($I_{CH}$ — blue) is also plotted. Thus, $I_D$ is limited by the channel region in the OFF state and limited by the source contact in the ON state. In the OFF state/subthreshold regime, an electron injected from the source terminal faces the same electron barrier (at $x = 5\lambda$) for both source and channel regions, as shown in the left inset. However, $T_{CH}$ is significantly smaller than $T_{SB}$ due to the presence of diffusive channel region with significant back-scattering events. *Simply put, for channel lengths longer than the mean-free-path of carriers, the back-scattering events reduce the carrier transmission through the channel region rendering channel region as the bottleneck for current flow.* Conversely, for devices with reasonable channel lengths, the source contact dominates in the ON state since the channel becomes more conductive due to addition of drift current caused by large free carrier density and the presence of significant electric field in the channel region, indicated by the downward slope of $E_C(x)$ in the right inset. This electric field increases $T_{CH}$ by increasing the inelastic scattering events in the channel region which effectively reduces the effective channel length [83, 89]. For an electron injected from the source contact, the electron quickly looses its energy closer to the source junction and thus have fewer chance of back-scattering. Note that the contact also becomes more conductive in the ON state due to additional tunneling component however, the increase in tunneling current is considerably slower than that of the increase in the channel current.

Figure 3.3(c) provides another insight into the device operation where we focus on
the applied potential drop across the source contact ($V_{SB}$) as a function of $\psi_S$. Since the source SB is reverse biased, the current across the source SBFET can also be expressed as $I_{SB} = I_{SB,\text{sat}}(\psi_S) \left(1 - \exp\left(-\frac{qV_{SB}}{k_BT}\right)\right)$. Here, $I_{SB,\text{sat}}(\psi_S)$ is the $V_{GS}$ dependent saturation current — the maximum current allowed by the reverse-biased source junction at a given $V_{GS}$. In presence of the channel region, $V_{SB}$ becomes a fraction of $V_{DS}$ that depends upon the channel length and gate-defined operating regime. Note that, the source-contact dominates the current flow when $qV_{SB} > 4k_BT$ as the current saturates with respect to $V_{SB}$ at this point. Here, this happens only for $\psi_S > 0.2$ V which is well into the ON state, thus supporting our previous analysis.

We further validated our load-line method with self-consistent drift-diffusion transport simulations performed using an in-house developed simulator which correctly captures the 2D density of states, electrostatics, and Schottky contacts to 2D materials [90]. The simulator considers both drift-diffusion and hopping transport in the channel region which allows accurate modeling of 2D devices, especially in the presence of significant trap states. Further details of the simulation tool and methodology are provided in Appendix C. The schematic of the simulated device is shown in Figure C.1 where the top dielectric environment is chosen to be large enough such that the simulation boundaries do not artificially impact the electrostatics around the active region [35, 91]. Figure C.2 depicts the total device current for $N_{IT}$ ranging from $10^{11}$ to $10^{13}$ cm$^{-2}$ which indicates that the hopping transport is dominant over band transport for $N_{IT} \geq 6 \times 10^{12}$ cm$^{-2}$. Even though earlier experimental works with unencapsulated and CVD MoS$_2$ illustrated hopping-dominant transport due to large $N_{IT}$ [92, 93, 94, 95], recent advancements in device fabrication techniques, better substrate dielectrics, and intrinsic channel material have resulted in 2D devices with $N_{IT} < 10^{12}$ cm$^{-2}$ and thus more accurately modeled using band transport [96, 97, 98, 99]. We also verify our simplification of Equation (3.8) by comparing analytically calculated $I_D$ (using effective mobility given by Equation (3.9)) with simulated $I_D$ for different trap densities, as shown in Figure C.3. The plot shows
good agreement between the analytical and simulation results thus validating our simplification. Finally, Figure 3.3(d) shows comparison of the drift-diffusion simulation results with analytically calculated $I_D$ using the load-line method. Our method provides excellent fits to the simulation data for any channel length. Moreover, the device current in the subthreshold region shows strong variation with channel length, thus confirming our hypothesis that it is the channel region which dominates the OFF state.

3.1.3 Summary

In this section, we analyzed 2D transistors using load-line analysis of lumped-circuit BGFET model and self-consistent drift-diffusion simulations. Contrary to previous works, we show that the channel region limits the overall device current in OFF state whereas the source-contact limits current in the ON state of device operation. This was attributed to reduced transmission through the channel region in the OFF state and near unity transmission in the ON state.

3.2 Understanding current transport in top-contacted 2D semiconductors

Having understood the role of contacts in 2D transistors, it is imperative to understand the current transport at metal-TMDC interface to reduce the contact resistance in-order to increase the ON-current of the device. However, as explained in the Chapter 1, low-resistance contacts to few-layer sheets of 2D semiconductors are notoriously difficult to achieve. The typical method to reduce contact resistance in 2D-semiconductor devices is by applying a large back-gate bias to accumulate carriers in the semiconductor to ‘turn on’ the contacts. Even when the back gate ‘turns-on’ the contacts, the specific contact resistivity is still considerable, often in the range of 10 kΩ·µm or more [80]. Further complicating the picture is that it is still not firmly established how current flows through a contact to a TMDC in the commonly used top-contacted architecture — does the current flow vertically over some contact area and then laterally into the
semiconductor channel [100, 74], or does all the current flow diagonally at the edge of the contact, tunneling directly into the semiconductor channel [75]? Equivalently, is the transfer length of the metal contact long [101] or short [102]? An improved understanding of contacts to TMDCs is necessary to model current flow and develop techniques to reduce contact resistance.

For understanding this, we model the unique electrostatics of top contacts to monolayer TMDCs and show how a back gate ‘turns on’ the contact, not by accumulating carriers under the contact, but by reducing the lateral depletion width into the semiconductor channel (as shown in Figure 3.6). Increasing the back-gate bias reduces the tunneling distance from the contact into the semiconductor channel, which increases the transmission probability and turns the contact ‘on’. These concepts will be developed in the manuscript by first analyzing a simple structure — a top metal contact to an n-type 2D semiconductor without a back gate. From this structure, a simple analytical model will be developed to explain the potential profile across the device. This model will then be extended to a back-gated structure, more representative of typical devices. We will show that the very presence of a vertical junction to a 2D semiconductor — regardless of a back gate — gives rise to the corner effect, a sharp change in the electrostatic potential at the corner of the junction to 2D semiconductors. The corner effect has been observed in simulations [100, 103] presented in previous works, but its physical origin has not been elucidated. Here, we show that the corner effect results from the fundamental electrostatics of 2D vertical junctions and is therefore universally present in all vertical junctions to 2D semiconductors.

3.2.1 Corner effect in a metal/2D-semiconductor junction

To derive the corner effect, we first examine a simple vertical metal/2D-semiconductor Schottky junction, where a thin metal contact lies on top of a 2D semiconductor, depicted in Figure 3.4. This is typical of high work function metal contacts to TMDCs along with
contacts with dielectric interlayers [104, 105]. The difference in the metal and semiconductor work functions induces a built-in potential ($\phi_{bi}$) across the metal/semiconductor junction. The built-in potential depletes a portion of the semiconductor as shown in Figure 3.4a. The semiconductor depletion charge induces mirror charge in the metal contact (Figure 3.4b). As will be derived, the charge profile yields a sharp change in the potential profile at the corner of the junction, which we call the corner effect (Figure 3.4c).

While exact values for $\phi_{bi}$ for a given metal/semiconductor combination are difficult to theoretically predict due to non-idealities at the metal/semiconductor, Schottky barrier heights ($\phi_{Bn}$) are often experimentally extracted for a range of metal/semiconductor combinations. The Schottky barrier for an n-type semiconductor is related to built-in potential by

$$\phi_{Bn} = \phi_{bi} + \phi_n,$$  \hspace{1cm} (3.10)

where $\phi_n$ is the separation of the conduction band edge and Fermi level in the channel.
far from the metal contact defined as

$$\phi_n \equiv \frac{E_c - E_F}{q} \bigg|_{\text{channel}}. \quad (3.11)$$

Since the contact resistance is most important when the device is biased in the on-state with lots of free carriers in the channel, which implies that $\phi_n$ is small and, therefore, $\phi_{Bn} \approx \phi_{bi}$. Given the large contact resistances reported experimentally [106], we expect $\phi_{Bn}$ (and, therefore, $\phi_{bi}$ as well) to be substantial (i.e. $>100$ mV) for most contacts to few-layer 2D semiconductors. This suggests that Schottky contacts deplete the semiconductor near the contacts. While small Schottky barriers to 2D materials have been previously reported (e.g. $<100$ mV), the high specific contact resistivity of these devices ($>3 \text{k} \Omega \cdot \mu\text{m}$) is inconsistent with such low barriers. In the remainder of this section, we derive the corner effect in terms of the built-in potential with understanding that it is intrinsically related to the Schottky barrier through Equation (3.10). We first split the built-in potential drop across the structure into two components as shown in Figure 3.5:

1. a vertical potential drop ($\phi_\perp$) across the metal/semiconductor overlap region, and

2. a lateral potential drop ($\phi_{\text{lat}}$) at the right edge of the junction.

These potential drops must sum to the built-in potential of the junction, i.e. $\phi_\perp + \phi_{\text{lat}} = \phi_{bi}$. We show that the vertical potential drop is quite small under most conditions, which leads to most of the potential dropping laterally at the corner of the metal/semiconductor junction.

**Vertical Potential Drop**

For a vertical metal/2D-semiconductor junction with a large overlap area, we can calculate the vertical potential drop between the layers ($\phi_\perp$) from a simple capacitive model:

$$\phi_\perp = \frac{Q_s}{C_{\text{gap}}} = \frac{qN_{D,2D}}{\epsilon_{\text{gap}}/t_{\text{gap}}}, \quad (3.12)$$
Figure 3.5: The total electrostatic potential across the structure is equal to the built-in potential of the junction ($\phi_{bi}$), which is divided in the vertical potential drop ($\phi_{\perp}$) and lateral potential drop ($\phi_{lat}$). $\phi_{\perp}$ is approximated as the potential drop between two equal and opposite sheet charges. $\phi_{lat}$ is approximated as the lateral potential drop due to the dipole consisting of a line charge at the edge of the metal and a sheet charge representing the depletion width in the semiconductor.

where $Q_s$ is the net sheet charge density on the 2D semiconductor, and $C_{gap}$, $\epsilon_{gap}$, and $t_{gap}$ are the capacitance per area, permittivity, and thickness of the gap between the layers. Assuming the semiconductor is fully depleted, the net sheet charge density on the 2D semiconductor is equal to the elementary charge times the donor sheet density ($Q_s = qN_{D,2D}$). For vertical junctions to 2D semiconductors without a spacer layer, $t_{gap}$ is equal to the van der Waals gap. Previous works have extracted the van der Waals gap to be $\sim0.3 \sim 0.35$ nm with a permittivity of $\sim10\epsilon_o$ giving a gap capacitance of $C_{gap} = 25 \mu F/cm^2$ [107]. For such a large capacitance, $\phi_{\perp}$ is small even for large sheet charge densities. For example, $\phi_{\perp}$ is only 64 mV for $N_{D,2D} = 10^{13}#/cm^2$, a measly 6.4 mV for $N_{D,2D} = 10^{12}#/cm^2$, and a negligible 0.64 mV for $N_{D,2D} = 10^{11}#/cm^2$, assuming the semiconductor under the contact is fully depleted.

Lateral Potential Drop

The lateral potential drop can be found from the difference of the built-in potential of the total junction and the vertical potential drop, since $\phi_{\perp} + \phi_{lat} = \phi_{bi}$. Since the vertical potential drop is surprisingly small for most semiconductor doping densities, nearly all the built-in potential drops laterally across the semiconductor at the corner of
the metal/semiconductor junction.

The shape of the lateral potential profile can be deduced by considering the charge dipole that comprises the lateral extent of the junction. This dipole can be simplified as a uniform sheet charge with density \( N_{D,2D} \) representing the semiconductor depletion charge, extending from \( x = 0 \) to \( x_D \), where \( x = 0 \) is at the right edge of the metal/semiconductor junction and \( x = x_D \) is at the right extent of the lateral depletion width into the semiconductor. The corresponding mirror charge occurs at the corner of the metal, which is approximated as a 1D line charge with a charge density equal and opposite to the lateral depletion charge in the semiconductor. Due to the asymmetry of the lateral charge dipole, nearly all the lateral potential drops in the vicinity of the line charge. In Appendix D, the lateral potential along the semiconductor is derived and the result is plotted in Figure D.1. The plot clearly shows that for \( t_{gap} \ll x_D \), the majority of the potential drop occurs near the corner of the metal contact at \( x = 0 \). Note that the non-monotonic potential near \( x = x_D \) is an artifact from representing the depletion charge as a sheet with hard boundaries instead of allowing a more gradual roll off of the depletion charge near \( x = x_D \). Nevertheless, the key observation — that a sharp change in the lateral potential occurs near the corner of the metal contact — is true even when a more complex depletion charge profile is used.

3.2.2 Vertical Schottky junctions with global back gates

So far, we have considered a vertical metal/2D-semiconductor junction in isolation (i.e. without any gate electrodes). Figure 3.6 depicts a metal/semiconductor junction with the addition of a 300-nm-SiO\(_2\) global back gate, common for devices made from 2D materials. The capacitive coupling to the semiconductor layer under the contact is \( \sim 3,000 \times \) higher for the top contact as compared to the back gate, due to the large capacitance of the van der Waals gap and the low capacitance of 300 nm of SiO\(_2\). Even a large back gate voltage does little to significantly change the potential of the semicon-
ductor under the contact. Therefore, the 300-nm-SiO$_2$ global back gate — used in most 2D-semiconductor devices — is ineffective at accumulating carriers under the contact. Instead, the global back gate turns the contact on by accumulating carriers in the semiconductor channel immediately to the right of the top contact. As shown in Figure 3.6, an positive back-gate voltage pulls the bands down in the semiconductor channel and enables tunneling at the edge of the contact as will be derived in the following text. When 0 V is applied to the back gate, the results for the back-gated structure are similar to those without the back gate (shown in Figure 3.4) since the back gate is far away and a small potential is applied. When 100 V is applied to the back gate, the back gate potential might be expected to accumulate electrons across the entire semiconductor layer. Indeed, electrons are accumulated to the right of the top contact and the depletion region shrinks; however, the semiconductor remains depleted under the contact.

**Current flow across metal/2D-semiconductor junction**

Using the electrostatic analysis from the previous section, we now model how current flows across the metal/2D-semiconductor junction. We divide the structure into a resistor network, as shown in Figure 3.7(a), with current flowing under the metal contact through the red resistors and at edge of the contact through the green resistor. Here, we focus on the current flow under the contact but wish to point out that tunneling current at the edge of the contact can also take place. First, we investigate the case for where we assume that the applied potential drop does not vary underneath the contacts. In this condition, the sheet resistance ($R_{sk}$ in $\Omega/\square$) and the specific interfacial resistivity ($\rho_{C,\perp}$ in $\Omega \cdot \text{cm}^2$) are nearly constant as a function of position under the contact. We then used standard transmission line theory to lump the resistor network under the contact into a single element — the perpendicular resistance ($R_{C,\perp}$) — and to determine the transfer length of the contact. Next, we perform a similar analysis but for a case where we let the applied potential drop to vary with position underneath contact. The applied
Figure 3.6: A 2D semiconductor with a top contact and global back gate. (top) Zero bias is applied to the back gate with respect to the top contact. The 2D semiconductor is depleted due to the built-in potential of the metal/semiconductor junction. The depletion width extends laterally into the 2D layer. (bottom) A bias of 100 V is applied to the back gate. The back gate accumulates electrons in the n-layer to the right of the top contact, which decreases the extent of the lateral depletion width; however the semiconductor remains depleted directly under the top contact. The strong capacitive coupling of the top contact to the 2D layer severely limits the ability of the back gate to affect the potential (and, therefore, the electron density) directly under the top contact.

bias causes a change in the electron quasi-Fermi level as a function of position, which thereby causes $R_{sk}$ and $\rho_{C,\perp}$ to vary with position under the contact. Note that in both cases the electrostatic potential or conduction/valence band edges remain constant due to the corner effect. We solve this more complicated resistor network to find a solution in terms of the same transfer length concept.

3.2.3 Current flow under the contact

Under the metal contact, electrons face two resistances before they reach the channel of the device: the tunneling resistance across the van der Waals gap into the semiconductor, given by the specific interfacial resistivity under the contact ($\rho_{C,\perp}$) and the sheet resistance of the semiconductor layer under the contact ($R_{sk}$). As shown in Figure 3.7(b),
the resistor network of $\rho_{C,\perp}$ and $R_{sk}$ reduces to a single element, $R_{C,\perp}$. As discussed before, we solve the vertical transport in two distinct cases:

1. **No variation of applied potential drop under the metal**
   This is equivalent to the Ohmic contact assumption that almost no potential drops across the metal-semiconductor junction. Using a simple transmission line model (TLM), we can write the perpendicular resistance in a simplified manner (by assuming that the
length of the contact ($L_C$) is much longer than the transfer length ($L_T$)) as [108]

\[ R_{C,\perp} \simeq \frac{\rho_{C,\perp}}{L_T W_C}, \]  

(3.13)

where $W_C$ is width of the contact and

\[ L_T = \sqrt{\frac{\rho_{C,\perp}}{R_{sk}}}. \]  

(3.14)

Combining these expressions gives

\[ R_{C,\perp} = \sqrt{\frac{\rho_{C,\perp} R_{sk}}{W_C}}. \]  

(3.15)

The sheet resistance under the contact is given by

\[ R_{sk} = \frac{1}{q \mu_{\text{eff}} n_{\perp 0}}, \]  

(3.16)

where $\mu_{\text{eff}}$ and $n_{\perp 0}$ are the electron mobility and near-equilibrium sheet density in the semiconductor under the metal contact. Note that experiments have shown that the mobility under the contact (used in Equation (3.16)) is significantly degraded as compared to the channel [102].

All that remains to calculate $R_{C,\perp}$ is the determination of $\rho_{C,\perp}$ which for a monolayer channel can be attributed to the interfacial resistance across the van der Waals gap (or the interlayer distance for contacts with tunnel barriers). To derive an expression for the tunneling current across the gap, we imagine that the electrons are confined and bouncing back-and-forth in a quantum well with a width equal to the semiconductor thickness, as shown in Figure 3.7(c). Assuming infinite barriers on both sides of the well, the ground-state electron confinement energy is equal to

\[ E_{\perp} = \frac{\pi^2 \hbar^2}{2 m^*_{\text{eff}} t_{\text{semi}}^2}, \]  

(3.17)
where $\hbar$ is the reduced Planck constant, $m^*_e$ is the effective mass for electrons in the conduction band, and $t_{semi}$ is the thickness of the semiconductor layer. Using the typical expression for kinetic energy ($E_\perp = \frac{1}{2}m^*_e v^2_\perp$), the perpendicular velocity at which the electrons bounce back-and-forth in the well is given by

$$v_\perp = \sqrt{\frac{2E_\perp}{m^*_e}} = \frac{\pi \hbar}{m^*_e t_{semi}}. \quad (3.18)$$

We determine the frequency at which the electron impinges on the barrier of the van der Waals gap, commonly called the attempt frequency, as

$$f = \frac{v_\perp}{2t_{semi}} = \frac{\pi \hbar}{2m^*_e t^2_{semi}}. \quad (3.19)$$

Each time the electron impinges on the barrier, it tunnels through the van der Waals gap with transmission probability $T_\perp$. Note that we have assumed that the transmission probability $T_\perp$ is independent of energy. Even though this is an approximation, the small $\phi_\perp$ across the van der Waals (vdW) gap justifies the assumption. This assumption also allows to determine a simple analytical form for $T_\perp$ using WKB approximation of rectangular barrier [109]

$$T_\perp = \exp \left( -\sqrt{\frac{2m^*_e (\chi_{TMDC} - \phi_\perp)}{\hbar^2}} t_{vdW} \right) \quad (3.20)$$

Thus, the electron current density ($A/cm^2$) from the semiconductor to metal is given by

$$J_{s \rightarrow m} = -qfT_\perp n_{\perp 0} \exp \left( \frac{-qV_a}{kT} \right) \quad (3.21)$$

where, $V_a$ is the applied potential to the semiconductor with respect to the metal.
The total perpendicular current density is given by

\[ J_\bot = J_{s \rightarrow m} - J_{m \rightarrow s}, \]  

(3.22)

where \( J_{m \rightarrow s} \) is the electron current from the metal to the semiconductor. By near-equilibrium, we mean that a constant negligible bias voltage is applied laterally across the structure yielding a flat Fermi level in the semiconductor, but a non-zero back gate voltage may be present. In this case, \( n_{\bot o} \) can be given as

\[ n_{\bot o} = N_{C,2D} \exp \left( -\frac{(\phi_B - \phi_\bot + E_\bot)}{k_B T} \right) \]  

(3.23)

In equilibrium (i.e. \( V_a = 0 \)), the net current must be zero so that

\[ J_{m \rightarrow s} = J_{s \rightarrow m}(V_a = 0) = -q f T_\bot n_{\bot o}, \]  

(3.24)

which yields the net perpendicular current density as

\[ J_\bot = q f T_\bot n_{\bot o} \left( 1 - \exp \left( \frac{-q V_a}{kT} \right) \right). \]  

(3.25)

For small applied voltages at the metal-contact, we take the first order Taylor series expansion of the exponential term to arrive at

\[ J_\bot = q f T_\bot n_{\bot o} \left( \frac{q V_a}{kT} \right) \text{ for small } V_a. \]  

(3.26)

From this expression, we can easily calculate the specific interfacial resistivity under the contact as

\[ \rho_{C,\bot} = \left. \frac{\partial V_a}{\partial J_\bot} \right|_{V_a=0} = \frac{1}{q f T_\bot n_{\bot o}}. \]  

(3.27)
Substituting Equations (3.16) and (3.45) into Equation (3.13) yields

\[ R_{C,\perp} = \frac{1}{W_C} \cdot \frac{1}{qn_{\perp o}} \sqrt{\frac{kT}{q}} \mu_{eff} T_{\perp} \]  

(3.28)

or, equivalently, in terms of conductance as

\[ G_{C,\perp} = W_C \cdot qn_{\perp o} \sqrt{\frac{\mu_{eff} T_{\perp}}{kT}} \]  

(3.29)

From the derivation in the previous section, we can calculate the transfer length (defined in Equation (3.14)) as

\[ L_T = \sqrt{\frac{kT \mu_{eff}}{q \cdot f T_{\perp}}} = \sqrt{\frac{2}{\pi \hbar} \frac{kT}{q} \frac{\mu_{eff} m^*_{\text{semi}} t_{\text{semi}}^2}{T_{\perp}}}. \]  

(3.30)

2. Applied potential drop varies under the metal

We start with defining a lateral current \( I_{\text{lat}} \) in the semiconductor underneath the metal which changes with \( x \), where \( x = 0 \) is the start of the interface between metal and semiconductor from the semiconductor side. As shown before, the electrostatic potential underneath the contact is nearly constant, and thus we can assume that most of the current flows due to diffusion rather than drift.

\[ I_{\text{lat}}(x) = qD_e W_C \frac{dn(x)}{dx} \]  

(3.31)

Due to the applied potential drop, the carrier density \( n_x \) will change with position which changes the form of \( J_{\perp} \) to

\[ J_{\perp}(x) = qf T_{\perp} (n_{\perp o} - n(x)) \]  

(3.32)
which gives

\[
\frac{dJ_\perp(x)}{dx} = -q f T_\perp \frac{dn(x)}{dx}
\]  \hspace{1cm} (3.33)

Using Equations (3.31) and (3.33) we get,

\[
I_{lat}(x) = -\frac{D_e W_C}{f T_\perp} \frac{dJ_\perp(x)}{dx}
\]  \hspace{1cm} (3.34)

We can rewrite Equation (3.34) as

\[
I_{lat}(x) = -L_T^2 W_C \frac{dJ_\perp(x)}{dx}
\]  \hspace{1cm} (3.35)

where, \( L_T \) is the transfer length has the form

\[
L_T = \sqrt{\frac{D_e}{f T_\perp}} = \sqrt{\frac{kT \mu_{eff}}{q f T_\perp}}
\]  \hspace{1cm} (3.36)

where, we have used the Einstein’s relation within Maxwell-Boltzmann approximation. This is same as Equation (3.30) which is interesting and shows that for contacts to 2D semiconductors, the transfer length is independent of the carrier concentration under the metal contact.

Taking second derivative of Equation (3.35), we get

\[
\frac{d^2I_{lat}(x)}{dx^2} = -L_T^2 W_C \frac{d^2J_\perp(x)}{dx^2}
\]  \hspace{1cm} (3.37)

Using nodal analysis at point \( x \) in Figure 3.7, we get

\[
I_{lat}(x) - I_{lat}(x + dx) = J_\perp W_C \Delta x
\]  \hspace{1cm} (3.38)
which gives

\[
\frac{dI_{\text{lat}}(x)}{dx} = -J_\perp W_C
\]  

(3.39)

Combining Equation (3.39) with Equation (3.37) provides

\[
\frac{d^2 J_\perp(x)}{d^2x} \frac{J_\perp(x)}{L_T^2} = 0
\]

(3.40)

Equation (3.41) is of the form of second-order linear differential equation whose solutions can be expressed in terms of hyperbolic functions as

\[
J_\perp(x) = A \sinh \left( \frac{x}{L_T} \right) + B \cosh \left( \frac{x}{L_T} \right)
\]

(3.41)

We find \(A\) and \(B\) by using the boundary conditions that \(J_\perp(x = L_C) = 0\) and \(J_\perp(x = 0)\) is given by Equation (3.25), which gives

\[
J_\perp(x) = q f T_\perp n_\perp \left( 1 - \exp \left( -\frac{q V a}{kT} \right) \right) \left[ \cosh \left( \frac{x}{L_T} \right) - \coth \left( \frac{L_C}{L_T} \right) \sinh \left( \frac{x}{L_T} \right) \right]
\]

(3.42)

Next, we use Eq. (3.39) to derive \(I_{\text{lat}}(x)\) as

\[
I_{\text{lat}}(x) = qf L_T T_\perp n_\perp \left( 1 - \exp \left( -\frac{q V a}{kT} \right) \right) \left[ \coth \left( \frac{L_C}{L_T} \right) \cosh \left( \frac{x}{L_T} \right) - \sinh \left( \frac{x}{L_T} \right) \right]
\]

(3.43)

We can then calculate the contact resistance \(R_{C,\perp}\) Using Equation (3.42), we can then define \(\rho_{C,\perp}\) for two specific situations
1. If $V_a < \frac{kT}{q}$ (typical of Ohmic contacts):

$$\rho_{C,\perp}(x) = \frac{\partial V_a}{\partial J_\perp(x)}\bigg|_{V_a=0} = \frac{kT}{q f T_\perp n_{\perp o}} \left[ \cosh \left( \frac{x}{L_T} \right) - \coth \left( \frac{L_C}{L_T} \right) \sinh \left( \frac{x}{L_T} \right) \right].$$

(3.44)

2. If $V_a \geq \frac{kT}{q}$ (typical of Schottky contacts):

$$\rho_{C,\perp}(x) = \frac{V_a}{J_\perp(x)} = \frac{V_a}{q f T_\perp n_{\perp o}} \left[ \cosh \left( \frac{x}{L_T} \right) - \coth \left( \frac{L_C}{L_T} \right) \sinh \left( \frac{x}{L_T} \right) \right].$$

(3.45)

Since $\rho_{C,\perp}$ changes with $x$, we define an average quantity to get a quantity similar to $R_{C,\perp}$. We begin with defining $<\rho_{C,\perp}>$ as

$$<\rho_{C,\perp}> = \frac{\int_0^{L_T} \rho_{C,\perp}(x) \, dx}{L_T}$$

(3.46)

which gives

$$<\rho_{C,\perp}> = \frac{V_a}{q f T_\perp n_{\perp o}} \sinh \left( \frac{L_C}{L_T} \right) \log \left( \tanh \left( \frac{L_C}{2L_T} \right) \coth \left( \frac{L_C - L_T}{2L_T} \right) \right)$$

(3.47)

where $V_a$ can be replaced by $\frac{kT}{q}$ for Ohmic contacts. Finally, using Equation (3.13) we get

$$R_{C,\perp} \approx \frac{V_a}{q f T_\perp n_{\perp o} L_T W_C} \sinh \left( \frac{L_C}{L_T} \right) \log \left( \tanh \left( \frac{L_C}{2L_T} \right) \coth \left( \frac{L_C - L_T}{2L_T} \right) \right)$$

(3.48)

For $L_C >> L_T$, the above equation can be simplified to

$$R_{C,\perp} \approx \frac{V_a}{q f T_\perp n_{\perp o} L_T W_C} \frac{e}{e - 1}$$

(3.49)

where, $e$ is the Euler’s constant. Thus, $R_{C,\perp}$ increases than the near-equilibrium
case when we consider that potential drop changes along the semiconductor under the contact.

3.2.4 Current flow at the edge of the contact

At the edge of the contact, electrons tunnel across the van der Waals gap directly into the semiconductor channel. We employ the Landauer transport formulation to calculate the edge current as

$$I_{\text{edge}} = -\frac{2q}{h} \int_{-\infty}^{+\infty} T(E) M(E) (f_{\text{ch}}(E) - f_{\text{m}}(E)) \, dE$$  \hspace{1cm} (3.50)

where $h$ is the Plank constant, $T(E)$ is the transmission probability, $M(E)$ is number of modes, and $f_{\text{ch}}(E)$ and $f_{\text{m}}(E)$ are the Fermi-Dirac distributions of the channel and metal as a function of energy. Within Maxwell-Boltzmann approximation for non-degenerate carrier concentrations, this equation is similar to one given by Equation (3.2) [81].

We thus approximate the Fermi-Dirac distributions of the channel and metal using the Maxwell-Boltzmann approximation,

$$f_{\text{ch}}(E) - f_{\text{m}}(E) \approx \exp \left( \frac{E_{f,\text{ch}} - E}{kT} \right) - \exp \left( \frac{E_{f,m} - E}{kT} \right)$$ \hspace{1cm} (Maxwell-Boltzmann)

$$= \exp \left( \frac{E_{f,\text{ch}} - E}{kT} \right) \left( 1 - \exp \left( \frac{E_{f,m} - E_{f,\text{ch}}}{kT} \right) \right)$$

$$= \exp \left( \frac{E_{f,\text{ch}} - E}{kT} \right) \left( 1 - \exp \left( \frac{qV_a}{kT} \right) \right)$$

$$\approx \exp \left( \frac{E_{f,\text{ch}} - E}{kT} \right) \left( -\frac{qV_a}{kT} \right) \text{ for small } V_a.$$  \hspace{1cm} (3.51)

Here, $E_{f,\text{ch}}$ and $E_{f,m}$ are the Fermi levels of the channel and metal, and $V_a$ is the voltage applied to the channel with respect to the metal contact. The last line of Eq. (3.51) uses a Taylor series expansion to linearize the expression for small $V_a$. 

82
Substituting Eq. (3.51) into the Landauer expression for current gives

\[ I_{\text{edge}} = \frac{2q}{h} \left( \frac{qV_a}{kT} \right) \int_{-\infty}^{+\infty} T(E)M(E) \exp\left( \frac{E_{f,ch} - E}{kT} \right) dE, \]  

which can be expressed in terms of edge conductance as

\[ G_{\text{edge}} = \frac{2q^2}{h} \frac{1}{kT} \int_{-\infty}^{+\infty} T(E)M(E) \exp\left( \frac{E_{f,ch} - E}{kT} \right) dE. \]

As apparent in Eq. (3.53), the edge conductance depends on the transmission probability and the number of modes. Note that, we can always tune Equation (3.53) to Schottky contacts by following the methodology provided in earlier discussion. The number of modes is calculated analytically as described in Appendix E, and the WKB approximation is used to determine the transmission probability at a given energy based on a simulated energy band diagram (obtained using self-consistent drift-diffusion simulation) along the semiconductor at the edge of the metal as detailed in previous section.

**Edge transport or Vertical transport**

We then calculate \( G_{C,\perp} \) and \( G_{\text{edge}} \) as a function of Schottky barrier height and carrier concentration in the channel (far away from the metal contact) assuming small \( V_a \). The results are shown in Figure 3.8. We can clearly see that expect for very small \( \phi_{Bn} \) and low carrier density \( (n_{2D}) \), the edge conductance is always higher than the vertical path. Therefore, most current flows through the edge rather than the vertical path. This is further supported by the calculation of true transfer lengths in 2D semiconductors. Using approximate values for 2D semiconductors (i.e. \( m^*_v = 0.5m_o, \mu_{eff} = 10 \text{ cm}^2/\text{V} \cdot \text{s}, \) \( t_{\text{semi}} = 0.7 \text{ nm}, \) and \( T = 300 \text{ K} \)), we estimate \( L_T \) to be 0.19 nm for a perpendicular transmission probability \( (T_{\perp}) \) of 1. We use a degraded mobility due to increased scattering caused by metallization-induced damage and polymer residue. Determining the exact value of \( T_{\perp} \) is challenging since the details of the interface between the metal and 2D
Figure 3.8: Edge \((G_{\text{edge}})\) and vertical \((G_{C,\perp})\) conductance as function of carrier density \(n_{2D}\) sufficiently far from the contact with different Schottky barrier height, \(\Phi_{Bn}\). Relative dielectric constant at van der Waals gap between metal and TMDCs is determined from 1 (dashed line) to 10 (solid line). The \(G_{\text{edge}}\) is larger than \(G_{C,\perp}\) for monolayer TMDCs for all practical SBH and carrier densities.

semiconductor are typically unknown, but given the extremely small van der Waals gap between the materials, it is reasonable to assume a rather high transmission probability in the perpendicular direction. If we use Equation (3.20) for evaluating \(T_{\perp}\), we get \(T_{\perp} = 0.12\) for which \(L_T\) is estimated to be 0.6 nm, which is still quite small. This analysis suggests that regardless of semiconductor doping, the transfer length will be quite small, estimated to be less than 20 nm for a variety of semiconductor parameters. We further plotted total \(R_C\) as a function of carrier density in Figure 3.9. The figure clearly shows that to reduce carrier density in ON state, degenerate doping \((> 10^{13} \# /\text{cm}^2)\) is required to reduce \(R_C\) below the required 100 \(\Omega.\mu m\) [110]. Also, SBH reduction would help in reducing contact resistance which can be achieved using cleaner contacts [29].

3.2.5 Summary

This analysis suggests that nearly all contacts to 2D semiconductors operate as ‘edge contacts’ due to the intrinsic electrostatics that give rise to the corner effect. Therefore, optimization of contacts to 2D materials require a more significant emphasis on what is
Figure 3.9: Contact resistance ($R_C$) as function of carrier density ($n_{2D}$) shows that degenerate doping is required to lower $R_C$

happening at the edge of the contact as compared to under the contact. Strong doping can result in substantial decrease in contact resistance irrespective of the Schottky barrier height. Moreover, ultra-clean contacts can provide reduction in the contact resistance by reducing the overall Schottky barrier.
Chapter 4: Pitfalls in Electrical Measurements of 2D Transistors

Till now, we have primarily focused on understanding carrier transport in 2D transistors. The simulation and modeling results presented in previous chapters show the importance of both channel and contact regions in different operating regimes of 2D transistors. In this chapter, we analyze different experimental methods used to evaluate important device parameters such as mobility, carrier density, contact resistance, and Schottky barrier height. Over the years, conventional techniques used for extracting these parameters from bulk devices have been adopted by researchers to study 2D transistors; however, a detailed discussion on possible pitfalls associated with application of these techniques to 2D materials is not available in literature. Here, we try to study the interdependence of these parameters and provide best practices for accurate analysis of 2D transistors.

4.1 Contact resistance and Schottky barrier extraction

Lack of simple, efficient, and controllable doping techniques for 2D materials results in large contact resistance \( R_C \) at the metal-semiconductor junction. \( R_C \) depends on the nature of the barrier, i.e., its width and height, since carrier transport across the barrier is severely affected by these parameters. For the conventional semiconductors, e.g. Si and GaAs, \( R_C \) is known to approach near the quantum mechanical limit [110]. However, as discussed in earlier chapters, 2D semiconducting materials such as MoS\(_2\), WSe\(_2\) that have a sizable bandgap in the range of \( \sim 2 \) eV, show very high \( R_C \) > 10 times that of the conventional semiconductor materials. The large \( R_C \) at the metal-

\(^1\)This chapter is based on [73, 111] and reproduced with the permission of IOP Publishing.
semiconductor interface is often attributed to the formation of Schottky barriers due to mid-gap Fermi level pinning arising from intrinsic material defects and processing conditions [112, 113]. These Schottky barriers not only limit the ON current of the 2D FETs, but also determine their polarity [114, 115]. As discussed in chapter 3, typical back-gated 2D devices involves simultaneous gating of contact and channel regions, which further convolutes the underlying physics. Thus, the output and transfer characteristics of such FET devices represent contact properties rather than channel properties. This limits the performance of scaled 2D FETs and affects extraction of important device parameters such as field-effect mobility and carrier density, as shown in the following sections. Thus, accurate estimation of $R_C$ is critical for understanding, improving, and benchmarking 2D devices.

In this section, we discuss the widely employed TLM technique used to estimate contact resistance in 2D FETs. We discuss the advantages and disadvantages of the method and highlight important considerations that should be taken into account when applying it to 2D materials. We further discuss the different methods of SBH extraction and show that most experimental techniques results in severe underestimation of extracted SBHs in 2D devices.

4.1.1 Transfer line method

The TLM/transfer length method is conventionally used to determine $R_C$ for metal contacts on bulk semiconductors, such as Si and Ge [116, 117]. In this method, multiple devices are fabricated with TLM geometry (Figure 4.1(a)), where the channel length/spacing (denoted by $L_1, L_2$, etc) is varied between different contacts, while the contact length is kept constant. As shown in the inset of Figure 4.1, $R_{Total}$ between any two contacts can be expressed as a linear combination of $R_C$ and the length-dependent $R_{CH}$ of the
Figure 4.1: Extraction of $R_C$ and $L_T$. (a) Top view of the TLM configuration showing different channel lengths (L1, L2,...). The enlarged view shows the distribution of total resistance in terms of $R_C$ and $R_{CH}$. (b) A linear fit of the plot of $R_{Total}$ versus channel length giving rise to $R_C$, $R_{SH}$, and $L_T$. (c) Schematic of a MoS$_2$-based TLM device with back gating through SiO$_2$. (d) Schematic of a bilayer MoS$_2$ device and the TLM plot for the device showing the linear trend of $R_{Total}$ versus channel length as a function of carrier density (gate voltage). The inset shows the extracted $R_C$ values as a function of carrier density, demonstrating contact region gating.

semiconductor in between the contacts, i.e.

$$R_{Total} = R_{CH}(L) + 2R_C$$

which can be further rewritten as

$$R_{Total}W = R_{SH}(L) + 2R_CW.$$  \hspace{1cm} (4.2)

where, $R_{SH}$ is denoted as the sheet resistance of the channel region.

Note that the term $R_CW$ is sometimes used to refer to $R_C$ in literature, where it represents width-normalized $R_C$. Several 2-probe resistance measurements are made between an adjacent pair of contacts with different channel lengths and $R_{Total}$ is plotted.
as a function of channel length. Figure 4.1(b) shows a typical plot of $R_{Total}$ versus $L$ from which $R_C$ can be extracted by finding the y-intercept using a linear fit. Other relevant parameters are also highlighted in the plot. Figure 4.1(c) shows a schematic of a typical TLM structure with a 2D material as the channel material and conventional back-gated geometry. Unlike bulk semiconductors, 2D materials generally do not conduct well without gating due to large $R_C$. Thus, Equation (4.2) needs to be modified to show the effect of global back gating, in which case both the channel and contact regions are modified simultaneously, i.e.

\[ R_{Total}(V_{GS})W = R_{SH}(V_{GS})(L) + 2R_C(V_{GS})W. \]  

(4.3)

Figure 4.1(d) illustrates the use of TLM to extract contact resistance for Au contacts on a bilayer MoS$_2$ where the channel length was varied from 200 to 1000 nm [102]. The measured total resistance ($R_{Total}W$) was plotted as a function of channel length; the corresponding y-intercept provides the contact resistance ($R_CW$). As discussed earlier, the contact resistance shows clear gate voltage dependence (highlighted by carrier density in the channel using Equation (4.3)), as contact resistance decreases with an increase in gate voltage.

**Transfer length and contact resistivity extraction**

TLM also provides a simple way to study the scaling properties of contacts, which is crucial to determine the fundamental limits to scaling of 2D materials-based FETs. As the channel length is scaled to enable better electrostatics and achieve higher device density, a large portion of total resistance corresponds to the contact resistance resulting in contact-dominated behavior of scaled devices. Using a distributed resistive network model for the contact region, analytical expressions for contact resistance can be obtained in terms of specific contact resistivity ($\rho_c$), sheet resistance under contact ($R_{SK}$), and
transfer length \((L_T)\) which can be given as

\[
R_C W = \sqrt{\rho_c R_{SK}} \coth \frac{L_C}{L_T}; \quad L_T = \sqrt{\frac{\rho_c}{R_{SK}}}. \tag{4.4}
\]

Here, \(L_C\) denotes the physical contact length and \(L_T\) represents the transfer length introduced in chapter 3. Further insight can be gained by considering two limiting cases:

1. \(L_C \geq L_T\):

\[
R_C W = \sqrt{\rho_c R_{SK}} = L_T R_{SK} \tag{4.5}
\]

2. \(L_C \leq L_T\):

\[
R_C W = \frac{\rho_c}{L_C} \tag{4.6}
\]

Experimentally, these parameters are extracted from TLM by assuming that \(R_{SK} = R_{SH}\) and \(L_C \gg L_T\) which allows us to extract \(L_T\) by finding the \(x\)-intercept of the curve of \(R_{Total}\) versus \(L\). Once \(L_T\) is known, \(\rho_c\) can be determined by using the either Equation (4.5) or Equation (4.6).

### 4.1.2 Challenges with TLM

Over time, the TLM has become the most commonly employed method of determining \(R_C\) and \(R_{SH}\) in 2D devices due to the ease of device fabrication and straightforward nature of the analysis. Furthermore, the TLM has an advantage over 4PP as current transport is not disrupted by the presence of inner electrodes, which are used as voltage probes in typical 4-probe measurements [102, 76]. However, a few potential pitfalls must be considered when applying TLM to 2D FET analysis:

1. Reliable TLM requires linear dependence of channel resistance on channel length and low spatial variation of contact resistance. Fabrication issues such as irregu-
lar device geometry due to non-patterned 2D flakes, inhomogeneous non-laminar current flow due to polymer contamination, lithography-induced damage, and unknown contributions from sample edges, can cause deviation from linear scaling of channel resistance and therefore result in erroneous contact resistance measurements.

2. TLM is also problematic when contact resistance is substantially higher than channel resistance, since a small amount of inter-device variation in contact resistance can cause large errors in the linear fit. For Schottky contacts with non-linear $I-V$ characteristics, $R_C$ becomes drain bias-dependent, which needs to be carefully considered when examining scaling behavior. The impact of non-linearity in the plot of $R_{Total}$ versus $L$ is severe when the extracted transfer lengths are small.

3. Extracting transfer length and specific contact resistivity requires that $R_{SK} = R_{SH}$ holds true, which is hard to justify for few-layer devices. To see this, we rewrite Equation (4.2) in terms of $R_{SK}$ and $R_{SH}$ using Equation (4.5) as

$$R_{Total}W = R_{SH}(L) + 2L_TR_{SK} \tag{4.7}$$

From this, we can extract expression for $L_T$ by setting $R_{Total} = 0$ as

$$L_T = -\frac{R_{SH}}{2R_{SK}}L \tag{4.8}$$

If we set $R_{SK} = R_{SH}$, we can extract $L_T$ as indicated in Figure 4.1(b). However, unlike conventional semiconductors, in which lateral transport occurs far ($\sim 10–100$ nm) from the metal-semiconductor interface, transport in 2D materials occurs right at the interface and the material properties are substantially changed by the metal contacts (e.g. contact doping, fabrication-induced damage, and change in bandgap) as discussed in previous chapter. In our recent experimental studies, we have seen
significant differences in $R_{SK}$ and $R_{SH}$ values where $R_{SK}$ could be easily $4\sim15 \times R_{SH}$. This observation is well-supported our theoretical investigations in chapter 3 and by other experimental works [118]. For $R_{SK} > R_{SH}$, the actual transfer length is significantly smaller than extracted transfer lengths and thus also supports Corner effect phenomena. In this scenario, it is better to determine $L_T$ from end-resistance measurements [108].

4.1.3 Schottky barrier heights and Fermi-level pinning

Figure 4.2: Schottky barrier height and Fermi level pinning. (a) Band diagram of a metal-semiconductor junction. (b) SBH versus metal work function showing the Schottky–Mott rule and Fermi level pinning. (c) Schematic image of Fermi level pinning [119].

As discussed in the previous section, the large contact resistance in 2D devices can be attributed to the presence of Schottky junctions at the metal-2D semiconductor interfaces. Schottky junctions are characterized by SBHs, the relative values of which determine the current transport at the metal-semiconductor interface affecting the polarity, magnitude, and switching characteristics of the injected charge carriers. Figure 4.2(a) shows the SBH and conceptual band diagram of a metal-2D semiconductor interface. For an ideal metal-2D semiconductor junction, the SBH for $n$-type ($\phi_{Bn}$) or $p$-type ($\phi_{Bp}$)
semiconductors is given by

\[ \text{For } n - \text{type : } \phi_{Bn} = \phi_m - \chi, \]  \hspace{1cm} (4.9) \]

\[ \text{For } p - \text{type : } \phi_{Bp} = E_g + \chi - \phi_m, \]  \hspace{1cm} (4.10) \]

where \( \phi_m \) is the work function of a metal, \( \chi \) is the electron affinity and \( E_g \) is the bandgap of the 2D semiconductor. For such ideal systems, the SBH for electrons increases linearly with the metal work function, thus satisfying the Schottky–Mott rule as shown in Figure 4.2(b). However, non-ideal states such as interface and gap states at the metal-semiconductor interface can cause severe deviation from the Schottky–Mott rule, making it difficult to control electron/hole SBH by varying the metal work function [120]. Quantitatively, we can interpret this deviation by introducing a pinning factor (\( S \)) and charge neutrality level (CNL, \( \phi_{CNL} \)) [121]:

\[ \text{For } n - \text{type : } \phi_{Bn} = S(\phi_m - \phi_{CNL}) + (\phi_{CNL} - \chi), \]  \hspace{1cm} (4.11) \]

\[ \text{For } p - \text{type : } \phi_{Bp} = E_g + \chi - \phi_m, \]  \hspace{1cm} (4.12) \]

Here, \( S \) is defined as the slope \( S = \frac{\partial \phi_{Bn}}{\partial \phi_m} \) and can be calculated from the linear fit of \( \phi_{Bn} \) versus \( \phi_m \) plot. \( S = 1 \) represents an ideal metal-semiconductor interface whereas \( S = 0 \) represents almost no variation in SBH with a change in the metal work function, indicating a completely pinned interface at the charge neutrality level. The CNL for n-type can be estimated by the relation

\[ \phi_{CNL} = \frac{\chi + b}{1 - S} \]  \hspace{1cm} (4.13) \]

For \( S < 1 \), the semiconductor Fermi level is fixed near the CNL, which results in
Figure 4.3: The top panel compares the theoretical and experimental values of electron Schottky barrier height (SBH) for different metal contacts to few-layer MoS$_2$. The theoretical values are generally higher than the experimental values indicating a systematic underestimation by the temperature-dependent SBH extraction method. The bottom panel highlights the absolute error (disparity) in the measured SBH values compared to theoretical values determined by comparing the mean values of experimental and theoretical values for different metals, provided in top panel.

similar SBHs for different metal contacts often referred as Fermi level pinning, as shown in Figure 4.2.

4.1.4 Challenges with SBH extraction

Figure 4.3 shows the comparison between the experimentally extracted and theoretical values of electron Schottky barrier height ($\phi_{Bn}$) for various metal contacts to
few(< 10)-layer MoS$_2$. The theoretical values are calculated for metal-MoS$_2$ contacts with top contact geometry using \textit{ab initio} density functional calculations [122, 123, 124, 125], whereas the experimental values were extracted using the well-known Arrhenius method which uses temperature-dependent I-V measurements [126, 127, 71, 128, 129, 130, 131, 132]. The top panel shows the disparity between the experimental and theoretical values for any metal contacts. Unlike the theoretical calculations, the experimental values show minimal variation of $\phi_{Bn}$ with variation in metal work function, indicating severe underestimation of barrier height especially for metals with large work functions. To further emphasize on the erroneous nature of Arrhenius method of SBH extraction in 2D devices, we plotted the difference between the median of the experimental and theoretical values, respectively, in the lower panel. This data clearly show that across the board, the SBH extracted by the Arrhenius method is around 200 to 500 meV less than what is expected from DFT calculations. In this light, a deeper investigation of Arrhenius method is critical for evaluating its suitability for studying 2D devices. Here, we review the fundamental assumptions of the Arrhenius method and provide reasoning for their invalidity with respect to back-gated 2D transistors.

**Method of temperature-dependent Arrhenius extraction**

To understand the limitations of the Arrhenius method of barrier extraction, we first provide a detailed overview of this method by analyzing the temperature dependent I-V characteristics of a back-gated monolayer MoS$_2$ transistor with Co/1L $h$-BN contacts previously demonstrated by Cui et al. [129]. Cobalt on 1L $h$-BN provides high-quality contacts to monolayer MoS$_2$ with one of the lowest ($< 3 \text{k}\Omega \cdot \mu\text{m}$) reported contact resistance and smallest (18 meV) extracted barrier height. Additionally, 1L $h$-BN also provides excellent passivation, yielding in reliable device performance over more than 3 years. Figure 4.4(a) shows the optical image (left) and the circuit diagram (right) of the device used for this study.
Figure 4.4: **Arrhenius method of Schottky barrier extraction:** (a). Schematic and circuit diagram of the back-gated monolayer MoS₂ transistor. The 1L-hBN acts as a tunnel barrier and provides excellent contacts to monolayer MoS₂. (b). Temperature dependent transfer characteristics of the device shown in part (a). The expected region \((V_{GS} < V_{FBSB})\) which is dominated by thermionic transport is highlighted in grey. (c). Arrhenius plots corresponding to the transfer characteristics in part (b) at different gate voltages. The voltage-dependent slopes (highlighted in red) were used to extract activation energy \((E_A)\)as a function of gate voltage. (d). Activation energy as a function of gate voltage. The true electron barrier \((\Phi_{BN})\) was extracted by identifying the point at which the slope (shown by solid black line) deviates from the linear trend at low gate voltages.
In Arrhenius method, the barrier height is determined at the source contact by extracting the activation energy in the thermionic regime of the 2D transistor. The Schottky barrier height is then deduced using the following methodology.

1. Transfer characteristics ($I_D-V_{GS}$) of the back-gated 2D transistor are measured at different temperatures for $V_{DS} \gg \frac{k_B T}{q}$, as shown in Fig. 4.4(b).

2. The activation energies are then extracted from the slope of the gate-dependent Arrhenius curves ($\ln(\frac{I_D}{T^{1.5}})$ versus $\frac{1000}{T}$) using the thermionic emission model given by

\[
I_D = WA^*_2D T^{1.5} \exp \left( -\frac{E_A(V_{GS})}{k_B T} \right) \left[ 1 - \exp \left( -\frac{qV_{DS}}{k_B T} \right) \right], \tag{4.14}
\]

where, $W$ is the channel width, $A^*_2D$ is the 2D Richardson constant, $T$ is the temperature, $k_B$ is the Boltzmann constant, and $E_A(V_{GS})$ represents the gate-voltage-dependent activation energy that represents the apparent barrier height at the source contact. For $V_{DS} \gg \frac{k_B T}{q}$, Equation (4.14) simplifies to

\[
I_D = WA^*_2D T^{1.5} \exp \left( -\frac{E_A(V_{GS})}{k_B T} \right). \tag{4.15}
\]

The gate-dependent activation energies extracted from the slope of the Arrhenius curves are shown by the red lines in Fig. 4.4(c). Some works refer to the activation energy as the apparent barrier height.

3. The extracted activation energy is plotted as a function of gate voltage ($V_{GS}$), as shown in Figure 4.4(d). The source-to-body flat band voltage ($V_{FBSB}$) is identified using the linear dependence of $E_A$ and $V_{GS}$ in the thermionic regime ($V_{GS} < V_{FBSB}$).
which can be expressed as

\[ E_A = q\Phi_{Bn} - q\gamma(V_{GS} - V_{FBSB}), \quad (4.16) \]

where

\[ \gamma = \left(1 + \frac{C_{it}}{C_{ox}}\right)^{-1}, \quad (4.17) \]

with, \( C_{it} \) is the interface trap density and \( C_{ox} \) is the oxide capacitance [80]. Thus, the activation energy corresponding to \( V_{FBSB} \) represents the true (actual) barrier height, shown in the inset of Figure 4.4(d). For \( V_{GS} > V_{FBSB} \), the thermionic field emission and field emission components start to dominate which makes it incorrect to apply thermionic emission model.

Using the Arrhenius method, we extract a true barrier height of \( \Phi_{Bn} = 45 \) meV at \( V_{FBSB} = 1.5 \) V for this device.

**Self-consistency check on the extracted SBH**

The standard Arrhenius method, shown by first three steps of the flow chart shown in Figure 4.5(a), provides a straightforward way to extract Schottky barrier height in 2D transistors. However, the method lacks a self-consistency check between the measured current and the theoretical current predicted for the extracted SBH. In this section, we provide a simple self-consistent check as an additional step to the original flow chart to determine the same. We compare the room-temperature experimental \( (I_D(V_{FBSB})|_{exp}) \) and theoretical \( (I_D(V_{FBSB})|_{theo}) \) values of drain current at the extracted flat band voltage \( (V_{FBSB}) \). The theoretical drain current is calculated using the thermionic emission model given by Equation (4.15) for the extracted barrier height. If the calculated theoretical drain current is similar (a small variation can happen due to variation in the effective
electron mass assumed for the calculation) to the experimental drain current, then the extracted barrier height is self-consistent. Otherwise, the Arrhenius method provides an inaccurate barrier height for such devices. Figure 4.5(b) shows the experimental drain current at the extracted flat band voltages compared to the theoretical drain current calculated from Equation (4.15) as a function of SBH. For our device, the Arrhenius method results in a completely erroneous barrier height as the experimental drain current at $V_{FBSB}$ is $\sim 3000\times$ smaller than the respective theoretical value. The erroneous nature of Arrhenius method is further highlighted by the fact that the theoretical drain current at $V_{FBSB}$ is almost $\sim 500\times$ higher than the maximum current at measured at any gate voltage. Further, a more elaborate self-consistent check should also compare the $\gamma$ values extracted from the thermionic/subthreshold region of the transfer characteristics and slope of the linear region of the $E_A$ vs $V_{GS}$ curve. The disparity between the $\gamma$ values also bolsters our argument for erroneous extraction.

In Figure 4.5(b), we also show the SBH extracted using the Arrhenius method for other prominent works of various contact engineering to few-layer ($< 10$) MoS$_2$ devices. The black dots in Figure 4.5(b) represent the reported drain currents at the extracted $V_{FBSB}$ for these studies. Similar to the analysis of our device, we see a large discrepancy (typically at least 5-10000×) between the experimentally-measured and theoretically-calculated drain current values at $V_{FBSB}$. This shows that the erroneous SBH values predicted by the Arrhenius method are due to systematic and fundamental limitations of the Arrhenius method in extracting Schottky barrier height in 2D transistors.

In our analysis, there are three major reasons for failure of Arrhenius method for SBH extraction:

1. **Need for a clear transition from the thermionic regime to the tunneling regime:**

   Since the Arrhenius method depends upon proper identification of the flat band voltage, the device needs to show a clear transition from a thermionically dominated regime to a tunneling regime at all the measurement temperatures. However,
Figure 4.5: **Self-consistent check for the accuracy of extracted Schottky barrier height:**
(a) Flow chart for temperature dependent Schottky barrier extraction method. The first three steps are part of the standard/typical extraction method. A follow-up step involving comparison of theoretical and experimental drain current values is added to the flow chart to make the extraction self-consistent. (b) Comparison of drain current at extracted flat band voltages \( I_D(V_{FBSB}) \) for various studies involving few-layer MoS\(_2\) based transistors with the drain current predicted by thermionic emission model for extracted barrier heights. The model predicts \( \sim 3000 \times \) larger current at the extracted \( V_{FBSB} \) for our device. Similarly, the model also predicts significantly higher theoretical drain currents than the experimental values for other prominent contact engineering works.
this transition is often poorly defined in 2D devices due to the presence of non-idealities such as traps, non-homogeneous doping due to surface contaminants, and van der Waals gap [133, 134, 135]. Moreover, for doped contacts, devices with thick (> 2 nm) tunnel barriers, and few-layer (> 5) devices, the assumption of pure thermionic current is difficult to verify due to the high tunneling current arising from the channel region underneath the contact [136, 137].

2. **Weaker thermionic current at lower temperatures:** More often than not, the Arrhenius method for SBH extraction in 2D materials involves temperatures below 100 K. At such temperatures, the thermionic component is substantially smaller than the usual leakage floor for any considerable SBH ($\phi_{Bn} > 100$ meV). For example, a contact-dominated 2D FET with an SBH of 0.3 eV should result in a maximum thermionic current of 6 nA at flat-band condition at 300 K, which is reduced to less than 1 fA for $T < 77$ K. Thus, it is extremely difficult to measure any thermionic current at low temperatures below 100 K. This means that the currents observed at such temperatures usually come from TFE or FE components that show weak temperature dependence and therefore leading to erroneous SBH extraction [119].

3. **Role of channel region:** The Arrhenius method depend upon the accurate extraction of $V_{GS} = V_{FBSB}$ which fundamentally requires that the source contact dominates the current in the OFF state. This assumption leads to erroneous extraction of SBH in typical long-channel devices since it is the channel region which dominates the OFF state, as discussed in chapter 4. Figure 4.6(a) illustrates $L_{CH}$-dependent transfer characteristics of the MoS$_2$ device analyzed in the section 3.1.2 in chapter 4. To highlight the issue, we also plotted the transfer characteristics of corresponding source SBFET (shown with dashed line) and marked the ideal $V_{FBSB}$ (the gate voltage at which the current stops increasing exponentially with $V_{GS}$) and respective $I_{FBSB}$ ($I_D|_{V_{GS}=V_{FBSB}}$). For any channel length, the extracted source-body flatband
current ($I_{FBSB}^{ext}$) and source-body flatband voltage ($V_{FBSB}^{ext}$) differ significantly from the ideal value and become channel-length dependent. This is further illustrated in Figure 4.6(b) where the Arrhenius method result in large underestimation of SBH using Arrhenius method. To understand this underestimation, we express SBH in terms of $V_{FBSB}^{ext}$ as $SBH = \phi_{Bn} - q(V_{FBSB}^{ext} - V_{FBSB}(\text{ideal}))$. Our assumption of an ideal device allows us to do that. Using this, we extract $L_{CH}$-dependent SBH that varies from 290 to 40 meV for similar increase in channel length. Along with the Arrhenius method, the plot also shows the other known current-based method where the SBH is determined by identifying the $I_{FBSB}$. The current-based method of SBH extraction results in $\sim 60$ meV variation in extracted SBH values when the
channel length is varied from 0.01 to 10 µm. Thus, small channel-length (< 1 µm for 300 nm SiO$_2$) devices need to be fabricated to extract proper SBH values; preferably by the current-based method.

4.2 Mobility extraction

Figure 4.7: **Effective mobility and field-effect mobility.** (a) Effective mobility is extracted from the drain conductance near the origin of the output characteristics. For both mobility extraction techniques, both the transfer and output characteristics should show linear behavior around the extraction bias point. (b) Field-effect mobility extracted from the transconductance of a MOSFET, biased in the linear regime [114].

Two forms of mobility are typically extracted in 2D devices—MOSFET-based mobility and Hall mobility. Both extraction techniques have their pros and cons.

4.2.1 MOSFET mobility

MOSFET mobilities can be extracted from the measured transistor characteristics. MOSFET mobilities come in two flavors: effective mobility and field-effect mobility. Figure 4.7 illustrates the MoS$_2$-MOSFET characteristics employed to extract the effective and field-effect mobilities [138].

Effective mobility is extracted from the drain conductance of a MOSFET biased in the
linear regime. A general expression for the drain current of a MOSFET with a negligible diffusive current at small $V_{DS}$ can be written as

$$I_D = \frac{W}{L} \mu_{eff} Q_n V_{DS}, \text{ for } V_{GS} > V_{TH} \text{ and } V_{DS} \leq V_{GS} - V_{TH},$$

(4.18)

where, $Q_n = C_{ox}(V_{GS} - V_{TH})$ is the sheet charge density, $\mu_{eff}$ is the effective mobility, and $V_{TH}$ is the threshold voltage. Note that for thin gate oxides the relation for $Q_n$ needs to be modified to include the importance of quantum capacitance as

$$Q_n = \frac{C_{ox} + C_{S,Q}}{C_{ox} C_{S,Q}} (V_{GS} - V_{TH})$$

(4.19)

where, $C_{S,Q}$ is the degenerate quantum capacitance of the channel ($C_{S,Q} = q^2 DOS_{2D}$) discussed in Chapter 3 [67].

Ideally, $Q_n$ is determined through independent capacitance or Hall-effect measurements of the MOSFET structure; however, given the small size of many exfoliated samples, the capacitance of 2D MOSFETs is not typically measured as the signal is much too small and complex to reliably detect using conventional techniques. For an ideal device, effective mobility is then given by

$$\mu_{eff} = \frac{g_d L}{Q_n W}$$

(4.20)

where $g_d$ is the drain conductance given by $g_d = \frac{\partial I_D}{\partial V_{DS}} |_{constantV_{GS}}$, as shown in Figure 4.7(a). If the output characteristics do not exhibit a linear dependence on $V_{DS}$ around the bias point for which the mobility is extracted, the extracted mobility is suspect since the device characteristics do not follow Equation (16) from which $\mu_{eff}$ is derived. Similarly, if the transfer characteristics do not exhibit a linear dependence on $V_{GS}$ around the bias point for which the mobility is extracted, the use of the equation to determine $Q_n$ is highly suspect since the device behavior does not fit the charge model.
Field-effect mobility is derived from the transconductance \( g_m = \left. \frac{\partial I_D}{\partial V_{GS}} \right|_{constant V_{DS}} \) of a MOSFET biased in the linear regime as shown in Figure 4.7(b), which is generally given by

\[
\mu_{FE} = \frac{g_m}{C_{ox} V_{DS}} \frac{L}{W}.
\]  
(4.21)

where, \( \mu_{FE} \) is the field-effect mobility with value same as \( \mu_{eff} \) (\( \mu_{FE} = \mu_{eff} \)).

However, the derivation of Equation (4.21) assumes that \( \mu_{eff} \) does not change with gate voltage/electric field. When considering this dependence, the relation between \( \mu_{FE} \) and \( \mu_{eff} \) can be written as

\[
\mu_{FE} = \left( \mu_{eff} + (V_{GS} - V_{TH}) \frac{\partial \mu_{eff}}{\partial V_{GS}} \right).
\]  
(4.22)

For conventional MOSFETs, extracted \( \mu_{FE} \) is often less than the \( \mu_{eff} \) as \( \mu_{eff} \) decreases with increasing effective field and the measured transconductance is less than what would ideally be expected. Figure 4.8(a) shows the universal mobility curves for Si giving the different scattering mechanism as a function of effective electric field (\( \varepsilon_{eff} \)). If phonon and/or surface scattering dominates, \( \frac{\partial \mu_{eff}}{\partial V_{GS}} \) becomes negative and if Coulomb (impurity) scattering dominates, \( \frac{\partial \mu_{eff}}{\partial V_{GS}} \) becomes positive.

In bulk semiconductor, it is generally the phonon and surface scattering which dominates. However, this is not always true for 2D semiconductors as shown in Figure 4.8(b). Ma et al. calculated the electron mobility in monolayer MoS\(_2\) with 300nm SiO\(_2\) substrate as a function of carrier density (gate voltage) at different temperatures [67, 26]. A impurity/defect density (\( N_d \)) of \( 4 \times 10^{12} \) cm\(^{-2}\) was assumed in the calculations. Unlike Si, MoS\(_2\) and other TMDCs show impurity/defect limited mobilities at room temperature and thus show positive \( \frac{\partial \mu_{eff}}{\partial V_{GS}} \). Thus, the field-effect mobility is generally larger than the effective mobility in 2D materials. At high carrier densities, the impact of impurity scatterers are screened by the free electrons and thus \( \frac{\partial \mu_{eff}}{\partial V_{GS}} \) becomes nearly negligible and
Figure 4.8: **Relation between effective mobility and field-effect mobility.** (a) Typical universal $\mu_{\text{eff}}$ curves for Si for different doping densities ($N_{D_1} < N_{D_2} < N_{D_3}$). Coulomb (impurity) ($\mu_c$), phonon ($\mu_{\text{ph}}$), and surface scattering ($\mu_{\text{sr}}$) limited mobilities are also indicated [139]. (b) Calculated electron mobilities for monolayer MoS$_2$ as a function of carrier density (top panel) and gate voltage (lower panel) for different temperatures [67]. thus $\mu_{FE} \approx \mu_{\text{eff}}$.

### 4.2.2 Challenges due to contact resistance

Large contact resistance is a common problem in 2D devices that limits the accurate extraction of MOSFET mobilities, especially field-effect mobility ($\mu_{FE}$). To understand this, we rewrite Equation (4.18) to include the contribution due to contact resistance ($R_C$) as

$$I_D = \frac{W}{L} \mu_{\text{eff}} C_{\text{ox}} (V_{GS} - V_{TH}) (V_{DS} - I_D R_C).$$  \hspace{1cm} (4.23)

This allows us to use Equation (4.21) to understand the role of contact resistance on
\( \mu_{FE} \) as

\[
\mu_{FE} = \frac{L}{W C_{ox} V_{DS}} \frac{\partial I_D}{\partial V_{GS}} \tag{4.24}
\]

\( \frac{\partial I_D}{\partial V_{GS}} \) can then be evaluated using Equation (4.23) and can be expressed as

\[
\frac{\partial I_D}{\partial V_{GS}} = \frac{\kappa V_{DS} - \kappa^2 (V_{GS} - V_{TH})^2 V_{DS} \frac{\partial R_C}{\partial V_{GS}}}{(1 + \kappa R_C (V_{GS} - V_{TH}))^2}, \tag{4.25}
\]

where,

\[
\kappa = \mu_{eff} C_{ox} \frac{W}{L}. \tag{4.26}
\]

Thus, in the presence of gate-voltage dependent \( R_C \) the extracted field-effect mobility can be expressed as

\[
\mu_{FE}(R_C \text{ corrected}) = \frac{\mu_{FE} \left(1 - \kappa (V_{GS} - V_{TH})^2 \frac{\partial R_C}{\partial V_{GS}}\right)}{(1 + \kappa R_C (V_{GS} - V_{TH}))^2} \tag{4.27}
\]

Equation (4.27) indicates that neglecting \( R_C \) would result in underestimation of channel mobilities. Moreover, the mobility extraction from the contact limited devices can be problematic since \( V_{TH} \) is determined by the contacts rather than the channel region. In conventional MOSFETs, \( R_C \) is often determined from TLM structures and the extracted mobilities can be corrected for degradation due to \( R_C \). In principle, the same TLM can be applied to 2D MOSFETs; however, often large device-to-device variations make it difficult to achieve reliable and trustworthy results when applied to 2D materials as discussed in the previous section.

One way to circumvent the problem of \( R_C \) is to fabricate four-probe structures similar to those used for Hall effect measurements, shown in Figure 4.9(b) [108]. In such a structure, the voltage drop between the middle contacts is measured (\( V_{xx} \)), while \( V_{DS} \) is applied between the source and drain contacts. In this case, Equation (4.23) can be
modified to

\[ I_D = \frac{W}{L_{4p}} \mu_{eff} C_{ox} (V_{GS} - V_{TH}) V_{xx}, \]  

(4.28)

which results in accurate \( \mu_{eff} \) and \( \mu_{FE} \). Since \( V_{TH} \) is still determined by \( R_C \), it still results in slightly erroneous values if \( V_{TH} \) is determined using 2-probe transfer characteristics. In this light, a better method is to rewrite Equation (4.28) as

\[ I_D = \frac{W}{L_{4p}} \mu_{eff} C_{ox} (V_{GS} - V_{TH,4p}) V_{xx}(V_{GS}), \]  

(4.29)

From Equation (4.29), we can extract accurate channel mobility as

\[ \mu_{FE}(R_C \text{ corrected}) = \frac{L_{4p}}{W C_{ox}} \frac{\partial G_{4p}}{\partial V_{GS}}, \]  

(4.30)

where,

\[ G_{4p}(V_{GS}) = \frac{I_D(V_{GS})}{V_{xx}(V_{GS})}. \]  

(4.31)

We can directly determine \( G_{4p} \) using 4-probe measurements. Note that the measured potential across the channel may be quite small and perturbation of potential distribution due to the device geometry (e.g. size of voltage sensing probes). Nevertheless, due to the often large and variable \( R_C \) in 2D MOSFETs, four-probe measurement presents the best technique to accurately determine channel mobility with MOSFET mobility measurements.

4.2.3 Hall mobility

Hall effect measurements are widely carried out to extract the intrinsic material properties as it allows independent characterization of carrier density and mobility, without any direct effect of contact resistance. Figure 4.9(a) illustrates how an electron moves
in a conductive channel under applied longitudinal electric and perpendicular magnetic fields. The underlying principle of the Hall effect is based on the Lorentz force [140]. An electron flows (in the opposite direction to the current) along the channel in the presence of an electric field \(E_x\) with drift velocity \(v\). When a perpendicular magnetic field \(B_z\) is applied, the electron experiences Lorentz force, resulting in a voltage difference (Hall voltage, \(V_H\)) transverse to the flow of the electron. The sign of \(V_H\) depends on carrier type (electron or hole), and the value of \(V_H\) varies depending on the carrier density, current, and magnetic field.

The standard procedure to measure the \(\mu_H\) is to pattern the semiconductor into a Hall bar structure with contacts placed on the fingers, as shown in Figure 4.9(b). In the typical approach for measuring the \(\mu_H\) in 2D devices, a constant current is flowed between the source and drain contacts, while a magnetic field is applied normal to the plane of the semiconductor. Hall effect mobility measurements benefit from the independent extraction of the carrier concentration in the channel. In quasi-equilibrium, zero current flows along the width of the device.

Therefore, the total force along the width must be zero, satisfied when the Lorentz force is zero, which gives \(E_y = v_x B_z\), where \(x\) is along the length, \(y\) is along the width, and \(z\) is perpendicular to the 2D semiconductor channel. The general expression for current flow is given by \(I_D = q W v_x n_{2D}\). By defining the Hall voltage as \(V_H = E_y W\), we find that \(V_H = \frac{I_D B_z}{\mu_{2D}}\) and \(n_{2D} = \frac{I_D \Delta B_z}{q \Delta V_H}\). From the measurement of \(V_{xx}\) shown in figure 4(b), the \(R_{SH}\) of the channel can be determined by

\[
R_{SH} = \frac{V_{xx}}{I_D} \frac{W}{L_{4p}} \tag{4.32}
\]

Using \(R_{SH} = \frac{1}{q \mu_n n_{2D}}\), we find the Hall effect mobility to be

\[
\mu_H = \frac{V_H L_{4p}}{V_{xx} W B_z} \frac{1}{1} \tag{4.33}
\]
Figure 4.9: **Hall effect measurements of a bridge-type Hall bar structure.** (a) Illustration of the Hall effect of an electron. (b) Circuit configuration of a typical bridge-type Hall bar structure device. (c) $V_H$ versus B-field of a graphene device dependent on $V_{GS}$. (d) SdH effect in graphene showing oscillatory behavior of $\rho_{xx}$ and $\sigma_{xx}$ in the presence of B-fields [141].

where $\mu_n = \mu_H$ is assumed (which is only valid for a Hall scattering factor of 1). This assumption is further discussed in the following section. Note that since both $\mu_H$ and $\mu_{eff}$ depend upon the same drift current formulation, both should be give same value. However, since Equation (4.19) underestimates the carrier density in channel, $\mu_{eff}$ is often larger than $\mu_H$. Thus, for 2D semiconductors $\mu_{FE} > \mu_{eff} > \mu_H$ [142].

### 4.2.4 Challenges of Hall-effect measurements

In principle, the measurement of $\mu_H$ is straightforward, but in practice, several difficulties arise, complicating the measurement on 2D materials.

1. The first challenge is that the Hall effect measurement requires a specialized structure, ideally following the guidelines of ASTM Standard F76. The structure should
be designed such that the contacts lie as close to the edge of the sample as possible. The flakes can be etched into the desired geometry, but doing so has a negative consequence that the lithography and etch process may adversely decrease the mobility from its value in a pristine state. This is especially concerning for the mobility measurement of ultra-thin samples, where surface contamination can greatly affect the material’s mobility.

2. Another practical challenge for measuring Hall mobility in 2D materials is that $V_H$ can be quite small, making measurement difficult. $V_H$ is proportional to current per unit width, which is often less than $1 \mu A \mu m^{-1}$ for ultra-thin samples. $V_H$ can have an offset (i.e. $V_H \neq 0$ for $B=0$ given in Figure 4c) due to asymmetry in a Hall bar geometry so the difference in Hall voltage at different B-fields must be used instead of a single B-field measurement. A specialized probe station is typically required to obtain a large B-field, often involving the use of a cryostat with a cryogenic superconducting magnet. The AC Hall effect measurements, where a coil is used to generate the AC magnetic field, which is advantageous over DC measurement as it enables fast and low field measurements $< 0.1$ T, can also be used [143].

3. Although it is not often done for 2D materials, the sample (mostly graphene) can also be measured while placed atop a permanent magnet that is flipped between measurements to give a positive and negative B-field [144, 145]. Unfortunately, many back-gated devices that are pervasive across the 2D materials community show significant hysteresis [146, 147, 148] (or even worse, device degradation) from measurement to measurement, which makes the differential extraction between the positive and negative B-field measurements prone to hysteretic error. A solution to overcoming this problem is to perform repeated measurements, switching back and forth between $+B_z$ and $-B_z$, to verify that the data is stable.

4. Another, often overlooked, error in the measurement of $\mu_H$ arises from the assump-
tion of energy independent scattering in the semiconductor, which is generally only valid at very high magnetic fields ($\geq 1$ T) or for neutral impurity scattering. Energy dependent scattering is captured in the Hall scattering factor, $r = \frac{\langle r^2 \rangle}{\tau^2}$ ($1 < r < 2$), where $\tau$ is the mean time between carrier collisions and $\langle \tau \rangle$ is the average over energy. The Hall scattering factor can be determined at a specific B-field by $r = \frac{R_H(B_z)}{R_H(B_z=\infty)}$. Including this factor, the carrier concentration becomes

$$n_{2D} = \frac{r I_D B_z}{q V_H},$$

and the true electron mobility equals

$$\mu_n = \frac{\mu_H}{2}.$$  \hspace{1cm} (4.35)

Therefore, the $\mu_H$ can over-predict the conductivity mobility by up to a factor of 2.

All in all, Hall effect measurement is a powerful technique to measure carrier mobility in 2D materials; however, the technique is not without challenges and complications.

4.3 Summary

In this chapter, we looked at extraction techniques for two important device properties: (1) contact resistance ($R_C$) and (2) channel mobility ($\mu$). For $R_C$, we introduced the generally used TLM technique and understood its limitations when applied to 2D materials. Moreover, the difference between $R_{SK}$ and $R_{SH}$ leads to erroneous calculation of transfer lengths—$3\sim15\times$ larger than the ideal value. We also discussed the Arrhenius and current-based method of SBH extraction and showed that both the techniques underestimates the true SBH. For accurate extraction of SBH, devices with smaller channel lengths must be fabricated. Later, we discussed the different mobilities extracted in 2D transistors, namely, MOSFET mobilities and Hall mobilities. We analyzed the underes-
timation of channel mobilities in presence of large $R_C$ and proposed the use of TLM or four-probe measurements to extract $R_C$-corrected channel mobilities. A relation between $\mu_H$, $\mu_{eff}$, and $\mu_{FE}$ was determined for 2D materials where typically $\mu_H < \mu_{eff} < \mu_{FE}$. 
Chapter 5: Fabrication of High-Mobility 2D Devices Using Atomic Layer Etch

In the chapter, we demonstrate a novel etching method which allows fabrication of pristine/ultra-clean 2D devices with high channel mobilities. As discussed in Chapter 4, both mobility and contact resistance in 2D transistor is severely affected by disorder introduced during the fabrication process. In order to overcome that, we developed an atomic layer etching technique which allows controllable and selective removal of WSe$_2$ monolayers; thus allowing 1L WSe$_2$ to act like a sacrificial layer. This technique also enables reliable patterning, clean layer transfers, and high-quality ohmic contacts to 2D materials [149, 150]. Although etching of conventional bulk materials has demonstrated reliable sub-10-nm resolution, etching methods for two-dimensional (2D) materials are still in their infancy. Controllable etching of 2D materials has been a challenging task due to the presence of large surface-to-volume ratio (such that any surface damage readily impacts device performance), high structural similarity between different 2D materials, and lack of a suitable etch-stop layer [151].

To date, several etching techniques have been proposed to control the layer number and patterning of 2D materials. These techniques involve high-energy lasers, focused-ion beams, thermal/vacuum annealing, gaseous-phase surface treatments, and plasma-based etching. High-energy lasers and focused-ion beams have shown to be effective for controlled etching of MoS$_2$; however, these techniques are limited by their throughput and limited scalability [153, 154]. Thermal/vacuum annealing and gaseous-phase surface treatments offer a fast, repeatable, and wafer-scale etching process and have

---

1 This chapter is based on [152] and reproduced with the permission of ACS Publishing.
been use to etch various TMDCs such as MoS$_2$ and WSe$_2$ [155, 156]. However, detailed electrical and optical characterization of the etched flakes, required to support their suitability in device applications, has been lacking. Meanwhile, a wide range of plasma etching techniques has been developed for 2D materials due to their high throughput, low variability, and CMOS compatibility. However, most plasma-based layer-by-layer etching techniques are not self-limiting in nature due to associated physical etch and thus require careful optimization of several parameters such as plasma density, sample distance from plasma source, and etching time to achieve monolayer precision [157]. Furthermore, the inherent low selectivity associated with plasma-based methods makes it harder to etch one 2D material selective to another 2D material; a requirement useful for fabricating novel van der Waal heterostructures.

5.1 Introduction to atomic layer etch (ALE)

Our ALE process is a top-down approach that involves cyclical steps of self-limited oxidation and selective etch to achieve the desirable layer thickness. Our damage-free atomic layer etch that overcomes the aforementioned issues of existing techniques for etching 2D materials and thus provides high-quality etched flakes of deterministic thickness using comprehensive material, optical, and electrical characterization. Besides, our method employs a wafer-scale room-temperature process with the potential of having high throughput for use in back-end-of-the-line (BEOL) integration.

Furthermore, using graphene as a testbed, we demonstrate that our ALE process enables “etching and cleaning” of the top surface of the channel after device fabrication resulting in pristine high-quality 2D devices. Generally, high-quality 2D devices are fabricated using a dry/wet transfer of mechanically exfoliated or CVD-grown 2D materials on insulating hexagonal boron nitride ($h$-BN) [158]. However, these transfer processes are marred by large surface contamination, unintentional trapping of residue, and non-homogeneous topography (bubbles and blisters) which results in poor device
performance and large device-to-device variability [159]. Here, we use monolayer WSe$_2$ as a first-of-its-kind sacrificial 2D layer to protect the surface of a monolayer graphene transistor from any polymer residue during different fabrication steps. We finally remove the sacrificial WSe$_2$ layer with ALE processing (SWAP) technique and demonstrate high-quality graphene transistors with field-effect mobilities $3 \times$ than our control devices fabricated using direct polymer-based transfer.

5.1.1 Process flow

![ALE process flow and XPS characterization](image)

Figure 5.1: **ALE process flow and XPS characterization.** (a) The topmost layer of WSe$_2$ is removed by first oxidizing the flake using UV+O$_3$ for 30 minutes at room temperature, forming monolayer tungsten oxyselenide (TOS) that is subsequently removed with a mild 1M KOH dip for 10 s. The etching process selectively removes the oxidized topmost layer without damaging the underlying layers. (b) XPS spectra of the W (4f) peak for a 4L WSe$_2$ flake showing the appearance of W-O peaks after the UV-ozone process. The peaks corresponding to W-O bond formation disappears after the KOH treatment indicating the removal of the TOS layer.

As shown in Figure 5.1(a), our ALE process consists of two distinct steps.
1. **Self-limited oxidation of the topmost WSe$_2$ layer.** The sample is exposed to ozone (O$_3$) under ultraviolet (UV) light for 30 minutes at room temperature.

2. **Selective etching of the oxidized layer.** The sample is dipped in a 1M potassium hydroxide (KOH) solution for 10 seconds and then rinsed with DI water.

In the first step, the UV-ozone exposure converts the topmost WSe$_2$ layer into monolayer tungsten oxyselenide (TOS). The UV-ozone oxidation was performed in a SAMCO UV-2 system, a commercially available UV-Ozone cleaning/stripping tool. The process was performed at room temperature with an oxygen flow rate of 3 L/min. The oxidation process starts with the dissociation of O$_3$ into molecular oxygen (O$_2$) and singlet oxygen radical (O (1D)) in the presence of UV light. Then, the singlet oxygen radical (a strong oxidizing agent) reacts with the topmost WSe$_2$ layer to give monolayer TOS in an self-limited manner. This process is similar to that used by commercial UV-ozone cleaners to remove organic polymers from the surfaces of bulk semiconductors [160].

To confirm the presence of TOS, we performed X-ray photoelectron spectroscopy (XPS) on a 4L WSe$_2$ sample. The 4L WSe$_2$ flakes were stacked on a thick h-BN flake using the polycaprolactone (PCL) polymer-based dry transfer process [161]. Figure 5.1(b) shows the XPS spectra for the tungsten (W (4f)) core level for the same 4L WSe$_2$ sample in part (a) at each stage of the ALE process. After oxidation, the sample shows two additional weak doublet peaks in the W (4f) spectrum at higher binding energies. These higher energy doublet peaks correspond to formation of W-O bonds, similar to that of tungsten oxide (WO$_x$) observed in previous works [162]. Additionally, the presence of stronger doublet peaks corresponding to WSe$_2$ in the oxidized sample indicates that our oxidation process only affects the topmost layer. This is further confirmed by Raman and Photoluminescence (PL) measurements which are discussed in the subsequent sections. The slight red-shift in the binding energies of the WSe$_2$ peaks after oxidation can be attributed to the doping of the underlying layers due to presence of TOS layer on top whose utility will be discussed in Chapter 6 [60].
In the second step, we dip the oxidized samples in a 1M KOH solution to selectively remove the topmost TOS layer. KOH has been previously shown to selectively remove tungsten oxide (WO\textsubscript{x}) without etching the underlying WSe\textsubscript{2} layers [163]. We then rinse the etched samples with DI water to completely remove KOH traces from the surface. As detailed in the following section, the underlying layers are not affected by this mild etching process. XPS spectra, after the etching step, show that the doublet peaks corresponding to TOS layer are completely removed after KOH treatment while the peaks corresponding to WSe\textsubscript{2} peaks are restored to their pristine form (no shift in binding energy compared to as-exfoliated sample).

5.1.2 Monolayer precision of ALE process

To support our claim of monolayer precision for our ALE process, we investigated a WSe\textsubscript{2} flake exfoliated on a 285 nm SiO\textsubscript{2}/Si substrate with naturally abutting monolayer (1L) and bilayer (2L) regions. Figure 5.2(a) shows the optical and AFM images of the flake at each stage of the ALE process. The optical contrast of the entire flake changes significantly after oxidation such that the oxidized 1L region (labeled TOS) is nearly indistinguishable from the substrate and the oxidized 2L region (labeled 1L with TOS) is visually similar to 1L WSe\textsubscript{2}. The similarity in the optical contrast between the etched-to monolayer (bilayer etched to a monolayer) and the oxidized bilayer further confirms that only the top layer gets oxidized in our process. The corresponding AFM images for the oxidized samples show the presence of both bilayer and monolayer regions of the flake after oxidation, which confirms that the change in optical contrast is due to oxidation rather than etching of the top layer. This is also evident when we study the height (thickness) profiles where rather than a decrease, we observe an increase in thickness of both 1L and 2L regions after oxidation. Subsequently, the optical and corresponding AFM images also show that after the flake undergoes the etching step, the monolayer region is completely removed and the bilayer region is thinned down to monolayer with
Figure 5.2: **Monolayer precision of ALE process.** (a) Optical images (top), AFM scans (middle), and height profiles (bottom) of an exfoliated WSe$_2$ flake on SiO$_2$ at each stage of the ALE process. The as-exfoliated bilayer region (2L) is etched down to monolayer (1L) and the monolayer region is completely removed after the oxidation and etching step. (b) Raman spectra for the as-exfoliated, oxidized, and etched flake from (a). The Raman spectra were taken near the center of the 2L region of the flake. The removal of $B_{2g}^1$ peak (in the inset) corroborates the conversion from bilayer to a monolayer. (c) PL spectra are taken from the as-exfoliated bilayer region of the flake given in (a) at each stage of the ALE process. The PL spectra shows a clear transition from a broad, low intensity bilayer spectrum to a sharp, high-intensity monolayer spectrum, in agreement with our optical and AFM results.

no apparent change in the shape or size. Note that, the relative increase in height of etched-to monolayer to that of as-exfoliated monolayer can be attributed to the etching of neighbouring SiO$_2$ region. We further confirmed this by comparing the thickness of the
pristine and etched-to monolayers on an *h*-BN substrate as shown in Figure 5.3. Clearly, both pristine and etched-to flakes have similar thickness supporting our hypothesis.

Figure 5.3: **AFM height profiles of etched-to 1L WSe$_2$** (a,b) Color-corrected optical images of the pristine and etched-to 1L WSe$_2$ on *h*-BN. (c) Height profiles along A-A’ line cuts indicate similar heights (∼0.7 nm) for both the flakes. The pristine 2L used to obtain etched-to 1L is also shown as a reference.

Next, we performed Raman and photoluminescence characterization on the same flake to corroborate our optical and AFM results. Figure 5.2(b) shows the Raman spectrum of the as-exfoliated bilayer region at different steps of the ALE process. For the pristine bilayer (indicated in dark red), the dominant peak occurs at a Raman frequency of 249.42 cm$^{-1}$ which corresponds to the convoluted $A_{1g} + E_{2g}^1$ peak, where $A_{1g}$ represents out-of-plane and $E_{2g}^1$ represents in-plane vibrational modes. These modes are in-
distinguishable in our experiment since we have employed an unpolarized laser and can only be distinguished using curve fitting [164]. Along with this, a secondary peak occurs at 258.12 cm$^{-1}$, known as $2LA(M)$ peak, which represents the double resonance process involving two phonons from the longitudinal acoustic (LA) branch. Both $A_{1g} + E_{2g}^1$ and $2LA(M)$ peaks are present after oxidation, which confirms that the oxidation process only affects the top layer. The monolayer precision of our ALE process can be further corroborated by looking at the Raman mode ($B_{1g}^1$), which is active only in few-layer WSe$_2$ [165]. As shown in the inset, the $B_{1g}^1$ occurs at 308.32 cm$^{-1}$ for the as-exfoliated 2L flake but disappears after the oxidation step. Finally, both $A_{1g} + E_{2g}^1$ and $2LA(M)$ peaks remain strong after etching with 1.74 cm$^{-1}$ blueshift in the $2LA(M)$ peak. The strong blueshift of the $2LA(M)$ peak is consistent with bilayer to monolayer transition. This confirms that the oxidation step only affects the topmost layer and the etching step selectively removes the top TOS layer.

Figure 5.2(c) shows the PL spectrum for the same flake at each stage of the ALE process. Here, we observe a clear transition from a low intensity, double peak spectrum for the as-exfoliated bilayer region to a high intensity single peak spectrum (10× the peak intensity of original bilayer sample) after etching. This increase in intensity and the peak shift towards higher energy suggest a transition from indirect (bilayer) to direct bandgap (etched-to monolayer), confirming the monolayer precision of our ALE process. Meanwhile, the PL spectrum for oxidized bilayer looks similar to that of the etched-to monolayer; though with a large redshift (45 meV) that can be attributed to hole-rich trion formation [166]. Furthermore, we carried out our ALE technique on other TMDCs such as MoSe$_2$ where we could again achieve clean, damage-free etch; however, it does not with sulfide-based TMDCs (MoS$_2$ and WS$_2$) due to non-uniform oxidation as shown in Figure 5.4. Further theoretical/experimental studies are required to understand the difference between the oxygen passivating strength (barrier strength to oxygen diffusion) of transition metal sulfides and selenides. Thus, our PL characterization corroborates our
Figure 5.4: Application of ALE process to other TMDCs Optical images of MoSe$_2$, MoS$_2$, and WS$_2$ flakes processed using the ALE method. In the case of MoSe$_2$, the 1L and 2L regions are uniformly oxidized and removed by the ALE process leaving behind an etched-to 1L MoSe$_2$ with no apparent damage. However, the oxidation process is not self-limiting for MoS$_2$ and WS$_2$ samples possibly due to different diffusion barrier strengths of sulfur-based TMDCs.

optical, AFM, and Raman measurements and show that our ALE process removes precisely one layer per cycle.

5.1.3 Repeatability and patternability of ALE technique

The ALE process can be repeated in a cyclical manner to remove the desired number of layers of WSe$_2$ as shown in Figure 5.5. Exactly one layer is removed per ALE cycle with minimal damage (if any) to the etched layers as shown by the Raman and PL characterization. Furthermore, we also demonstrate that our ALE process is area-selective by removing WSe$_2$ layers from a particular section of a few-layer flake (Figure 5.6) by sim-
Figure 5.5: **Repeatability of the ALE process.** (a) Optical images show clear demonstration of repeatable ALE process. In the first ALE cycle, the (n) layer section becomes (n-1) layer, e.g., 3L region become 2L, 2L region become 1L, and 1L region is completely removed. The same process is repeated twice to completely remove the 3L region with clear change in optical contrast of thick area illustrating layer-by-layer etching of all the regions. No visible damage was observed on the thick or thin parts of the flake during the multiple runs of the ALE process. (b) Normalized Raman spectra of the different monolayers obtained from multiple ALE cycles showing no change in crystalline structure of WSe$_2$ with ALE processing. (c) Normalized photoluminescence (PL) spectra of the same monolayers illustrating slight increase (∼20 meV) in sample inhomogeneity with multiple ALE cycles possibly due to small size of the etched-to flakes, however no defects peaks were observed in the PL spectra.

ploy using a polymethyl methacrylate (PMMA) mask. This shows that our ALE process is patternable and can be used to selectively remove WSe$_2$ layers from a particular area to create novel 2D devices.
5.2 Detailed characterization of ALE-processed flakes

Having established the monolayer precision of the ALE process, we then characterized optical, electrical, and material properties of etched-to flakes (also referred as ALE-processed flakes) and compared them to pristine (as-exfoliated) flakes of similar thickness. The quality of ALE-processed layers needs to be comparable to exfoliated or CVD-grown layers for use in practical 2D devices.

5.2.1 Material characterization

The optical images in Figure 5.7(a)(i) shows clear thinning of the bilayer region to monolayer after the ALE process along with the complete removal of the pristine monolayer region. The corresponding high-resolution AFM images, shown in Figure 5.7(a)(ii), further support the premise of atomic layer etching. The AFM image of the etched flake clearly shows that the ALE process is highly uniform with no apparent damage. This is different to previous oxidation-enabled etching studies where large etch pits were formed on the flakes after the etching process [167, 168]. Note that the slight increase in
Figure 5.7: **Comparison of pristine and etched-to 1L WSe$_2$ flakes.** (a) (i). Optical images of WSe$_2$ flake confirming the monolayer etch. The as-exfoliated bilayer (2L) region is etched down to monolayer (1L) and the monolayer region is completely removed. (ii). AFM characterization confirms the atomic layer etch with no apparent change in shape, size, or topography of the flake. RMS surface roughness ($R_q$) of the etched-to 1L is similar to that of the as-exfoliated monolayer indicating that the ALE process minimally affects the underlying layers. (b) Raman spectrum of the etched-to 1L WSe$_2$ is similar to that of as-exfoliated 1L WSe$_2$. No additional defect peaks were observed in the etched-to 1L WSe$_2$ with negligible change in the peak positions confirming the crystalline nature of the etched layers. (c) STEM images of as-exfoliated and etched-to 1L WSe$_2$ taken at the same scale. The etched flake does not show any noticeable damage or defects when compared to pristine flake. The electron diffraction patterns also indicate a single-crystalline structure of the etched-to 1L flake with similar angle and distance between adjacent planes as that of pristine 1L WSe$_2$, indicated by arrows in the inset.

($\sim 1$ Å) in the root mean square value of surface roughness ($R_q$) in the etched-to monolayer, can also be attributed to substrate etching due to intercalated KOH as long term KOH exposure leads to flake liftoff [169, 170].

Figure 5.7(b) compares the normalized Raman spectrum of the etched-to monolayer WSe$_2$ with that of the monolayer region of as-exfoliated flake. No additional peaks were observed in the etched-to monolayer, indicating pure hexagonal symmetry and lack of macroscopic defects in the ALE-processed flakes [171]. We used Gaussian-Lorentzian
fits to deconvolve the curve into three distinct peaks, identified as $A_{1g}$ (cyan), $E_{2g}^1$ (pink), and $2LA(M)$ (green). Compared to pristine monolayer, all three peaks $A_{1g}$, $E_{2g}^1$, and $2LA(M)$ are slightly blueshifted by 0.14, 0.14, and 0.31 cm$^{-1}$, respectively. These small blueshifts are likely due to minimal residual p-type doping and sample inhomogeneity [172]. We further performed scanning transmission electron microscopy (STEM) and electron diffraction studies of etched-to and pristine 1L WSe$_2$ samples, as shown in Figure 5.7(c). The STEM characterization bolsters our Raman measurements by showing single-crystalline nature of etched-to 1L flakes without any discernible defects or damage. Moreover, the diffraction patterns (shown in the inset) of the etched-to and pristine 1L flake show a similarity in the distance and angle between the adjacent planes suggesting similar hexagonal symmetry. Thus, the Raman spectrum and STEM characterization of etched-to monolayers appear similar to that of pristine monolayers. No additional defect peaks are observed, which confirms that the ALE process maintains the quality of the etched-to monolayers.

5.2.2 Optical characterization

Next, we performed room and low-temperature PL characterization of ALE-processed flakes to compare the optical quality of etched-to (2L to 1L) monolayers with pristine flakes. The top panel of Figure 5.8(a) shows the optical image of an exfoliated WSe$_2$ flake with adjacent monolayer and bilayer regions. The optical images clearly show the thinning of bilayer region to monolayer, which is further supported by the PL spectra given in the lower panel that shows a large increase in PL intensity after etching due to bilayer to monolayer conversion. We also performed hyperspectral mapping of the entire flake, before and after etching, to determine the uniformity of the etch which shows less than 20 meV variation in the PL spectral median over the entire etched flake [173]. The PL intensity variation in the etched-to flake is similar to that of the as-exfoliated monolayer flake.

After confirming the atomic layer etch, we compared the PL spectra of as-exfoliated
Figure 5.8: **Optical characterization of etched-to 1L WSe\(_2\).** (a) Optical (gray-scaled) images of pristine WSe\(_2\) flake (before and after etch) show that only one layer is removed per ALE cycle. Corresponding PL spectra confirm this with a clear transition from a broad low-intensity bilayer WSe\(_2\) spectrum to a sharp monolayer spectrum for etched-to 1L WSe\(_2\) after the ALE process. The etched-to 1L WSe\(_2\) flake depicts PL characteristics similar to the pristine 1L WSe\(_2\) flake with a 0.55\(\times\) reduction in the peak PL intensity. (b) Deconvolved room-temperature PL spectra of the pristine and etched-to 1L WSe\(_2\) from (a). The PL spectrum for the etched-to 1L WSe\(_2\) shows the same number of deconvolved peaks (exciton peak \(X^0\) and hole trion peak \(X^+\)) and similar \(X^+ / X^0\) ratio. (c) Integrated PL intensity \(I\) for as-exfoliated and etched-to monolayers for different pump laser powers \(P\). Both as-exfoliated and etched-to flakes show similar sub-linear behavior. (d) Time-resolved PL decay also corroborates the similar optical characteristics of pristine and etched-1L WSe\(_2\) flakes with comparable PL dynamics and decay time constants \(\tau\).
a much larger linewidth of 70 meV that is attributed to positively-charged, hole-rich tri-
ons ($X^+$) (two holes and one electron) [178]. Compared to the as-exfoliated monolayer,
we see no apparent change in the intensity of the $X^0$ peak except a small redshift (12
meV) and increase in linewidth (9 meV). This increase in linewidth can be attributed
to several factors such as exciton-defect scattering, sample inhomogeneity, and residual
doping [179, 180, 181]. To understand the role of residual doping, we adopted the study
from Mccreary et al. where the shoulder to main peak intensity ratio was presented
as an important quantitative parameter to determine the pristine nature of any 2D ma-
terial [182]. We observe a slight increase in $X^+/X^0$ ratio (0.29 to 0.33) in our etched
monolayer indicating minimal residual doping; however, substantially better than pre-
vious works where this ratio almost doubles after etching [183]. To rule out the role
of exciton-defect scattering, we performed temperature-dependent PL measurements on
pristine and etched-to flakes, provided in Figure 5.9. Since exciton-defect scattering be-
comes much more apparent at lower temperatures, the lack of any defect peaks at low
temperatures ($\sim$77 K), further supporting our claim of defect-free etch. Moreover, the
difference between the excitonic-component linewidth for pristine and etched-to mono-
layer reduces with lowering of temperature which indicates slightly larger sample in-
homogeneity in etched-to monolayers which could lead to larger excitonic linewidth.

Furthermore, we performed power-dependent and time-resolved PL characterization
on the same flake to obtain insights into the dynamic behavior of PL characteristics in
the etched-to layer compared to that of the pristine WSe$_2$ layer. Figure 5.8(c) shows a
log-log plot of the integrated PL intensity ($I$) as a function of laser power ($P$) for both
as-exfoliated and etched-to monolayers. These logarithmic plots can be described by the
power-law equation, i.e, $I \propto P^\alpha$, where $\alpha$ represents the linearity factor [184, 185]. For
exciton-dominant ($X^0$) PL, $\alpha$ is close to unity at low excitation power ($<1.5 \mu$W) due to
radiative recombination processes governed by a first-order rate equation [186]. How-
Figure 5.9: Temperature-dependent PL characterization of etched-to 1L WSe₂ (a) PL spectra of as-exfoliated (pristine) and ALE-processed (etched-to) monolayer WSe₂ flakes measured at 77K. The ALE-processed flake has similar PL spectrum as pristine flake with no discernible defect peaks. (b) Excitonic linewidth as a function of temperature show a clear reduction in the linewidth and linewidth difference between the pristine and etched-to monolayers with lowering the temperature, indicating reduced sample inhomogeneity.

Ever, in presence of defects, the linearity factor deviates from this ideal value and thus become either highly sublinear ($\alpha = 0.3$) or superlinear ($\alpha = 1.45$), as shown in previous studies [187, 188]. Here, we extracted $\alpha = 0.72$ for the etched-to monolayers which is similar to that of pristine monolayer ($\alpha = 0.84$) indicating the similarity between the two. As discussed before, the slight reduction in the $\alpha$ value can also be attributed to higher sample inhomogeneity and trion (hole doping) concentration in the etched-to monolayer compared to pristine monolayer [189]. Thus, we do not observe significant reduction or increase in the $\alpha$ value as shown in previous defect studies, further confirming the nearly
defect-free nature of the etched-to monolayers.

Figure 5.8(d) shows the normalized time-resolved PL decay for both as-exfoliated and etched-to monolayers. Here, IRF stands for instrument response function which indicates the limits of our measurement system (∼20 ps). We used Gaussian-biexponential fits to extract fast ($\tau_1$) and slow ($\tau_2$) decay time constants [190]. Here, the fast decay time is associated with the dynamics of bright exciton, and the slow decay time is associated with dark exciton. The fast decay constant decreases slightly from 75 ps in the pristine monolayer to 65 ps in etched-to monolayers suggesting similar exciton recombination dynamics in both. Since we do not see any discernible defects in temperature-dependent PL, the slight decrease once again indicates slightly larger sample inhomogeneity and trion density in the etched-to monolayer. Thus, our detailed PL characterization supports our claim that the optical (PL) properties (both static and dynamic) of ALE-processed flakes are comparable to pristine WSe$_2$ layers.

5.2.3 Electrical characterization

Next, we investigated the electrical properties of ALE-processed WSe$_2$ flakes by fabricating an ideal p-type transistor [191]. For this, we fabricated a back-gated field-effect transistor on an ALE-processed trilayer (3L) WSe$_2$ flake, back-gated through a ∼60 nm $h$-BN/285 nm SiO$_2$ dielectric stack, as shown in Figure 5.10(a). The 3L WSe$_2$ was obtained from an as-exfoliated 5L flake by performing two cycles of ALE process. We chose 3L WSe$_2$ for our study as (i) thick flakes (>10 layer) are largely unaffected by etch processes since the majority of current flow occurs in the bottom layers and (ii) 1L/2L WSe$_2$ devices are limited by high contact resistance resulting in mainly n-type transport even with high work function metals [114, 74]. To obtain high-quality contacts, we used a newly developed technique called transferred via contacts consisting of platinum (Pt) embedded in $h$-BN vias patterned in a Hall-bar structure [29, 192]. Figure 5.10(b) shows the output characteristics of the device for low $V_{DS}$ indicating linear behavior at high gate voltages.
Figure 5.10: Electrical properties of ALE-processed WSe$_2$ flakes. (a) Enhanced optical image of a back-gated 2D device made from an etched-to 3L WSe$_2$ flake with transferred via contacts made from Pt in a Hall-bar pattern. The WSe$_2$ was exfoliated as a 5L flake and etched-to 3L using two cycles of the ALE process. (b) Output characteristics show linear behavior at low $V_{DS}$ for high gate voltages. (c) Transfer characteristics show good p-type conduction with high ON-current, low hysteresis, and large ON/OFF ratio. (d) Temperature-dependent transfer characteristics showing the linear region of device operation for $V_{GS} < -80$ V. (e) Extracted 2-probe and 4-probe field-effect mobility ($\mu_{FE}$) at $V_{GS} = -95$ V at different temperatures. High 4-probe hole mobility of 515 cm$^2$/V·s at room-temperature denotes the high-quality electrical properties of ALE-processed WSe$_2$ flakes.

which is useful for extracting accurate field-effect mobilities as discussed in Chapter 4. The transfer characteristics, shown in Figure 5.10(c), depict high ON-current ($\sim 10$ µA at $V_{DS} = -500$ mV), excellent ON/OFF ratio ($10^5$ — limited by the 100-pA leakage floor of our measurement setup), good sub-threshold swing, and small hysteresis. Overall, the transfer characteristics highlight the excellent switching characteristics of the etched-to 3L WSe$_2$ layers.

To further quantify the electrical properties of the ALE-processed WSe$_2$, we extracted
temperature-dependent field-effect mobility ($\mu_{FE}$) using four-probe measurements, as shown in the lower panel of Figure 5.10(a). Figure 5.10(d) shows the temperature-dependent transfer characteristics for our device at $V_{DS} = -500$ mV from which we extracted 2-probe and 4-probe conductances. We then extracted the field-effect mobilities from the respective 2-probe and 4-probe transconductances at $V_{GS} = -95$ V, corresponding to the linear region of device operation. For our 3L WSe$_2$ device, room-temperature 4-probe hole mobility of 515 cm$^2$/V·s at $V_{GS} = -95$ V corresponding to $p_{2D} = 1.2 \times 10^{12}$ cm$^{-2}$ was extracted which is the highest reported hole mobility in few (1~5)-layer WSe$_2$ devices.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Layer Number</th>
<th>Gate design</th>
<th>Extraction technique</th>
<th>Hole mobility (cm$^2$/V·s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fang et al.[193]</td>
<td>1L</td>
<td>top gated</td>
<td>2-probe</td>
<td>250</td>
</tr>
<tr>
<td>Pradhan et al. [194]</td>
<td>4L</td>
<td>back gated</td>
<td>4-probe</td>
<td>200</td>
</tr>
<tr>
<td>Movva et al. [191]</td>
<td>4L</td>
<td>top gated</td>
<td>4-probe</td>
<td>140</td>
</tr>
<tr>
<td>Wang et al. [195]</td>
<td>5L</td>
<td>back gated</td>
<td>4-probe</td>
<td>300</td>
</tr>
<tr>
<td>Jung et al. [29]</td>
<td>2L</td>
<td>back gated</td>
<td>4-probe</td>
<td>195</td>
</tr>
<tr>
<td>This work</td>
<td>3L</td>
<td>back gated</td>
<td>4-probe</td>
<td>240~515</td>
</tr>
</tbody>
</table>

Table 5.1: Room-temperature hole mobilities in p-FETs fabricated on pristine WSe$_2$ flakes.

Table 5.1 shows the comparison of field-effect mobilities of our device with several other reported high-quality WSe$_2$ devices which shows that our ALE-processed flakes show similar or higher mobilities than pristine WSe$_2$ based devices. Overall, our electrical characterization demonstrates high-electrical quality of the ALE-processed WSe$_2$ flakes.

5.3 Graphene devices using sacrificial WSe$_2$ with ALE processing

Having demonstrated the monolayer precision of our ALE method, we utilize the ALE process to provide a path for clean processing of 2D materials to enable high-performance 2D devices with large throughput. For enabling this, we use monolayer WSe$_2$ as a sacrificial layer between the pick-up polymer and the flake of interest. The
Figure 5.11: Sacrificial WSe₂ with ALE Processing (SWAP) technique for high-quality 2D devices. (a) Schematic and process flow of a graphene device made with direct PCL transfer (control device) and using SWAP technique. The ALE process removes the sacrificial WSe₂ layer along with the polymer residue accumulated in the transfer and lithography steps. (b) Raman measurements show the relative cleanliness of the samples. Raman spectra of the control device show a broadened 2D peak with a small $I_{2D}/I_G$ ratio (2.2) indicating the presence of polymer residue on top of the graphene. Raman spectra for the direct transfer device with WSe₂ sacrificial layer has sharper 2D peak, compared to that of control device. The subsequent ALE process removes of the top WSe₂ layer along with the polymer residue and leads to a clean graphene surface indicated by the large $I_{2D}/I_G$ ratio suggesting minimal damage to the underlying graphene layer during the ALE process. (c) Comparison of sheet resistance ($R_{sh}$) for graphene devices made using direct transfer and SWAP technique showing better electrical characteristics of the SWAP device with Dirac peak close to 0 V, 2.5× reduction in $R_{sh}$ at corresponding Dirac peaks, and 3× increase in effective mobilities extracted using the Kim’s model (dark grey lines). (d) Room-temperature 4-probe field-effect mobility ($\mu_{FE}$) of the device made using SWAP technique shows minimum 3× improvement at any carrier density ($n_{2D}$) over the devices fabricated with direct PCL transfer. The SWAP enabled device shows much better mobility from an unencapsulated graphene device.[196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 78, 206]

monolayer WSe₂ is subsequently etched after the transfer process, leaving the desired material of interest in its pristine form without any polymer residue. We illustrate the advantages of our ALE process in the clean fabrication of 2D devices by fabricating
two sets of graphene devices, one without WSe$_2$ (the control device) and another with the SWAP technique. The control device was made using standard PCL transfer where graphene was picked up directly by PCL and transferred onto bottom $h$-BN. Although this process is much faster than other van der Waals based dry transfer processes giving higher throughput, it results in significant polymer residue on the transferred layer that is hard to remove even with best cleaning techniques such as vacuum annealing and AFM cleaning [207, 208]. Moreover, subsequent device processing, such as channel patterning and contact formation, leave additional resist and solvent residues that contaminates the channel. Figure 5.11(a) shows a schematic of the control device made using direct PCL transfer indicating the presence of all kinds of residue on top of the graphene layer. Raman measurements, shown in Figure 5.11(b), confirm the presence of significant residue (even after vacuum annealing) as indicated by a small cleanliness ratio ($I_{2D}/I_G$) of 2.2, similar to that of other unencapsulated devices made on graphene on top of $h$-BN [209]. In our SWAP method, we first pick-up monolayer WSe$_2$ with PCL polymer and then pick up monolayer graphene using the PCL/monolayer WSe$_2$ stack, which is finally transferred on bottom $h$-BN. Subsequently, the Gr/WSe$_2$ stack is patterned in a Hall-bar geometry and edge contacts are formed using the same etch mask. Thus, the graphene remains in its pristine form and is not exposed to polymer along with clean patterning using standard lithographic techniques. Raman measurements of the device with WSe$_2$ on top of graphene show smaller full-width half maxima (FWHM) of the $I_{2D}$ peak compared to that of direct-transfer device indicating a clean graphene-WSe$_2$ interface. Next, we use the ALE process to remove the top WSe$_2$ layer without altering the properties of underlying graphene. *Only the top surface of the monolayer WSe$_2$ is exposed during device processing, so the ALE process removes this contamination along with the WSe$_2$ monolayer, leaving behind a pristine graphene surface.* Raman measurements show that the $I_{2D}/I_G$ ratio increases by 60% after removal of WSe$_2$ (2.1 to 3.5) and becomes comparable to that of previously calculated values for pristine graphene (3~4) demonstrating the advantages.
of SWAP technique [210].

Next, we performed electrical characterization to corroborate our claim of making high-quality graphene devices using our SWAP technique. Figure 5.11(c) compares the 4-probe electrical characteristics of our SWAP device with the control sample, i.e. the direct-transfer device. As discussed earlier, 4-probe characterization enables a comparison of the intrinsic properties of the channel layer without the effect of contact resistance. The direct-transfer device shows a large shift in the Dirac voltage \( V_D \) towards negative gate voltage \( V_D = -50 \) V along with a large sheet resistance \( R_{sh} \) away from the peak. This is also consistent with previous experimental works that show a large \( V_D \) shift and high \( R_{sh} \) values due to reduced carrier mobility and large residual doping due polymer, photoresist, and solvent residue [211, 212]. In contrast, the SWAP device shows nearly pristine characteristics with the Dirac point at \( V_{GS} \approx 0 \) V and a \( 2.5 \times \) reduction in \( R_{sh} \) at corresponding Dirac voltage \( V_D \). We model the \( R_{sh} \) vs \( V_{GS} \) characteristics using the effective mobility method \( (\mu_{eff}) \) provided by Kim et. al., indicated by the corresponding solid (dark grey) lines [213]. In this method, the gate only controls the carrier density in the device and thus the sheet resistance can be modelled as \( R_{sh} = \left( q(n_{eff}(V_{GS})\mu_{eff}) \right)^{-1} \), where \( q \) is the electron charge. The effective carrier density can be given as \( n_{eff}(V_{GS}) = \sqrt{(n(V_{GS} - V_D))^2 + (n_0)^2} \), where \( n_0 \) is the residual doping density at Dirac voltage. Using this method, we extract an \( \mu_{eff} = 8700 \) cm²/V·s and \( n_0 = 6.23 \times 10^{11} \) cm⁻² for the SWAP device along with an \( \mu_{eff} = 3100 \) cm²/V·s and \( n_0 = 8.73 \times 10^{11} \) cm⁻² for the direct-transfer device. Clearly, the SWAP device provides much better mobilities (\( \sim 3 \times \)) and lower residual doping (\( \sim 30\% \)) than the direct-transfer device, elucidating the importance of clean transfer and device patterning.

To further quantify this improvement, we extracted the room-temperature field-effect mobility \( (\mu_{FE}) \) for the direct-transfer and SWAP device using the technique described in previous graphene studies. For extracting four(4)-probe field-effect mobility in graphene,
we first extract the 4-probe resistance using the relation

\[ R_{4p} = V_{4p} / I_D. \]  

(5.1)

Here, \( V_{4p} \) represents the potential drop across the side Hall-bar potential probes on the device. We then extract the gate-induced electron carrier density \( n_{2D} \) in the graphene device using the expressions

\[ n_{2D} = \frac{1}{q} C_{ox} (V_{GS} - V_D) \] for \( V_{GS} > V_D \), 

(5.2)

where \( V_D \) refers to the Dirac voltage corresponding to the maximum resistivity. Finally, 4-probe electron field-effect mobility is determined by using the relations

\[ \mu_{FE} = \frac{L_{4p} \times 1}{W \times qn_{2D}R_{4p}}. \] 

(5.3)

As shown in Figure 5.11(d), the SWAP device illustrates a 3× increase in the field-effect mobility at any carrier density \( n_{2D} \) over the direct-transfer device, similar to the increase in \( \mu_{eff} \), thus highlighting the advantages of the SWAP technique. To compare the SWAP enabled device with published results, we plotted the maximum field-effect mobility extracted for graphene devices made using different cleaning and transfer techniques. Our SWAP technique provides extremely high mobilities (~200,000 cm²/V·s), higher than any other unencapsulated device and below only to fully h-BN encapsulated graphene devices. Even though fully-encapsulated graphene devices show better intrinsic characteristics, the fabrication method is extremely slow and also cannot be used for fabricating devices for sensing applications and near-field measurements that require direct access to the surface of the devices. Thus, our SWAP technique provides a faster, cleaner, and reliable method to fabricate high-quality 2D devices with no discernible degradation of pristine properties.

136
5.4 Summary

In this chapter, we present a clean, selective, and repeatable atomic layer etch for achieving high-quality WSe₂ flakes with precise layer control to optimize different regions—for example, thick at the contacts to reduce parasitic resistance and monolayer in the channel to exploit its direct band gap for optoelectronic applications. Several etching techniques have been demonstrated in the past, but they result in sub-par flakes with low optical and electrical quality rendering them useless for most device applications. Using a comprehensive set of optical, material, and electrical characterization techniques, we provided a detailed characterization of the ALE-processed WSe₂ layers and benchmarked them against pristine (as-exfoliated) flakes. We further use our ALE technique to demonstrate the fabrication of pristine high-quality 2D devices using WSe₂ as a 2D-sacrificial layer. We show that graphene transistors protected by a sacrificial monolayer of WSe₂, that is subsequently removed using our ALE process, have higher mobility than any other unencapsulated graphene device. Thus, our ALE process provides a path towards a controllable, and selective etch that can enable integration of 2D materials in CMOS technology for future applications.
6.1 Introduction

In this chapter, we introduce monolayer tungsten oxyselenide (TOS) as a tunable high-quality p-type surface-layer dopant. Monolayer TOS is formed by the same UV-ozone oxidation process discussed in the previous chapter on atomic layer etch. Since our oxidation process is self-limiting to monolayer limit and thus only affects the top \( \text{WSe}_2 \); the process allows direct processing of 2D homo- & hetero-structures while providing high mobilities at high carrier densities. This makes our method uniquely different from previous oxidation-based doping works [214, 215]. Unlike the traditional substitutional dopants, monolayer TOS does not require a high-temperature processing step for dopant activation and is therefore suitable for a variety of diverse semiconductor materials. In contrast to the substitutional doping, our TOS doping does not modify the constituent atoms of the crystal, allowing researchers to study its innate properties and low temperature characterizations, albeit at a much larger hole density. Here, we used graphene as a platform to investigate the doping properties of monolayer TOS. However, our doping technique can also be extended to a diverse array of multi-dimensional (1D to 3D) semiconducting materials such as \( \text{WSe}_2 \), single walled carbon nanotubes (SWCNTs), and organic semiconductor (DNTT — Dinaphtho[2,3-b:2',3'-f]thieno[3,2-b]thiophene) as shown in the following sections.

Figure 6.1(a) shows the process for TOS doping of a monolayer graphene device. The device is fabricated by first stacking 1L \( \text{WSe}_2 \), graphene, and bottom \( h \)-BN using a poly-

---

(i) This chapter is based on [152] and reproduced with the permission of ACS Publishing.

(ii) This chapter is also based on the paper submitted to Nature Electronics and reproduced with the permission of NPG Publishing.
caprolactone (PCL) polymer based dry transfer process [161]. The device is then etched into a Hall-bar pattern to allow accurate four-terminal resistivity and Hall-effect measurements. Finally, edge-contacted metal (Cr/Au) electrodes were formed using electron beam lithography and e-beam evaporation as shown in optical image of Figure 6.1(b) [216]. The monolayer WSe$_2$ is then converted into monolayer TOS by 30 minutes room-temperature UV-ozone oxidation process, without affecting the underlying graphene layer. We performed Raman characterization of the as-fabricated (gray) and doped (dark red) graphene device to qualitatively study the doping effects as shown in Figure 6.1(c). After doping, both G and 2D peaks show a clear blueshift (18.1 and 4.3 cm$^{-1}$, respectively) from their original values which indicates p-type doping in graphene [217, 218, 219]. Here, the $h$-BN peak was used as a reference for calculating peak shifts as it does not change with doping. Moreover, the $I_{2D}/I_G$ ratio shows a reduction from 3.0 to 1.1 and full-width half maxima (FWHM) of the G peak reduces from 7.3 to 5.9 cm$^{-1}$ after oxidation as shown in the insets, confirming the doping effects [217].

6.1.1 Formation of monolayer TOS

To further elucidate the doping effect, we used four-terminal resistance ($R_{4p}$) measurement of the graphene device at room temperature (RT), shown in Figure 6.1(d). For the as-fabricated device (before the formation of the TOS layer), the Dirac point ($V_D$) is at back-gate bias ($V_{GS}$) = +30 V; while it completely disappears from the measurement range after oxidation doping indicating an ultrahigh carrier density in graphene. The $R_{4p}$ decreases at larger negative voltages implying the p-type nature of induced charge carriers. If required, this doping effect can be removed by the ALE method. In the as-fabricated device, the semiconducting 1L WSe$_2$ is much more resistive than graphene as shown in Figure 6.2, such that its contribution to the measured resistivity can be neglected. We also measured the conductivity of 1L-TOS layer (after oxidizing the same 1L WSe$_2$) to confirm if the TOS itself is conductive. After oxidation, the current becomes
Figure 6.1: **Electrical and Raman characterization of the TOS-doped graphene.** (a) Device structure at different points in the measurement process. UV-ozone exposure only oxidizes the top WSe$_2$ layer creating monolayer TOS, which strongly p-dopes the underlying graphene layer. (b) Optical image of the TOS-doped graphene device with Hall-bar geometry used for Hall-effect and four-terminal measurements. (c) Raman characterization of the graphene device shows blueshifts in G and 2D peaks after oxidation, indicative of p-doping effect. The insets show a clear reduction in $I_{2D}/I_G$ and FWHM (G peak) after doping. (d) Comparison of $R_{4p}$ as a function of $V_{GS}$ for the graphene device in part b, before and after doping, depicts the strong reduction in $R_{4p}$ after doping. The Dirac peak shifts to higher $V_{GS}$ (beyond our measurement range) after doping indicating the degenerate p-doping of graphene. (e) Zoomed-in view of $R_{sh}$ shows 118 Ω/□ for doped sample at zero gate voltage.
~1 pA, which is our measurement system’s noise floor. Thus, we conclude that carrier transport only takes place in the underlying graphene layer before and after oxidation. This conclusion is further corroborated by the carrier density and mobility extraction in the following section. Figure 6.1(e) shows the sheet resistance ($R_{sh}$) as a function of $V_{GS}$. TOS-doped graphene showed an extremely low $R_{sh}$ of 118 Ω/□ at zero $V_{GS}$.

![Transfer curve of WOx-100 -50 0 50 100 10^-13 10^-12 10^-11 10^-10 10^-9 10^-8 ID (A) VGS (V) V_DS = 1 V 1L-WSe2 on hBN Before UV-ozone After UV-ozone IGS](image)

Figure 6.2: **Transfer curves of a 1L WSe$_2$ device.** Transfer curves of a 1L WSe$_2$ device before and after UV-ozone at $V_{DS} = 1$ V. The Cr/Au contacts were formed by the edge contact method used for graphene devices. It shows an n-type semiconducting behavior before doping and the current drops to the noise floor of our measurement system (gray line indicates the leakage current, $I_{GS}$) after doping.

### 6.2 Detailed characterization of TOS doping

Having confirmed the hole-doping nature of the TOS, we performed comprehensive material and detailed electrical characterization to shed light on the structural properties and doping mechanism.

#### 6.2.1 Material characterization

To study the material properties of TOS layer, we oxidized a monolayer flake of WSe$_2$ as shown Figure 6.3(a). The oxidation produces a clean, continuous oxide film after UV-
Figure 6.3: **Material characterization of monolayer TOS.** (a) Optical micrographs of monolayer (1L) WSe$_2$ before and after oxidation. After oxidation, 1L WSe$_2$ is transformed into monolayer tungsten oxyselenide (TOS) which makes it completely indiscernible from the surrounding SiO$_2$ due to higher transparency. (b) AFM images and height maps confirm that the change in optical contrast in part (a) is due to oxidation. Similar to 1L tungsten oxide (WO$_x$), 1L TOS also show increase in height after oxidation [156]. The AFM scans show that the UV-Ozone process results in uniform oxidation of sample without any pits or damage [167].

ozone exposure which was confirmed using optical and atomic force microscopy (AFM) techniques as shown in Figure 6.3(b). Figure 6.4 shows the color-corrected optical images of a WSe$_2$ flake used to demonstrate the self-limiting nature of our oxidation process. Here, even after employing a substantially longer oxidation process ($5 \times$ than the usual
Figure 6.4: **Self-limiting nature of UV-ozone oxidation.** Optical images highlighting the self-limiting nature of our oxidation process. Only a single layer of WSe$_2$ is removed despite using ($5 \times$) longer oxidation process (2.5 hrs).

30 min duration), only the top 1L WSe$_2$ (indicated as 1L TOS) layer got affected. The self-limiting nature of the oxidation process has been previously studied using first-principle density functional theory (DFT) calculations where it was shown that the top layer (surface) of TMDCs have high-affinity for oxidation in presence of oxygen radicals; however, once the top layer gets fully oxidized, it creates a high vertical diffusion barrier for further oxidation due to strong and directional bonding between oxygen and chalcogen atoms, akin to native SiO$_2$-Si interface [220, 221].

We further analyzed the structural characteristics of TOS layer Raman and electron diffraction studies. Raman measurements on the same monolayer sample used for AFM study (Figure 6.3), depicts the amorphous nature of the top TOS layer as no Raman peak between 790-810 cm$^{-1}$ (indicative of crystalline WO$_x$ [222]) was observed after oxidation as shown in Figure 6.5. To bolster our Raman measurements, we performed transmission electron microscopy (TEM) and selected-area electron diffraction (SAED) study of mono- and few-layer WSe$_2$ flakes as shown in Figure 6.5(b). For the monolayer region, Figure 6.5(c) and (d) show SAED patterns of 1L WSe$_2$ flake before and after oxidation, respectively. Before oxidation, SAED pattern corresponding to pristine 1L WSe$_2$ can be clearly seen with hexagonal symmetry along the [0 0 0 1] zone axis. How-
Figure 6.5: **Structural characterization of monolayer TOS.** (a) Raman spectra of the 1L WSe$_2$ flake given in Figure S1 before and after oxidation. The Raman spectrum after oxidation does not show any additional peak near 800 cm$^{-1}$ which indicates amorphous nature of 1L TOS [222]. (b) TEM image of monolayer and few-layer WSe$_2$. (c,d) SAED patterns of 1L WSe$_2$, before and after UV-ozone process. Single-crystal diffraction patterns with zone axis [0 0 0 1] are clearly visible before UV-ozone and are completely removed after UV-ozone process, confirming the amorphous nature of the TOS layer. (e,f) Corresponding SAED patterns for few-layer WSe$_2$ before and after UV-ozone treatment. Unlike monolayer WSe$_2$, few-layers still show single crystalline patterns even after UV-ozone treatment, indicating the underlying layers are protected by self-limited TOS layer.

However, the diffraction pattern is completely removed after the formation of 1L TOS after oxidation, confirming the amorphous nature of the TOS. Unlike 1L WSe$_2$, the diffraction pattern corresponding to few-layer WSe$_2$ region, shown in Figure 6.5(e) and (f), still shows hexagonal single-crystal patterns even after oxidation, confirming that the underlying WSe$_2$ layers are not damaged due to the oxidation process.

Energy dispersive X-ray spectroscopy (EDS) study of the 1L WSe$_2$ flake, before and oxidation, provides further information about the elemental composition of TOS layer. EDS data provided in Figure 6.6(a) confirms the existence of selenium atoms after UV-
Figure 6.6: **Stoichiometry of monolayer TOS.** (a) The table shows the obtained atomic percentage of C, W, Se, O atoms, and Se/W and O/W ratio for monolayer WSe$_2$ from EDS measurements. The atomic ratio of Se/W for monolayer is significantly reduced (although not completely removed) while O/W ratio increases by the oxidation process, indicating that mostly the Se atoms are replaced by oxygen atoms. (b,c) XPS characterization of 1L WSe$_2$ and 1L-TOS indicating the W (4f) and Se (3d) peaks, respectively. A comparison of W (4f) peaks show a creation of new oxidation states in the TOS layer confirming the formation of W-O bonds and substoichiometric nature of the oxide. Se (3d) peaks also show a formation of Se-O bonds after oxidation with the appearance of broad peaks at higher binding energies.

ozone oxidation of 1L WSe$_2$, suggesting that the oxidation process forms substoichiometric tungsten oxyselenide such as W$_x$Se$_y$O$_z$, which is why we denote it as TOS. Self-limiting nature of the oxidation process further suggests that oxygen atoms mostly replace Se atoms or are bound on top of Se atoms, preventing further penetration of oxygen molecules by high diffusion barrier. It should be noted that the exact stoichiometry was difficult to determine due to atomic thickness of the 1L TOS layer and presence of large oxygen & carbon background signals even before TOS formation. Nevertheless, the Se/W ratio clearly decreases while O/W ratio increases, indicating that it is the selenium atom which is mostly replaced by oxygen atoms. Further details regarding the stoichiometry of the 1L TOS was determined using X-ray photoelectron spectroscopy.
Figure 6.6(b) and (c) provides the XPS spectra of W (4f) and Se (3d) for 1L WSe$_2$ and 1L TOS, respectively. Here, the presence of Se atoms even after oxidation supports our EDS results. XPS measurements for the 1L TOS also depict additional peaks at higher binding energies in both W and Se spectra, confirming the formation of W-O and Se-O bonds after oxidation. The presence of W-O bonds indicates the replacement of Se atoms by O atoms and Se-O bonds indicates the O atoms bound to top Se atoms. Moreover, the higher oxidation states (W$^{5+}$ and W$^{6+}$) in W spectra suggests that 1L TOS is sub-stoichiometric ($W_xSe_yO_z$—$z < 3$) in nature.

6.2.2 Electrical characterization

Having understood the material properties of TOS layer, we further investigated the controllability of our TOS doping method. The self-limiting nature of our oxidation process allows use to tune the doping density using interlayers between the TOS and graphene sheet. We then characterized these devices using Hall measurements for accurate extraction of induced carrier density and mobility. Figure 6.7(a) shows the hole density ($p$) of TOS-doped graphene with 0, 1, 3, and 4L-thick WSe$_2$ interlayers at room temperature (RT). The TOS-doped graphene without the interlayer shows a hole density of $3.2 \times 10^{13}$ cm$^{-2}$ at $V_{GS} = 0$ V (dark red). The applied $V_{GS}$ can further increase the density up to $3.7 \times 10^{13}$ cm$^{-2}$. As we vary the initial layer thickness of WSe$_2$ from 1L to 5L (which becomes 0L to 4L after UV-ozone treatment since the topmost layer is oxidized into TOS), the zero-bias hole density in graphene decreases from 3.2 to $0.4 \times 10^{13}$ cm$^{-2}$. The Fermi-level energy ($E_F$) in the graphene is related to its hole density by

$$E_{CNP} - E_F = \hbar v_F \sqrt{\pi p}, \quad (6.1)$$

where $E_{CNP}$ is the energy of charge neutrality point (Dirac point) in graphene, $\hbar$ is the reduced Planck constant, and $v_F$ is the Fermi velocity of graphene ($10^6$ m/s) [223].

146
Figure 6.7: **Tunable p-type doping and hole mobilities in graphene with WSe$_2$ interlayers.** (a) Hole density, (b) corresponding Fermi-level of graphene with respect to its $E_{\text{CNP}}$, and (c) sheet resistance as a function of back-gate bias, extracted from Hall-effect measurements. Increasing the number of WSe$_2$ interlayers between the TOS doping and graphene reduces the hole density in the graphene. High hole doping densities push the Fermi-level deep into the valence band as shown in the inset. (d,e) Hole mobility as a function of hole density for TOS-doped graphene with WSe$_2$ interlayers at room temperature. LA phonon-limited mobility is achieved at high hole densities in doped graphene with 3L and 4L WSe$_2$ interlayers, suggesting phonon-limited scattering due to the TOS doping. These mobilities are significantly higher than other chemical doping techniques with similar doping densities, highlighted in the bottom gray zone. This also extends the ideal hole mobility, previously achieved in $h$-BN-encapsulated graphene (black points), to higher hole densities. At extremely high hole densities for TOS-doped graphene without and with 1L WSe$_2$ interlayer, our work shows a significant mobility improvement compared to electrolyte gating.

Depending on the WSe$_2$-interlayer thickness and the applied $V_{\text{GS}}$, the Fermi-level of the TOS-doped graphene lies between 0.1 and 0.7 eV below the charge neutrality point as shown in Figure 6.7(b).

Note that the achieved hole density for the TOS-doped graphene without the interlayer is equivalent to approximately 1% of the graphene atomic density ($3.82 \times 10^{15} \text{ cm}^{-2}$), which is comparable to the maximum limit of heavily-doped silicon ($0.6 \sim 1.2\%$) using substitutional doping techniques [224]. Due to the non-substitutional nature of TOS-
doping technique, we achieve almost similar doping densities even at cryogenic temperatures (1.5 K) indicated as open symbols. Thus, TOS doping remains electrically-active even at low temperatures whereas widely used substitutional dopants such as niobium (Nb) in MoS$_2$ leads to significantly lower (electrically-active) carrier density at lower temperatures than the substituted atom density [225, 226]. High doping density and significantly enhanced Hall mobility of TOS-doped graphene at cryogenic temperatures (less than 4 K) is of another great merit for fundamental physics and exploring quantum systems, including quantum computing [227]. More importantly, this high doping density is beyond the limit of conventional electrostatic gating techniques due to the occurrence of dielectric breakdown. For example, an ideal metal gate with a perfect dielectric (i.e. no defects, no leakage current, and a high electrical-breakdown dielectric strength of 1 V/nm) can only accumulate $\sim 2 \times 10^{13}$ cm$^{-2}$ before catastrophically failing; however, in practice, the maximum hole density obtained by electrostatic gating is typically much smaller due to large leakage currents that flow through the gate dielectric when high fields are applied. Although high carrier densities in graphene can be obtained by electrolyte gating, it results in large disorder and low mobilities [228]. Electrolyte gating is also sensitive to environments and unsuitable for batch manufacturing [229, 230].

Figure 6.7(c) shows gate-dependent resistivity of the same samples. All devices show p-type behavior with resistivity ranging from 40 to 250 $\Omega/\square$. It is noted that a weak secondary peak in the TOS-doped graphene with 1L-WSe$_2$ interlayer (green) can be attributed to a slight inhomogeneity in the sample. A significant advantage of the TOS-doping over other doping techniques is highlighted by the comparison of hole mobilities shown in Figure 6.7(d) and (e), calculated from the measured carrier density and resistivity. For reference, Figure 6.7(d) shows the carrier mobility for an ideal h-BN-encapsulated graphene device that represents the highest room-temperature mobility that has been previously demonstrated [216]. When viewed together, our results for
the TOS-doped graphene with 3L and 4L WSe$_2$ interlayers extend the mobility trend of Ref. [216] to higher hole densities that were previously not obtainable. Furthermore, the mobility almost perfectly fits the longitudinal-acoustic (LA) phonon-scattering limit as indicated by gray dashed line suggesting minimal scattering due to the TOS doping [231, 232]. The extracted hole mobilities are significantly (∼10×) higher than other p-type chemical doping techniques at similar doping densities [233, 234, 235, 236, 237]. Consequently, a reduced $R_{sh}$ (∼50 Ω/□) was obtained for the 3L WSe$_2$ interlayer sample further supporting our claim of the intrinsic transport by WSe$_2$ passivation, resulting in reduced scattering and thus higher mobility. Figure 6.7(e) shows the same mobility vs. carrier density plot for TOS-doped graphene without and with 1L WSe$_2$ interlayer, where carrier densities $> 1 \times 10^{13}$ cm$^{-2}$ were achieved. Here, since such high carrier densities were previously achieved only by electrolyte gating, we compare the hole mobilities of our doped graphene to those of electrolyte-gated graphene devices [238, 239, 240, 241, 242]. The hole mobility for TOS-doped graphene is much higher than for electrolyte-gated graphene across the measured hole density range, with a 12.5× mobility improvement at $p \approx 4 \times 10^{13}$ cm$^{-2}$. Although electrolyte gating can induce a wide range of carrier densities, the ionic liquid introduces charged impurities that limit mobility, even at RT [228]. In contrast to electrolyte gating, the charge impurity scattering seems to be less dominant with TOS doping giving rise to the higher mobility.

### 6.3 Understanding charge transfer mechanism in TOS-induced doping

Next, we delve deeper to understand the mechanism behind TOS-based doping. Prior to this, several works on surface charge transfer (SCT) based doping have been demonstrated in the field of 2D semiconductors where the SCT mechanism was often associated with three major processes: (1) SCT due to work function mismatch [243], (2) SCT due to fixed charges [244], and (3) SCT due to dipole interactions [245]. Here, we used a combination of experiment and detailed simulations to prove that the SCT occurs due
to work function mismatch in TOS-based doping. The structural similarity between TOS and WO$_x$ also indicates that TOS can be hypothesised to possess a large work function resulting in further p-type doping of underlying layers [246, 247].

To confirm work function based SCT, let’s consider the schematic of the TOS-doped Gr device with WSe$_2$ interlayers as shown in Figure 6.8(a). Due to work function mismatch between graphene and TOS layer ($W_{TOS} > W_{Gr}$), holes accumulate in the graphene layer and electrons accumulate in the TOS layer in equilibrium conditions. Considering the WSe$_2$ interlayers + van der Waal gap as ideal dielectric (which we will verify shortly) system, we can model the system as a parallel plate capacitor with charges set by difference in work function along with the total separation between Gr and TOS layer ($L$). Using Gauss’s law and charge conservation, we can relate the two work functions and the drop across the dielectric ($\Delta V_{diel}$) as

$$W_{TOS} - W_{Gr} = q\Delta V_{diel}, \quad (6.2)$$

where, $W_{Gr} = E_{ea,Gr} + \Delta E_F$ and $\Delta V_{diel} = \frac{Q_{Gr}}{C_{diel}}$.

Here, $E_{ea,Gr}$ is the electron affinity of pristine Gr (4.5 eV), $\Delta E_F$ is the change in Fermi level in Gr with respect to CNL, $Q_{Gr}$ is the charge in the graphene layer ($Q_{Gr} = qp$), and $C_{diel}$ is the equivalent capacitance of the separation between Gr and TOS layer which can be expressed as

$$\frac{1}{C_{diel}} = \frac{1}{C_{WSe2}} + \frac{1}{C_{vdW}}. \quad (6.3)$$

Since for graphene layer, $\Delta E_F$ can be directly expressed in terms of hole density ($p$) using the relation $\Delta E_F = h\nu_F \sqrt{\frac{4\pi p}{g_s g_v}}$, Equation (6.2) can be rewritten as

$$p = \frac{C_{diel}}{q} \left( W_{TOS} - E_{ea,Gr} - h\nu_F \sqrt{\frac{4\pi p}{g_s g_v}} \right). \quad (6.4)$$

The black square dots in Figure 6.8(b) show the extracted hole density in our TOS-
Figure 6.8: **Charge transfer mechanism for TOS-based doping.** (a) Schematic for TOS-Gr system indicating the origin of charge transfer due to work function mismatch. In equilibrium, the high work function of TOS layer results in hole doping of the graphene layer. (b) Fits (gray line) using Equation (6.4) to the hole density extracted using Hall measurements (black squares). A TOS work function of 5.6 eV was extracted from the fits. The thickness of the films is indicated in parentheses. Our results for the TOS-doped graphene are highlighted in purple and red, showing the highest transmittance and comparable sheet resistance at remarkably small thickness (<5 nm). (c) Simulated sheet charge density in different layers in the TOS-WSe$_2$-Gr stacks for different WSe$_2$ interlayers. Most of the charge is mirrored in the Gr and TOS layers. (d) Raman map of 2D vs. G peak frequencies for pristine graphene (gray), graphene on 1L-TOS (red), and 1L-TOS on graphene (dark red). The gray and orange dotted lines indicate a peak shift of G and 2D peaks with different carrier density and strain, respectively [248]. A clear blueshift obtained from the G and 2D peaks in graphene along the $\epsilon = -0.2\%$ line can be seen after the TOS doping, irrespective of the TOS location. The shifts are in good agreements with previous work on p-type carrier modulation in graphene devices by Das et al.(shown in sky blue) [217].
doped graphene with different interlayer separation ($t_{\text{die}}$) with the charge transfer schematic given in the inset. We fit this data with Equation (6.4) (shown in gray line) to extract the TOS work function. For this, we used Fermi velocity ($v_F$) of $10^8$ cm/s, thickness of van der Waal gap ($t_{vdw}$ of 0.3 nm), and thickness of WSe$_2$ monolayer ($t_{\text{WSe}_2}$ of 0.7 nm). Since the dielectric constant of van der Waal gap is hard to predict, we used the same dielectric constant for both WSe$_2$ and van der Waal gap ($\epsilon_{\text{WSe}_2} = \epsilon_{vdW} = 3.8$) [249]. We extracted a work function of 5.6 eV for TOS layer which matches the theoretically predicted work function values for oxygen-deficient TOS layers [250]. Some previous experimental works with bulk WO$_3$ often performed high-resolution UV photoelectron spectroscopy of either freshly oxidized samples or sputtered samples in vacuum conditions and got a range of work function from 4.8 to 6 eV [251, 252]. It is extremely hard to perform such experiments on TOS layer due to several factors such as vacuum-free oxidation and atomically-thin dimensions. In the absence of these experimental technique, our simple analytical model allows accurate extraction the TOS work function which matches both experimental and theoretically predicted values.

Our simple analytical model assumes that the WSe$_2$ interlayers does act like nearly-perfect insulators allowing all charges to mirror in the graphene layer. To verify that the WSe$_2$ interlayers act like insulators, we performed self-consistent electrostatic simulations to determine the charge distribution along the various layers [253]. Figure 6.8(c) shows the charge stored in each layer for different devices. The simulations confirm that most of the charge is stored in the TOS and graphene layers with minimal ($< 10 \times$) charge being stored in the interlayers. Another possible SCT mechanism which have been reported in literature is based on fixed charge-based doping where the TOS layer can act as a floating gate with lots of fixed negative charge [244]. However, given that our oxidation conditions/mechanism remains same for all the devices with different interlayers, a fixed-charge based doping should have resulted in constant charge irrespective of thickness in Figure 6.8. Since this is not true, we can rule out such doping mechanism.
in our devices.

We further corroborate our claim of work function-based SCT doping mechanism by studying the directionality of TOS doping. Figure 6.8(e) shows the Raman G and 2D band shifts of three samples — pristine graphene, doped graphene with TOS on the bottom (Gr/TOS), and doped graphene with TOS on the top (TOS/Gr). Clearly, the p-type doping effects (blue-shifts of G and 2D bands) can be seen for both stacks (Gr/TOS and TOS/Gr) indicating that the doping is not directional which rules out any fixed dipole-based doping mechanisms associated with self-assembled monolayers and ferroelectric insulators [245, 254]. From the Raman shifts, we can estimate a doping density $>2.5 \times 10^{13} \text{ cm}^{-2}$ for TOS/Gr sample and $\sim 1 \times 10^{13} \text{ cm}^{-2}$ for Gr/TOS sample; in good agreement with previous works [217]. Slight decrease in doping density of the Gr/TOS sample can be attributed to above-room temperature chemical processing during the polymer based stacking and cleaning processes.

6.3.1 Universality of TOS doping

The work-function based SCT mechanism should also allow TOS doping of any other semiconductor/semi-metal with a smaller work function than TOS. To show its universality in this aspect, we built devices with and without TOS doping using a 1D semiconductor (SWCNT), and 2D semiconductor (4L WSe$_2$).

For the SWCNT device, SWCNT were first grown on a SiO$_2$/Si substrate, identified using scanning electron microscope (SEM), and then contacts were made using standard e-beam lithography and e-beam evaporation. The device was then measured before oxidation to get the pristine properties, i.e., data without TOS layer indicated by gray lines in Figure 6.9a. Without the TOS layer, the SWCNT showed a p-type transistor characteristics with an ON/OFF ratio of $\sim 500$. Subsequently, a monolayer of WSe$_2$ was dry-transferred on top of the entire device (channel + near-contact area), which was then oxidized to form TOS, as depicted in the inset. As a consequence, the hole current at $V_{GS}$
Figure 6.9: **Universal p-type doping of various semiconductors with TOS.** Transfer characteristics for (a) SWCNT, and (b) 4L WSe₂, with and without 1L-TOS. A significant increase in the hole current is observed in all the devices with TOS layer. The insets show the structure of each device.

$V_{DS} = 100 V$ increases by $\sim 10 \times$ and becomes nearly independent of back-gate voltage after TOS formation, indicating degenerate hole doping in SWCNT due to the TOS layer.

Finally, we also fabricated a 4L WSe₂ device and measured its characteristics before and after UV-ozone treatment, as shown in Figure 6.9b. Before UV-ozone oxidation, the device showed n-type transistor characteristics due to Fermi-level pinning [255, 256, 257]. After UV-ozone oxidation, the WSe₂ channel thickness was reduced from 4L to 3L, and the device transitioned to completely degenerate-doped, resistor-like behavior due to the high hole density in the channel and near-contact region. We further extracted the hole density and mobility using the Hall-effect measurements where the doping density ranges from 7 to $15 \times 10^{12} \text{cm}^{-2}$ with mobility ranging from 30 to 40 cm$^2$/V·s as a function of applied $V_{GS}$. Overall, the electrical characterizations show that TOS can provide universal p-type doping to a wide range of semiconducting materials.
6.4 Transparent graphene electrodes

In this section, we evaluate TOS-doped Gr as a potential replacement of Indium Tin-Oxide (ITO) for optoelectronic and photonic applications. Graphene has been extensively studied for the past decade for applications in high-speed electronics and photonics due to its extremely high carrier mobility and unique optical properties such as universal light absorption independent of wavelength [216, 258]. Its high conductivity relative to its atomic thickness also enables the possibility of using graphene as transparent and flexible electrodes [259]. Moreover, the linear dispersion in graphene allows simultaneous increase in conductivity and transmissivity with higher doping densities, extending the use of graphene electrodes into near- and mid-infrared (IR) applications. However, the existing techniques to achieve high carrier densities in graphene — based on chemical doping, electrolyte gating, and light and plasma exposure — cause significant material degradation, charge impurities and increased disorder [228, 260]. On the contrary, we showed that our TOS doping technique can achieve a high-quality, degenerate hole densities of $3.2 \times 10^{13} \text{cm}^{-2}$ resulting in an ultra-low sheet resistance ($118 \Omega/\square$) and near-ideal transmittance (99.2%) at IR.

It is expected that this high doping density induced in TOS-doped graphene should lead to a strong change in the interband absorption in graphene by Pauli blocking [266]. Due to p-type doping, the Fermi level shifts deep into the valence band in the TOS-doped graphene that reduces absorption (increases transmittance) for photon energies less than twice the difference between Fermi-level and charge neutrality point (i.e. $E_{ph} < 2 \times |E_{CNP} - E_F|$). Moreover, preservation of high intrinsic mobility of graphene suggests that TOS doping could enable high-speed transparent conductors for IR applications. To confirm this, we measured the transmittance of chemical vapor deposition (CVD)-grown 1L WSe$_2$-graphene films on quartz, before and after oxidation, as shown in Figure 6.10(a). The details of the measurement setup is provided in Appendix E.
Figure 6.10: **Transmittance of TOS-doped graphene.** (a) Transmittance of CVD-grown 1L-WSe$_2$ on graphene before and after the UV-ozone oxidation. Shaded area indicate the standard deviation. The dashed line indicates the transmittance of intrinsic graphene (97.7%). Before UV-ozone, the transmittance remains around graphene’s intrinsic absorption for photon energies less than 1.4 eV. An excitonic band gap peak of WSe$_2$ is also seen at 1.67 eV. After UV-ozone treatment, the peak is reduced significantly along with a increment in the transmittance from 97.2 to 99.2% at telecommunication wavelength (1550 nm). The insets show different CVD stacks on quartz substrate to compare transparency in visible regime. (b) Optical transmission at 1550 nm as a function of sheet resistance for widely-used transparent conducting films [261, 262, 263, 264, 265]. The thickness of the films is indicated in parentheses. Our results for the TOS-doped graphene are highlighted in purple and red, showing the highest transmittance and comparable sheet resistance at remarkably small thickness (<5 nm).

Before oxidation, the transmittance spectrum shows an absorption peak at 1.67 eV that corresponds to excitonic band gap of WSe$_2$ [267]. As expected, the transmittance hovers around graphene’s intrinsic value (97.7%) for photon energies less than 1.4 eV since the top WSe$_2$ layer is expected to be transparent in the near-IR region. In contrast, the spectrum after oxidation shows improved transmittance over the entire region. Specifically, the TOS doping increases the transmittance to more than 99% in the telecommunication band with 99.2% transmission at wavelength ($\lambda$) = 1550 nm, demonstrating its potential as a transparent conductor for near-IR optoelectronic applications. From the transmittance data, we infer that the Fermi level of the TOS-doped graphene is $\sim$0.6 eV below the charge neutrality point of graphene, in reasonable agreement with the hole density extracted from Hall-bar measurements for exfoliated samples (0.65 eV). Furthermore,
the TOS-doped graphene is highly transparent even in the visible regime indicated by the reduction of the WSe$_2$ absorption peak. The increased transparency of TOS-doped graphene in the visible regime compared to pristine WSe$_2$ and WSe$_2$-graphene samples can be clearly seen in the photographs provided in the inset.

To further emphasize on the potential use of TOS-doped graphene as transparent electrode especially in telecommunication regime, we compared the transmittance of our TOS-doped graphene with other widely used transparent conducting films, including CVD graphene (CVD Gr, measured at 2300 nm), ITO, zinc-doped indium oxide (IZO), zirconium-doped indium oxide (IO:Zr), hydrogen-doped indium oxide (IO:H), zinc oxide (ZnO), and aluminium-doped zinc oxide (AZO) [261, 262, 263, 264, 265]. Figure 6.10(b) shows the transmittance at 1550 nm and sheet resistance for different electrodes. The CVD TOS/Gr and exfoliated TOS/3L-WSe$_2$/Gr samples provide exceptionally high transmittance at a given sheet resistance; >99% transmittance at 197 $\Omega$/□ and >97.2% at 48 $\Omega$/□, respectively. More importantly, the TOS-doping method deviates from the general trend (shown in shaded gray area) where higher doping and greater thickness is required to reduce sheet resistance, but it leads to a significant reduction in optical transmittance. Thus, the superior transmittance of TOS-doped graphene, reasonable sheet resistance, and atomically-thin dimensions makes it particularly appealing for use as a transparent electrode for LED applications [236].

6.5 Summary

In this chapter, we demonstrate that monolayer TOS can be used to achieve universal, tunable and, stable p-type doping effects. Monolayer TOS is obtained using room-temperature UV-ozone oxidation of 1L WSe$_2$, similar to the process explained in Chapter 5. Monolayer TOS doping technique overcomes the limitations of previous chemical and substitutional doping techniques as it does not require high-temperature processing, and remains active even at cryogenic temperatures. TOS doped graphene (1L TOS/1L
Gr) shows excellent hole mobilities (>2000 cm²/V·s) even at high hole doping density (3.2 × 10¹³ cm⁻²) which results in low sheet resistance (118 Ω/□) at room temperature. By adding WSe₂ interlayers between TOS and graphene channel, the sheet resistance is further reduced (48 Ω/□) and the hole mobility is significantly enhanced (>50000 cm²/V·s) reaching the upper limit set by LA phonon scattering. We then used an analytical model to prove that the doping occurs through work function mismatch between the TOS and Gr layer and used that to extract the work function of TOS layer. The high work function of TOS allows doping of other semiconducting materials such as SWCNT and WSe₂; demonstrating its use as a universal p-type dopant for wide-range applications, specially with limited thermal-budgets. Finally, the ultrahigh doping density enables high optical transmittance (>99%) at near-IR wavelength; demonstrating the potential of TOS-doped graphene as a transparent conductor for integrated photonic applications at telecommunication wavelength.
Ultra-thin-body van der Waals layered 2D crystals such as graphene and transition metal dichalcogenides (TMDCs) have received significant attention because of their great potential to enable post-silicon nanoelectronics and high-performance optoelectronics. Moreover, the layer-dependent electronic band structure of 2D materials along with the ability to form van der Waals heterostructures make them a unique platform for studying novel physics. Hitherto the lack of analytical models, accurate characterization techniques, and novel fabrication methods appropriate for 2D materials restricted the use in developing high-performance 2D devices. In this thesis, we provided efficient solutions to these issues by providing an analytical formulation of lateral electrostatics, contact resistance, and current-voltage characteristics. We also demonstrate novel etching and doping methods for 2D materials that can enable next-generation electronic and photonic devices. The key results of each chapter are summarized below:

1. **Chapter 2**: entails detailed and systematic derivation of analytical expressions for the electric field, electrostatic potential, and depletion width for 2D lateral p-n junctions. We also extended our method for finding depletion width for 2D lateral heterostructures and metal-2D semiconductor junctions. The analytical model shows that the lateral depletion width at the metal-semiconductor junction is strongly dependent on the barrier height and semiconductor doping density. We further used this model to derive the optimal out-of-plane dielectric thickness required to accurately simulate 2D devices.

2. **Chapter 3**: provides valuable insight into the working of 2D materials-based back-
gated field-effect transistors using a combination of analytical modeling and self-consistent drift-diffusion simulations. Contrary to the general understanding, we show that the channel region (and not the source-contact) limits the device current in OFF state and vice versa. Furthermore, we modeled the two-dimensional current transport at the top-contacted metal-TMDC junction and proved that the edge transport dominates over the vertical current injection. Thus, to reduce the contact resistance in 2D devices, the development of efficient doping techniques while maintaining high mobility of the channel region is of paramount importance.

3. **Chapter 4**: evaluates the applicability of typically used experimental methods to evaluate device parameters such as contact resistance, Schottky barrier height, and channel mobility. We establish that large device-to-device variations, dissimilar sheet resistances in the contact and channel regions, and low sensitivity to channel length variations result in erroneous contact resistance and transfer length extraction using the standard TLM method. Furthermore, based on our renewed understanding of channel region in 2D devices, we explained the erroneous SBH extraction in experimental 2D devices. For mobility measurements, the use of four-probe and Hall-bar measurements are critical to extracting accurate values in the presence of large contact resistance in 2D devices.

4. **Chapter 5**: demonstrates a damage-free atomic layer etch (ALE) for WSe$_2$ using self-limiting oxidation and selective oxide etch for the fabrication of pristine, high-performance 2D devices. The ALE technique produces flakes of deterministic shape and thickness while retaining the properties of pristine flakes. P-FETs fabricated from ALE-processed WSe$_2$ flakes show the highest reported room-temperature hole mobility of 515 cm$^2$/V·s for few (1~5)-layer WSe$_2$ devices. The monolayer precision of the ALE technique enables its application as a universal method for cleaning 2D device surfaces through a process we call the SWAP
technique. Using the SWAP process, we demonstrated graphene devices with the highest reported mobility for unencapsulated graphene (>200,000 cm²/V·s at room temperature).

5. **Chapter 6:** establishes a novel doping technique using a monolayer tungsten oxy-selenide (TOS) layer. TOS-based doping is degenerate (>3×10¹³ cm²), universal (dopes many semiconducting materials such as CNT, WSe₂, MoS₂, graphene), active at cryogenic (1.5 K) temperatures, and stable over 6 months. We achieved record-high transmission (>99%) in telecommunication bandwidth with low sheet resistance (∼120 Ω/sq.) which makes TOS-doped graphene the best-in-class transparent conductor for telecommunication applications.
References


Appendix A: Analytical formulation of the electric field at the junction interface of a symmetrically-doped 2D lateral p-n junction

Figure A.1: Analyzing a symmetrically-doped 2D lateral p-n junction surrounded by an infinitely thick dielectric. The depletion region is divided into infinitely many line charges, which extend into the width of the plane containing the p-n junction. The vertical components of the electric field at the junction interface for two symmetrically placed line charges across the junction interface (n - line and p - line) cancel each other; resulting in only the horizontal component along the x-direction.

The left surface shown in Figure 2.12 of chapter 2 coincides with the y – z plane containing the junction interface. We derive the x-directed electric field at the junction interface of a symmetrically doped 2D lateral p-n junction surrounded with infinitely thick dielectric, by dividing the 2D depletion region into infinitely many line charges extending into the width of the 2D-plane, as shown in Figure A.1. The electric field for
such a single infinite line charge is given by

\[
\vec{E}(\vec{r})|_{\text{line}} = \frac{q\lambda}{2\pi\epsilon_{ox}|\vec{r}|}\hat{r},
\]

(A.1)

where \(\vec{r}\) is the radial vector from the line, \(q\lambda\) is the line charge density (in C/length), and \(\epsilon_{ox}\) is the dielectric constant of the surrounding material.

Further, the line charge density for a single line charge is given by

\[
q\lambda = \lim_{k \to \infty} \frac{qN_{S,2D}(x_2 - x_1)}{k} = qN_{S,2D}dx,
\]

(A.2)

where \(N_{S,2D}\) is the sheet density of the symmetrically doped 2D lateral p-n junction.

As illustrated in Figure A.1, the \(z\)-directed components of the electric field at the junction interface for two symmetrically placed line charges across the junction interface (\(n-\text{line}\) and \(p-\text{line}\)) cancel each other. Thus, the electric field due to two such symmetrically placed line charges lies only along the \(x\)-axis whose \(x\)-component can be expressed as

\[
dE_x(z) = -2|\overrightarrow{\vec{d}E}(z)|_{n-\text{line}} \sin \theta,
\]

(A.3)

where \(\theta\) is indicated in Figure A.1.

Moreover, Equation (A.3) can be expanded using Equation (D.1) and Equation (A.2) as

\[
dE_x(z) = -\frac{qN_{S,2D}dx}{\pi\epsilon_{ox}|r|}, |r| = \sqrt{x^2 + z^2}.
\]

(A.4)

To calculate the accurate total electric field at the junction interface (\(E_x(z)\)), it is essential to accurately capture the impact of the carriers in the quasi-neutral regions. Building upon the approach developed in Section 2.1, Figure A.2 illustrates the total depletion charge and the image charges which captures the screening effects of the quasi-neutral region. The image charges leads to accurate determination of the theoretical 2D depletion width (\(x_{D,inf}\)) by enforcing zero \(x\)-directed electric field at the depletion region.
Figure A.2: Application of method of image charges to model symmetrically-doped p-n junction. A symmetrically doped 2D lateral p-n junction with a depletion width of $x_{D,\text{inf}}$. The image charges enforces zero $x$-directed electric field at the depletion region edges; resulting in accurate calculation of $x_{D,\text{inf}}$. The blocks of charges were chosen in an order that enables easy analytical formulation of the $x$-directed electric field at the junction interface.

Thus, we extend our method to calculate the total $x$-directed electric field due to all the charge blocks given in Figure A.2. We choose the charge blocks such that an easy analytical formulation of the total $x$-directed electric field at the junction interface can be obtained. Hence, the $x$-directed electric field at the junction interface due to the zeroth depletion charge block shown in Figure A.2 can be found by integrating Equation (A.4) and can be expressed as

$$
E_x(z)|_{\text{zeroth block}} = \int_0^{x_{D,\text{inf}}} dE_x(z) = -\frac{q N_{S,2D}}{2\pi \epsilon_{ox}} \ln \left( \frac{z^2 + (x_{D,\text{inf}})^2}{z^2} \right). \quad (A.5)
$$

Extrapolating this methodology to evaluate the $x$-directed electric field at the junction interface due to other charge blocks shown in Figure A.2 results in

$$
E_x(z)|_{\text{other blocks}} = -\frac{q N_{S,2D}}{2\pi \epsilon_{ox}} \left[ \ln \left( \frac{z^2 + (3x_{D,\text{inf}})^2}{z^2 + (2x_{D,\text{inf}})^2} \right) - \ln \left( \frac{z^2 + (2x_{D,\text{inf}})^2}{z^2 + (x_{D,\text{inf}})^2} \right) + \ldots \right]. \quad (A.6)
$$

Furthermore, a compact form of Equation (A.6) can be obtained by using the symmetry of the involved terms and can be expressed as

$$
E_x(z)|_{\text{other blocks}} = -\frac{q N_{S,2D}}{2\pi \epsilon_{ox}} \sum_{n=1}^{\infty} \ln \left[ \frac{(c^2 + (2n + 1)^2)(c^2 + (2n - 1)^2)}{(c^2 + (2n)^2)^2} \right], \quad (A.7)
$$
where
\[ c = \frac{z}{x_{D,\text{inf}}}. \] (A.8)

We can further simplify the infinite summation in Equation (A.7) using the infinite product representations of two elementary Euler functions, which are [52]

\[ \frac{\sin(x)}{x} = \prod_{n=1}^{\infty} \left(1 - \frac{x^2}{(n\pi)^2}\right) \] (A.9)

and

\[ \frac{\cosh(x) - \cos(a)}{1 - \cos(a)} = \left[1 + \frac{x^2}{a^2}\right] \prod_{n=1}^{\infty} \left[1 + \frac{x^2}{(2n\pi - a)^2}\right] \left[1 + \frac{x^2}{(2n\pi + a)^2}\right], \] (A.10)

where, \( a \) is a constant.

Using Equation (A.9) and Equation (A.10) along with some algebraic manipulations, Equation (A.7) can be simply expressed as

\[ E_x(z)|_\text{other blocks} = -\frac{qN_{S,2D}}{2\pi\epsilon_0x} \ln \left[\coth^2 \left(\frac{\frac{\pi z}{x_{D,\text{inf}}}}{2}\right)\right], \] (A.11)

which using our Equation (A.8) can be rewritten as

\[ E_x(z)|_\text{other blocks} = -\frac{qN_{S,2D}}{2\pi\epsilon_0x} \ln \left[\coth^2 \left(\frac{\frac{\pi z}{x_{D,\text{inf}}}}{2}\right)\right]. \] (A.12)

Finally, Equation (A.12) and Equation (A.5) can be combined and further simplified to obtain an analytical expression for the \( x \)-directed electric field at the junction interface due to all charge blocks which can be expressed as

\[ E_x(z) = -\frac{qN_{S,2D}}{\pi\epsilon_0x} \ln \left[\coth \left(\frac{\frac{\pi z}{x_{D,\text{inf}}}}{2}\right)\right]. \] (A.13)
Appendix B: Schottky current in 2D transistors

The total current per unit width \( I_{SB} \) at a metal-2D semiconductor junction can be expressed as

\[
I_{SB} = \int_{-\infty}^{\infty} q v_x DOS(E) [f_M(E) - f_S(E)] T_{SB}(E_x) dE, \quad (B.1)
\]

where \( q \) is elementary charge, \( v_x \) is the velocity of electron in \( x \) direction, \( DOS(E) \) is the density of states for \( E \geq E_C \), \( f_M \) is the Fermi distribution function of the metal, \( f_S \) is the Fermi distribution function of the semiconductor, and \( T_{SB}(E_x) \) is the \( x \)-directed transmission probability.

The 2D density of states can be given as \( DOS(E) = \frac{g_s g_v m_e^*}{2\pi\hbar^2} \) for \( E \geq E_C \), where \( E_C \) is the conduction band minima at the location \( x \), \( g_s \) is the spin degeneracy factor, \( g_v \) is the valley degeneracy factor, and \( m_e^* \) is the effective mass of electron. Thus,

\[
I_{SB} = \frac{2\pi g_s g_v m_e^* q}{\hbar^2} \int_{E_C}^{\infty} v_x [f_M(E) - f_S(E)] T_{SB}(E_x) dE. \quad (B.2)
\]

Since \( E = \frac{m_e^* v_x^2 + m_e^* v_y^2}{2} \),

\[
I_{SB} = \frac{2\pi g_s g_v m_e^* q}{\hbar^2} \int_0^{\infty} v_x \left[ f_M\left( \frac{m_e^* v_x^2}{2} + \frac{m_e^* v_y^2}{2} \right) - f_S\left( \frac{m_e^* v_x^2}{2} + \frac{m_e^* v_y^2}{2} \right) \right] T_{SB}\left( \frac{m_e^* v_x^2}{2} \right) dE, \quad (B.3)
\]

where \( v = \sqrt{v_x^2 + v_y^2} \). Using the conversion \( dv_x dv_y = 2\pi vdv \),

\[
I_{SB} = \frac{g_s g_v m_e^* q}{\hbar^2} \int_0^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \times

v_x \left[ f_M\left( \frac{m_e^* v_x^2}{2} + \frac{m_e^* v_y^2}{2} \right) - f_S\left( \frac{m_e^* v_x^2}{2} + \frac{m_e^* v_y^2}{2} \right) \right] T_{SB}\left( \frac{m_e^* v_x^2}{2} \right). \quad (B.4)
\]
Here, $v_0$ is the minimum velocity in the transport direction ($x$). Using the integral,

$$
\int_{-\infty}^{\infty} dv_y \frac{1}{1 + \exp[m_e^* v_y^2/2k_B T + C]} = \sqrt{\frac{2\pi k_B T}{m_e^*}} F_{-1/2}(C), \quad (B.5)
$$

we integrated out the variable in $v_y$ direction as

$$
I_{SB} = \frac{g_s g_v m_e^* q}{\hbar^2} \sqrt{\frac{2\pi k_B T}{m^*}} \int_{v_0}^{\infty} dv_x
\times v_x \left[ F_{-1/2} \left( \frac{E_{Fn,M}}{k_B T} - \frac{m_e^* v_x^2}{2k_B T} \right) - F_{-1/2} \left( \frac{E_{Fn,S}}{k_B T} - \frac{m_e^* v_x^2}{2k_B T} \right) \right] T_{SB} \left( \frac{m_e^* v_x^2}{2} \right), \quad (B.6)
$$

where $E_{Fn,M}$ is the Fermi level in the metal and the $E_{Fn,S}$ is the Fermi level in the semiconductor. Here, $F_{1/2}$ is the Fermi integral

$$
F_j(\eta) = \frac{1}{\Gamma(j+1)} \int_0^{\infty} \frac{\xi^j d\xi}{1 + e^{\xi - \eta}}. \quad (B.7)
$$

Finally, using $dv_x = \frac{dE_x}{m_e^* v_x}$,

$$
I_{SB} = \frac{g_s g_v q}{\hbar^2} \sqrt{2\pi k_B m_e^* T} \int_{E_C}^{\infty} dE_x \left[ F_{-1/2} \left( \frac{E_x - E_{Fn,M}}{k_B T} \right) - F_{-1/2} \left( \frac{E_x - E_{Fn,S}}{k_B T} \right) \right] T_{SB}(E_x)(B.8)
$$
Appendix C: Self-consistent simulations with drift-diffusion and hopping transport

Fig. 1 shows the schematic of the MoS$_2$ BGFET used for the simulations. Here, the $x$-axis is aligned along the transport direction of the channel. The end of the channel at the source side is set to $x = 0$. In $z$-direction, the top of the 2D semiconductor is set to $z = 0$. We then start by solving the current continuity equation which can expressed as,

$$\frac{1}{q} \nabla J(x) - R(x) = 0. \quad (C.1)$$

Here, $J$ is the electron current density and $R$ is the recombination rate. In most long-channel 2D devices, the current transport in the channel region can be significantly affected by trap states or structural disorder resulting in variation of the transport mechanism from continuous band transport (also known as drift-diffusion) to trap/disorder-assisted hopping transport. In order to simulate this effect, we model $J(x)$ as

$$J(x) = -q[\mu_{\text{eff}}(x)n_s(x) + \mu_{\text{trap}}(x)n_{\text{trap}}] \nabla \Phi(x,0) + qD_n(x) \nabla n_s(x) + qD_{\text{trap}}(x) \nabla n_{\text{trap}}(x), \quad (C.2)$$

where, $\mu_{\text{eff}}(x)$ is the effective mobility at position $x$ in the 2D semiconductor given by Eq. 5, $n_s(x)$ is the free electron density, $\mu_{\text{trap}}(x)$ is the trap mobility (which is related to the hopping mobility), $n_{\text{trap}}(x)$ is the electron density in the band tails due to traps, $\Phi(x,0)$ is the potential, $D_n(x)$ is the diffusion constant for band transport, and $D_{\text{trap}}$ is the diffusion constant for the hopping current. Both diffusion constants, $D_n$ and $D_{\text{trap}}$,
Figure C.1: Schematic of the MoS$_2$ BGFET used for self-consistent simulations.

are related to respective effective mobility by using the general Einstein’s relation

$$D_n(x) = g_n(x) \frac{k_B T}{q} \mu_{\text{eff}}(x) \quad \text{and} \quad D_{\text{trap}} = g_{\text{trap}}(x) \frac{k_B T}{q} \mu_{\text{trap}}(x), \quad (C.3)$$

where,

$$g_n(x) = \frac{1}{k_B T} n_s(x) \left[ \frac{\partial n_s(x)}{\partial E_{Fn}(x)} \right]^{-1} \quad \text{and} \quad g_{\text{trap}}(x) = \frac{1}{k_B T} n_{\text{trap}}(x) \left[ \frac{\partial n_{\text{trap}}(x)}{\partial E_{Fn}(x)} \right]^{-1}. \quad (C.4)$$

For calculating $n_{\text{trap}}(x)$ and $\mu_{\text{trap}}(x)$, we follow the analysis provided by Wang et al. where variable range hopping was used to model hopping transport [268]. Following their methodology, the 2D density of states ($\text{DOS}_{2D}$) considering the band tails can be expressed as

$$\text{DOS}_{2D}(E) = \begin{cases} \text{DOS}(E) = \frac{2\pi \tau \xi g_0 m^*}{\hbar^2}, & \text{for } E \geq E_c, \\ \text{DOS}_{\text{trap}}(E) = \frac{N_{IT}}{k_B T_0} e^{\frac{E-E_c}{k_B T_0}}, & \text{otherwise.} \end{cases} \quad (C.5)$$

Here, $N_{IT}$ is the interface trap density and $T_0$ is the disorder magnitude parameter. The free electron density ($n_s(x)$) and trapped electron density ($n_{\text{trap}}(x)$) can now be obtained
Figure C.2: Simulated total current ($I_D$) for different $N_{IT}$ (thick solid line). The dotted line denotes the hopping current, and the thin solid line denotes the band (drift-diffusion) current. Note that the hopping current for $1 \times 10^{12}$ cm$^{-2}$ and the band current for $1 \times 10^{13}$ cm$^{-2}$ and $6 \times 10^{12}$ cm$^{-2}$ are smaller than $1 \times 10^{-12}$ A/µm. The dominant current component changes from band to hopping transport for $N_{IT}$ between $1 \times 10^{12}$ cm$^{-2}$ and $6 \times 10^{12}$ cm$^{-2}$.

by integrating Eq. 13

$$n_s(x) = \frac{2\pi g_s g_v q m^* k_B T}{\hbar^2} \log\left\{ 1 + \exp\left[ \frac{-(E_C(x) - E_{Fn}(x))}{k_B T} \right] \right\} \quad (C.6)$$

$$n_{\text{trap}}(x) = \int dE \frac{N_{IT} e^{\frac{E-E_C(x)}{k_B T}}}{k_B T_0} \frac{1}{1 + e^{\frac{E-E_{Fn}(x)}{k_B T}}} \quad (C.7)$$

Next, we calculate the trap mobility using the relation

$$\mu_{\text{trap}}(x) = \mu_{\text{hop}} \exp\left( -\frac{E_{\text{trap}}(x)}{k_B T} \right), \quad (C.8)$$

where $\mu_{\text{hop}}$ is hopping parameter and $E_{\text{trap}}$ is the effective energy level visited by the
charge carriers relative to the Fermi level inside the bandgap. This energy level can be calculated as

\[ E_{\text{trap}}(x) = 3k_B T_0 \left[ \left( \frac{2k_BT}{\xi} \right)^2 / \pi \right]^{1/3} e^{- \frac{E_C(x) - E_F(x)}{3k_B T_0}} \left[ \frac{9}{(k_BT_0)^2 N_{IT}} \right]^{1/3}. \] (C.9)

where, \( \xi \) is the localized length.

The Schottky current is then included in the recombination term as \( R(x) = \frac{dJ_{SB}(x)}{dx} \), where \( J_{SB}(x) \) is the current densities injected between \( x = 0 \) to \( x \). Further details are provided in our earlier work [90].

Finally, the Poisson equation was solved to find \( \Phi(x, z) \) as

\[ \nabla \epsilon_\alpha \cdot \nabla \Phi(x, z) = -qC(x, z), \] (C.10)

where \( C(x, z) = -n_s(x) - n_{\text{trap}}(x) \) for \( -t_{\text{ch}} < z < 0 \), and \( C(x, z) = 0 \) for \( z < -t_{\text{ch}} \) and \( z > 0 \). \( \epsilon_\alpha \) is the dielectric constant at the materials (\( \epsilon_{s,x} \), \( \epsilon_{s,z} \) for the 2D semiconductor, \( \epsilon_{ox} \) for the oxide layer and \( \epsilon_0 \) for air). Here, \( t_{\text{ch}} \) is the thickness of the 2D semiconductor.
The boundary condition is set to satisfy the continuity of the normal vector of the electric flux density at the interface between materials. The other boundary conditions are constrained by the Neumann boundary condition except at the metal gate. The boundary condition at the metal contact is set as $\Phi(x, -t_{ch} - t_{ox}) = V_{GS} + \Phi(0, 0)$, where $t_{ox}$ is the thickness of SiO$_2$. The carrier densities in the current equation and the potential in the Poisson equation are determined self-consistently in the numerical calculations.

In the simulation for the MoS$_2$ BGFET, we set $t_{ox} = 3$ nm. The dielectric constant of MoS$_2$ is set to $\epsilon_{s,x} = 5.21\epsilon_0$ in the $x$-direction is $\epsilon_{s,z} = 1.35\epsilon_0$ in $z$-direction. We set $\mu_{\text{hop}} = 50$ cm$^2$/V·s, $\xi = 3.4$ nm, and $T_0 = 2000K$ [268, 269, 270]. Fig. S2 shows the dominant channel current components for varying interface trap densities ($N_{IT}$). Clearly, the subthreshold swing degrades with increase in $N_{IT}$. Moreover, the dominant transport mechanism switches from the band to hopping transport for $N_{IT}$ between $1 \sim 6 \times 10^{12}$ cm$^{-2}$ and thus the device can no longer be modeled considering only drift-diffusion model at high trap densities. However, the drift-diffusion model is perfectly valid for $N_{IT} \leq 1 \times 12$ cm$^{-2}$. Thus, we adopt the simple drift-diffusion model for the channel transport to understand the current flow in Schottky transistors with $N_{IT} \leq 1 \times 10^{12}$ cm$^{-2}$.

Fig. S3 shows the comparison of simulated $I_D$ for different $N_{IT}$ (shown with dots) with $I_D$ calculated using our load-line model which assumes simplified $\mu_{\text{eff}}$ (independent of carrier density ($n_s$)) given by Eq. 6. The analytical model provides excellent fits to the simulation data.
Appendix D: Shape of the Lateral Potential Profile

As discussed in the text, we approximate the lateral dipole of the structure as a line charge at the corner of the metal contact at \((0, t_{\text{gap}})\) and a sheet charge representing the depletion in the semiconductor extending from \((0, 0)\) to \((x_D, 0)\), which is illustrated in Figure D.1. Here, we solve for the potential profile due to this dipole.

First, we begin with the line charge on the metal. The electric field due to a line charge is given by

\[
\vec{E}(\vec{r})|_{\text{line}} = \frac{q_{N1D}}{2\pi\varepsilon_{\text{eff}} |\vec{r}|} \hat{r},
\]

(D.1)

where \(\vec{r}\) is the radial vector from the line, \(q_{N1D}\) is the line charge density (in C/length), and \(\varepsilon_{\text{eff}}\) is the effective dielectric constant equal to \(\frac{\varepsilon_{\text{top}} + \varepsilon_{\text{bottom}}}{2}\) as detailed in [232]. This allows the x-directed field along \(y = 0\) to be written as

\[
E_x(x)|_{\text{line}} = \frac{\lambda_{1D}x}{x^2 + t_{\text{gap}}^2}, \quad \text{where} \quad \lambda_{1D} = \frac{q_{N1D}}{2\pi\varepsilon_{\text{eff}}}.
\]

(D.2)

From electric field, we obtain the potential due to this line charge as

\[
\phi(x)|_{\text{line}} = \int_0^x -E_x(x')|_{\text{line}} dx' = -\lambda_{1D} \ln \left(\sqrt{\frac{x^2}{t_{\text{gap}}^2} + 1}\right),
\]

(D.3)

where zero reference potential is defined at \(x = 0\).

The electric field along the semiconductor layer due to the sheet of depletion charge on the semiconductor is given by

\[
E_x(x)|_{\text{sheet}} = \lambda_{2D} \ln \left(\frac{x}{x_D - x}\right), \quad \text{where} \quad \lambda_{2D} = \frac{q_{N2D}}{2\pi\varepsilon_{\text{eff}}}.
\]

(D.4)
The potential due to the depletion charge is determined to be

\[ \phi(x)_{\text{sheet}} = \int_0^x -E_x(x')_{\text{sheet}} \, dx' = -\lambda_{2D} x_D \left[ \frac{x}{x_D} \ln \left( \frac{x}{x_D} \right) + \left( 1 - \frac{x}{x_D} \right) \ln \left( 1 - \frac{x}{x_D} \right) \right]. \] (D.5)

To find the lateral potential along the semiconductor, we add the potential contributions from both the line charge and sheet charge given in Eqs. (D.3) and (D.5). The line charge on the metal must be equal and opposite to the integrated sheet charge on the semiconductor,

\[ qN_{1D} = -qN_{2D} x_D, \] (D.6)

allowing the lateral potential profile along the semiconductor to be written as

\[ \phi_{\text{lat}}(x) = \phi_{\text{lat}} \left[ \ln \left( \sqrt{\frac{x^2}{t_{\text{gap}}} + 1} \right) - \frac{x}{x_D} \ln \left( \frac{x}{x_D} \right) - \left( 1 - \frac{x}{x_D} \right) \ln \left( 1 - \frac{x}{x_D} \right) \right]. \] (D.7)

where,

\[ \phi_{\text{lat}} = \frac{qN_{2D} x_D}{2\pi \epsilon_{\text{eff}}} \] (D.8)

Figure D.1 shows the lateral potential profile for different sheet densities \( (N_{D,2D}) \) for fixed \( t_{\text{gap}} \).
Figure D.1: Normalized lateral potential profile for different sheet densities ($N_{D,2D}$) for fixed $t_{gap}$. The $x$-axis is scaled with respect to lateral depletion width ($x_D$)
Appendix E: Number of modes in 2D semiconductors

For carrier flow in a two-dimensional channel, $M(E)$ is expressed by

$$M(E) = W \frac{h}{4} \langle v^+_x (E) \rangle D_{2D}(E) \quad (E.1)$$

where $\langle v^+_x (E) \rangle$ is the average electron velocity along the channel direction and $D_{2D}(E)$ is the density of states of the semiconductor (including spin). For parabolic circular energy bands in a 2D plane, the average velocity along the $+x$-direction (i.e. along the channel) is

$$\langle v^+_x (E) \rangle = \frac{\int_{-\pi/2}^{+\pi/2} v(E) \cos(\theta) \, d\theta}{\pi} = \frac{2}{\pi} v(E), \quad (E.2)$$

where $v(E)$ is the magnitude of the velocity at a specific energy, given by

$$v(E) = \sqrt{\frac{2(E - E_c)}{m^*_e}} \quad \text{for} \ E \geq E_c, \quad (E.3)$$

where $E_c$ is the energy of the conduction-band edge. The conduction-band density of states for a two-dimensional semiconductor (including spin degeneracy) is

$$D_{2D} = g_v \left( \frac{m^*_e}{\pi \hbar^2} \right) \quad (E.4)$$

Putting this all together, we find that the number of modes in a 2D semiconductor is given by

$$M(E) = W \frac{g_v}{\pi \hbar} \sqrt{2m^*_e (E - E_c)} \quad \text{for} \ E \geq E_c, \quad (E.5)$$

where $W$ is the width of the device and $g_v$ is the valley degeneracy.