2-D Materials: Low Power Electronics and Novel Device Possibilities

Varun Mishra
Sayeef Salahuddin, Ed.
Vivek Subramanian, Ed.
Tsu-Jae King Liu, Ed.
Junqiao Wu, Ed.

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2019-13
http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-13.html

May 1, 2019

Acknowledgement

First, I would like to thank Professor Sayeef Salahuddin for providing me the opportunity to work with him and supporting me through the course of my PhD. Professor Salahuddin's persistent encouragement for me to push my comfort zone played a very vital role as I navigated my way through various projects in this work. Observing his methods on how to attack the different challenges involved in a project was a great learning experience which has provided me with important perspectives for the rest of my life. I would also like to thank Professors Vivek Subramanian, Tsu-Jae King Liu and Junqiao Wu for taking time to be part of my qualifying examination committee as well as the dissertation committee.

2-D Materials: Low Power Electronics and Novel Device Possibilities

by

Varun Mishra

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Sayeef Salahuddin, Chair
Professor Vivek Subramanian
Professor Tsu-Jae King Liu
Professor Junqiao Wu

Fall 2017

2-D Materials: Low Power Electronics and Novel Device Possibilities

Abstract

2-D Materials: Low Power Electronics and Novel Device Possibilities

by

Varun Mishra

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Sayeef Salahuddin, Chair

Semiconductor modeling assumes a critical role in understanding the physics of nanoscale devices as semi-classical approaches start breaking down for highly scaled devices. In this thesis, a quantum transport simulator is designed within the non-equilibrium Greens function (NEGF) formalism to explore the application of layered materials in low-power electronic and novel spintronic devices.

The quantum transport simulator is used to calculate the optimal performance of transition metal dichalcogenides (TMD) field effect transistors at the ultimately scaled limit of 5 nm gate lengths. MoS_2 based transistors are studied in further detail using a rigorous tight-binding Hamiltonian in order to include the effects of non-parabolicity of the electronic structure on the transistor performance. Large band gaps and effective masses were shown to result in excellent switching performance with ON/OFF ratios of 10^5 even at 5 nm gate lengths showing potential applicability in low power electronics, while screening effects were found to affect the drive current in few-layer devices.

Possibility of a one-dimensional spin channel is also discussed at the junction of a lateral heterostructure of two dimensional TMDs when the system is in non-equilibrium in the absence of any external magnetic field. The one-dimensional spin channel arises out of spin-orbit splitting in conjunction with the in-plane electric field that exists in the heterojunction of two materials with a conduction band offset. A spin polarization of 0.1 % was calculated at the junction of MoS_2/WSe_2, while the effect of effective mass and spin diffusion on the above is discussed. Such one dimensional spin channel could open up the possibility to probe spin dynamics in confined phase spaces while contributing to spintronics applications.

Contents

List of Figures

List of Tables

Acknowledgments

First, I would like to thank Professor Sayeef Salahuddin for providing me the opportunity to work with him and supporting me through the course of my PhD. Professor Salahuddin's persistent encouragement for me to push my comfort zone played a very vital role as I navigated my way through various projects in this work. Observing his methods on how to attack the different challenges involved in a project was a great learning experience which has provided me with important perspectives for the rest of my life. I would also like to thank Professors Vivek Subramanian, Tsu-Jae King Liu and Junqiao Wu for taking time to be part of my qualifying examination committee as well as the dissertation committee. Their valuable inputs during the qualifying examination helped me understand the importance of going beyond the minute details in order to attain a more holistic view of the current state of the art to motivate future work.

In all of my work I received significant help from the members of Professor Salahuddin's group at UC Berkeley, especially the members of the modeling and simulation sub-group. I am grateful to Kartik Ganapathi for mentoring me in the first year of graduate school by helping me understand the intricacies of quantum transport simulations and getting me up to speed to contribute in the field of device research. Samuel Smith and Sangwook Kim helped me in tackling a lot of the technical difficulties that I had over the course of graduate school. I gained significant insights from our conversations over the years and I hope they did the same. I'd also like to thank Dominic Labanowski for the friendly rivalry that we shared over the years and for making my transition into a new country very easy. The discussions on Physics that I used to have with Khalid Ashraf in the first few years were always very exciting and it was inspiring to see his career progress post graduation. I'd also like to thank Justin Wong, Korok Chatterjee and Praveen Gowtham for the countless stimulating discussions and adventures that we had and for making graduate school a very enjoyable experience. I enjoyed the conversations and will cherish the time I spent with Debanjan Bhowmik, Niklas Roschwesky and all the other members of the LEED group.

I'd like to thank Somil Bansal, Vinay Ramasesh, Aamod Shanker, Ankit Singh, Raj Shekhar Singh, Vijay Kamble, Jason Poon and Adriana Bastardas for being great friends and always having a patient ear to lend. I am especially grateful to Juan Pablo Duarte, Gerd Brandstetter, Jeff Clarkson, Julie Anton and Nicole Repina who encouraged me to spend more time outdoors which became an important part of my life over the last few years.

Finally, I'd like to thank my brother, Vivek Mishra for being a constant pillar of support and the best possible partner in all my endeavours over the years. I wish him the best as we briefly part ways. I am forever indebted to my parents for providing me with all the kind words of encouragement, support and resources so that I could get to wherever I needed to. None of what I've done would be possible without them.

Chapter 1

Introduction

In this chapter, we look at the motivation behind the need for low power electronic devices as well as the current state of the art in ultimately scaled transistors. Subsequently, we'll look at the various challenges faced by industry as the transistors are scaled to increase density and speed of operation. In the process, we discuss the major advances in the field by looking at the available literature, while motivating our approach to the problem. Finally, we'll do a brief overview of all the studies that we did to tackle the problem, while understanding the questions that we're trying to answer in this thesis.

1.1 Overview of Low Power Electronics and Scaling

Low power electronics have gained a lot of importance in the last few decades as the demand for mobile computing increased. Mobile technology has come a long way from the first electronic wrist-watches and hand-held calculators to the interconnected technology of today as smart wearable devices and "internet of things" becomes ubiquitous in our daily lives. Eventually, internet nodes might reside in everyday use things starting such as food, transport or home appliances. [6]. With this, there is an increased demand of research on ultra-low power nano-electronic devices to enable the technology. The packing density of transistors in an integrated circuit and their speed had been increasing steadily over decades after the first conception in 1958 following the well-known trend described by Moore's law (Fig. 1.1) [55, 80]. This was sustainable for the first 40 years as Dennard scaling allowed substantial power savings per node (via reduction of capacitance and drive voltage) which allowed the operating frequency to be ramped up in order to improve the performance of integrated circuits [10]. This approach has had to be modified in the recent past as OFF current became a significant issue. The power dissipation of a CMOS circuit is given by [45]

$$P = f_{clk} a C_{tot} V_{DD}^2 + V_{DD} I_{SC} + V_{DD} I_{OFF} \tag{1.1}$$

where f_{clk} is the clock frequency, a the average switching activity, V_{DD} the supply voltage, I_{SC} the short-circuit current between the supply voltage rails and I_{OFF} the transistor OFF

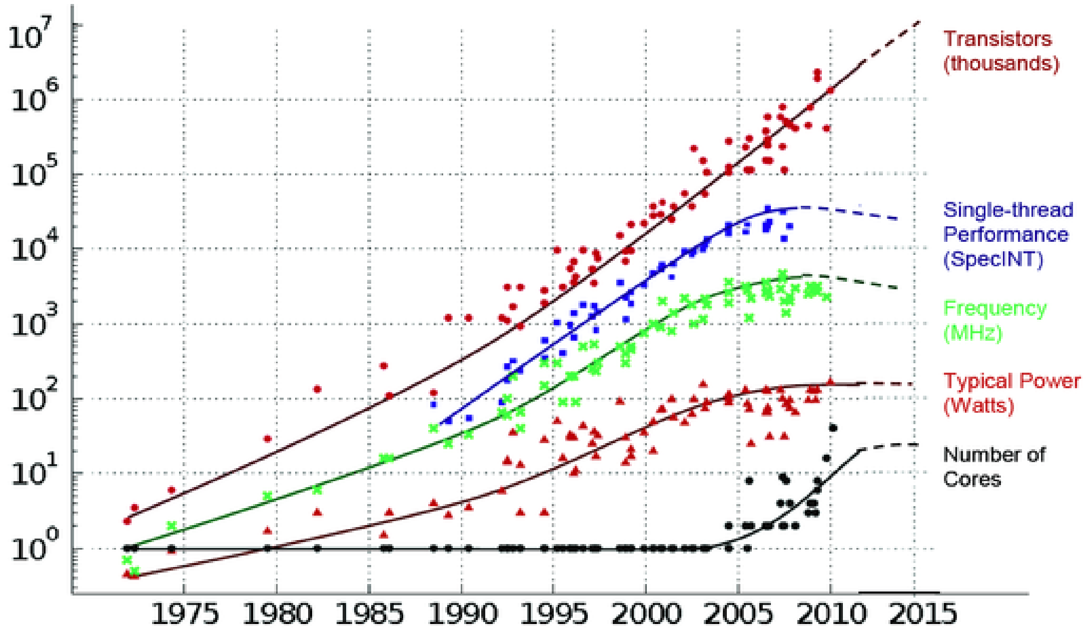


Figure 1.1: Illustration of current trends in the scaling of integrated circuits. Number of transtors per die has kept on increasing exponentially - doubles every 24 months in accordance with Moore's Law [80]. On the other hand, frequency of operation has tapered off as drive current levels haven't improved significantly in the last decade since the breakdown of Dennard scaling which allowed clock frequency to be increased every generation [10]. The reason for the same can be explained from the power dissipation plot in the same figure as it got to levels where it cannot be drawn out effectively using heat sinks. This is because the transistor OFF current poses significant challenges at the scaled nodes. This led to an increase in number of cores to continue performance improvements in the early 21st century.

current. In the early days of I.C development, it was possible to scale V_{DD} and C_{tot} to reduce the power dissipation. This was not possible after the threshold voltage was reduced to below 300 mV as the OFF current increased by a lot. This arises from a thermodynamic limit as the switching mechanism in a traditional MOSFET is dictated by a thermal barrier. The limit is given by the Fermi distribution which dictates that there is a lower limit to the ratio of change in gate voltage to the change in channel mobile charged carrier density and hence the rate the transistor current can change, i.e, about 60 mV/decade at room temperature ($k_B T ln(10)/q$ where k_B is the Boltzmann's constant and T is the temperature). This is also known as the subthreshold swing (SS) of the transistor. Thus, a threshold voltage of 300 mV provides a maximum ON/OFF ratio of 10^5. As devices are scaled, further challenges arise known as short channel effects that increase the subthreshold swing. A significant factor contributing to these are the degraded electrostatics as channels become shorter. This is further illustrated in Fig. 1.2, where the electrostatic potential at the top of the barrier is determined by several capacitance connected to different electrodes [81]. Ideally, the barrier is controlled only by the gate capacitance, which was true when the channel was long and the source and drain capacitances were relatively small. This is not true anymore as the channel length is shorter and the source and drain electrodes significantly affect the barrier height (reduce the barrier height in OFF state) leading to higher OFF current. This leads to increased subthreshold swing and other short-channel effects such as DIBL (Drain Induced Barrier Lowering) [111]. In order to improve gate control, the gate oxide was thinned down resulting in tunneling current through the gate oxide which was inhibited using high-K dielectrics [78]. In addition to that, as the source-drain distance is reduced, direct quantum mechanical tunneling can occur between the two electrodes in off state, thus increasing OFF current. There is therefore a growing need to rethink transistor design to tackle the challenges of scaling. There are two types of ideologies that have been pursued for the same - 1) innovations within the same transport mechanisms as traditional MOSFETs (switching via the control of a thermal barrier) to improve electrostatics using either different geometry or channel materials and 2) use a different transport mechanism to overcome the thermodynamic barrier limit such as tunnel transistors and negative capacitance.

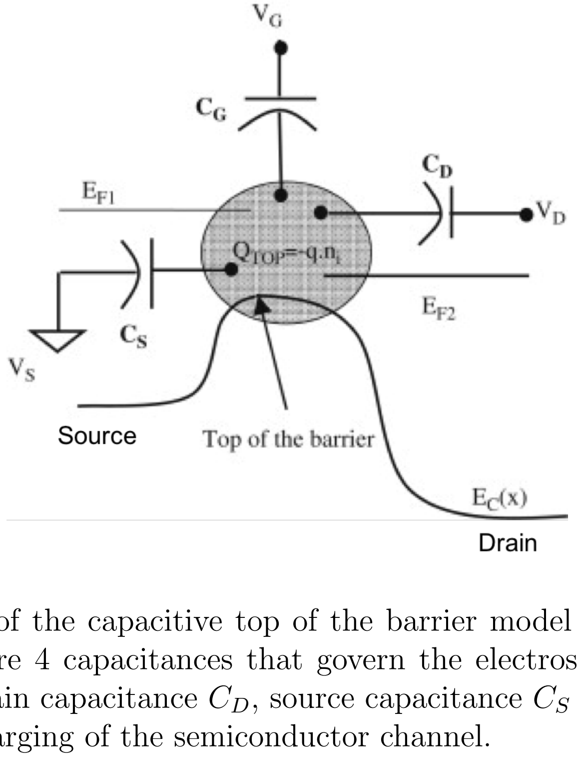


Figure 1.2: Illustration of the capacitive top of the barrier model to describe the working of a MOSFET. There are 4 capacitances that govern the electrostatics of the device, the gate capacitance C_G, drain capacitance C_D, source capacitance C_S and finally the quantum capacitance from the charging of the semiconductor channel.

1.2 Approaches for Future Transistor Design

Sub-60 mV/decade Technology

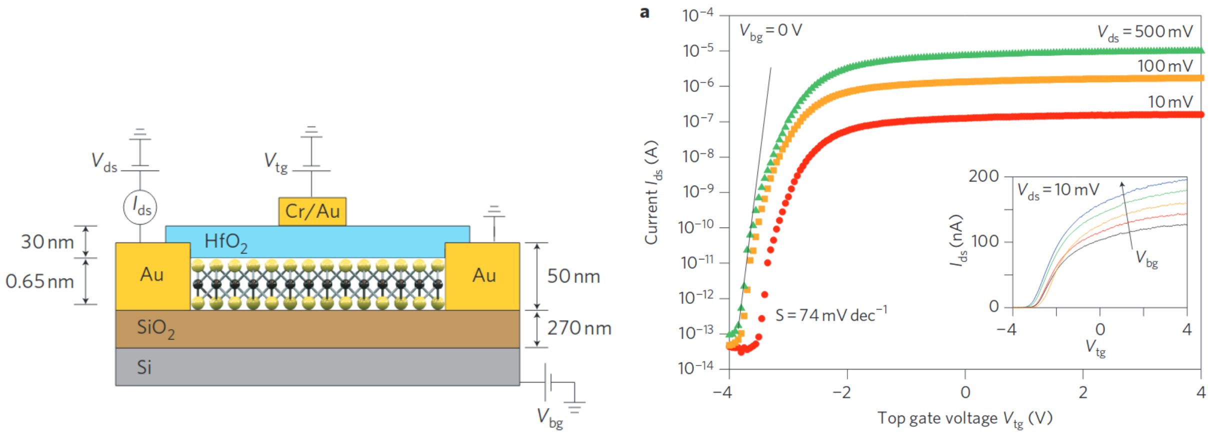
There has been a lot of ongoing work to break the thermodynamic limit which has governed the transistor behavior for the past 5 decades. One of the approaches is to exploit quantum-mechanical tunneling as the charge transport mechanism as it becomes relevant in the ultra-scaled regime. In TFETs (Tunnel FETs), instead of creating a thermal barrier to the flow of charged particles, the transistor is switched OFF by eliminating states for the charged particle to tunnel into. This is done in band-to-band tunneling transistors by modulating the crossing of the valence and conduction band on either side of a junction via a gate bias voltage [47]. The sub-60 mV/decade SS is achieved because the tail of the Boltzmann distribution is clipped by the gap of the semiconductor. Since, the first development of a carbon nanotube TFET by J. Appenzeller [5], which showed a SS of 40 mV/decade, various advances in the field has been made including the use of group III-V materials like InGaAs and InAs, use of staggered and broken-gap heterostructures as well as the introduction of resosnant diodes to sharpen the turn-on [8, 47, 7, 2]. Notwithstanding the progress, there are a few limitations of tunnel transistors which has impeded their application in the industry such as band-tails (due to high doping) which reduces the ON/OFF ratio [54] and low ON currents because of low probability of tunneling leading to low drivability [47]. In chapter 7, we'll look at another device which uses tunneling in Esaki diodes to result in magnetoresistance. Another device proposed to achieve an SS of sub-60 mV/decade uses the concept of negative capacitance [98]. Negative capacitance is based on stabilizing the polarization a ferroelectric by stacking it with a dielectric. This effective negative capacitance can result in lower SS as the gate capacitance effectively becomes negative in Fig. 1.2. Following the experimental evidence of negative capacitance in PZT-STO stacks by A. Khan [48], sub-60 mV/decade FETs have been demonstrated using PZT albeit with a large hysteresis [27]. Other ferroelectric materials such $\mathrm{HfZrO_x}$ are also being explored for hysterisis-free NCFETs [100, 66].

Confined Structures

The other approach to facilitate scaling is the use of non-classical CMOS structure, combined with the use of new materials for the gate stack in order to improve the electrostatic integrity of the transistors, thus enabling close to 60 mV/decade switching. The working theory of the above devices is to increase the effective gate capacitance relative to other capacitances in the system by confining the channel material along the thickness direction. As C_D and C_S in fig. 1.2 are both proportional to t_{ch} (thickness of the channel), the electrostatic short channel effects are inversely proportional to t_{ch}. Thus, ultra-thin body (UTB) SOI MOSFET were proposed consisting of very thin ($<$ 10 nm) Si channels to improve the electrostatics [30]. Other approach included the use of multi-gated devices such as double-gated transistors [22, 126] and gate-all-around devices [102] to further improve gate control on the electrostatic potential. One of the main challenges in any of these confined device concepts is to achieve a channel thickness significantly lower than the channel length [103], especially with bulk materials like Si where the mobility can fall precipitously for thin channels [34]. Thus, there was an increased need to work with materials which have a naturally thin structure which could reduce the effects of surface scattering and dangling bonds in these highly confined structures. There was a big push towards 2-D materials with the discovery that single layers of graphene could be exfoliated using a scotch tape [88] to produce a sheet of one atomic layer thick crystal. This along with with large estimate of graphene's mobility of up to $200,000\text{cm}^2/\text{V.s}$ [13] led to a thrust in research on Carbon-based electronic devices including carbon nanotube devices and graphene nanoribbons [9, 90, 68]. The main drawback of carbon based technology has been the inability to reliably produce a band gap leading to large band gap and device variability, as there is no band gap for an infinite sheet of graphene. Recently, another class of 2-D materials, namely the transition metal dichalcogenides (TMD) were explored for application in the transistor industry because of their naturally confined structure and the existence of a non-zero band gap in few-layer devices. We'll look into these materials in great detail in this thesis, while trying to understand their electronic structure and estimate their performance in low power applications.

1.3 Transition Metal Dichalcogenides - Application in Low Power Electronics

The first single-layer MoS_2 transistor was fabricated by B Radisavljevic in 2011 [93] where they showed ON/OFF ratio of up to 10^8 with a very low OFF current as shown in Fig. 1.3. This is possible because of the large band gap (1.8 eV) of MoS_2 when exfoliated as a monolayer resulting in inhibition of both direct S-D tunneling and band-to-band tunneling. This result led to an increased interest in the class of TMD materials which can similarly be exfoliated into few-layers to be used in the channel of materials. Subsequently, transistors were demonstrated using other TMDs such as $MoSe_2$ [62, 16], WS_2 [14, 50] and WSe_2 [26, 70]. As TMDs have a direct band gap in their monolayer phase, they have also found applications in optoelectronics, photovoltaics and photodetection, while their layered structure shows promise in high performance flexible electronics[105, 40, 65, 12, 44]. In most of the fabricated transistors, a high ON/OFF ratio has been observed because of the large band gap in its monolayer and few-layer crystalline phases, while the major drawback has been a significant contact resistance which reduces the effective mobility, thus limiting the ON current and drivability. K. Ganapathi investigated the drivability and ON/OFF ratio of scaled monolayer MoS_2 using an effective mass approach realizing that an ON/OFF ratio of 10^10 was achievable with SS close to 60 mV/decade and low DIBL [125]. We try to answer some other questions pertaining to ultimately scaled few-layer transistors in chapters 3-5 using an in-house quantum transport simulator (described in chapter 2). In chapter 3, we analyze four different TMDs -MoS_2, $MoSe_2$, WS_2 and WSe_2 to understand how the different electronic structure of different TMDs affect the transistor properties of the same. Another interesting



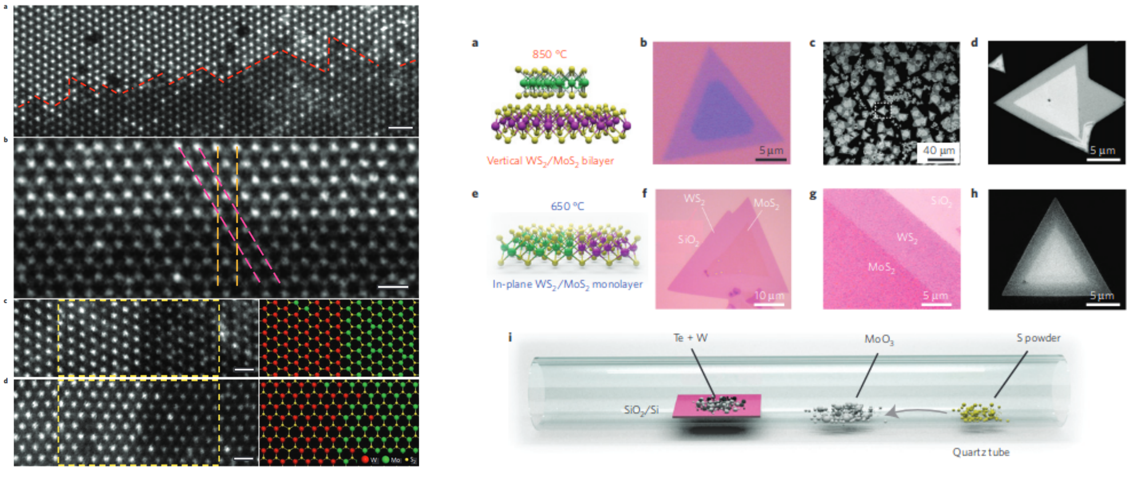
(a) (b)

Figure 1.3: a) Schematic of single-layer MoS$_2$ transistor implemented by Radisavljevic et al., and b) transfer characteristics for different drain biases and back-gate bias as a function of top-gate voltage. The channel thickness is 0.65 nm which is the thickness of a single-layer of MoS$_2$. An ON/OFF ratio of up to 10^8 is observed in the transfer characteristics. Adapted from [93]

question asked in the study was what's the optimum number of layers for transistor made of these few-layer materials. As the number of layers are increased from monolayer to few-layer, the electronic structure of the TMDs changes resulting in change of the curvature of the conduction band as well as relative positions of the valleys contributing to transport in the energy domain [69, 76]. Both these effects significantly affect the density of states and velocity of the particles contributing to transport, thus determining the transistor behavior. In addition to that, there was an electrostatic question which needed to be answered with regards to changing the number of layers - Increasing the number of layers adds the number of modes that could could contribute to current, while the gate control significantly diminishes with increase in channel thickness. Hence, it was important to understand the trade-off between the two, which is discussed in the chapter 3. In chapter 4, we study one TMD, MoS$_2$ in more detail while using a more rigorous tight-binding Hamiltonian for 1-3 layer transistors. The previous chapter, while using an effective-mass approach ignores the interaction between different valleys and the non-parabolicity of the electronic structure. Here, all the relevant physics pertaining to transport is included by using a sp^3d^5 orbitals with spin-orbit coupling to look at the electrostatics and transfer characteristics of few-layer MoS$_2$ transistors. In chapter 3, we used metal Schottky contacts while using doped contacts in chapter 4. We also realize the importance of doping concentration of the contacts on the transistor performance (both SS and ON current) of TMDs. Thus, there is a crucial need to look at the contact resistance in these materials and a lot of recent research has been focused towards obtaining low resistance contacts to TMDs [89, 24, 117, 32, 23]. It was still not understood what the intrinsic limit to contact resistance (as determined by the electronic structure assuming ideal contacts) of TMDs, which we'll attempt to study in detail in chapter 5 of the thesis. Overall, the above three chapters help in understanding the potential of confined materials in replacing Si for future low operating power applications, while providing insights on design constraints that'd govern their implementation.

Other Novel Applications

As TMDs gained prominence in the academic community, studies were done to realize novel device geometries out of monolayers. These included the development of both lateral and vertical heterostructures using CVD (Chemical Vapor Deposition) [39] (Fig. 1.4) and using MBE (Molecular-Beam Epitaxy) [114] or direct mechanical transfer [37]. Tunneling diodes have been made out of the vertical heterostructures showing NDR (Negative Differential Resistance) thus demonstrating possibility of band-to-band tunneling through the vertical van der Wals heterostructures of MoS_2/WS_2 [96]. Other interesting applications can be realized from these heterostructures given they are confined along one direction. One phenomenon is described in chapter 6, where an in-plane electric field perpendicular to a lateral heterostructure of 2-D monolayers like the one shown in Fig. 1.4 is used to generate a 1-D spin channel along the junction by spin-orbit coupling. Therein, the confinement in the thickness direction prevents the spins from having a momentum away from the channel. Another device



(a) (b)

Figure 1.4: a) Illustration of lateral heterojunction between WS_2 and MoS_2 heterostructures and b) schematic of the synthesis of vertical stacked and in-plane MoS_2/WS_2 heterostructures. The growth of in-plane (650°C) and vertical (650°C) heterostructures depends on the temperature at which the synthesis is carried out. (a) shows the atomically sharp lateral junction of the two TMDs. Adapted from [39]

is proposed in chapter 7 where we modulate the flux of tunneling current flowing through a vertical heterostructure or junction using a magnetic field to switch between two bias points of an Esaki diode. The above device can provide magnetoresistance (MR) of up to 10^2, which could find use in the information storage industry as the search for large MR devices to replace GMR (Giant Magnetoresistance) read/write heads continues.

Chapter 2

An Overview of TCAD and the Quantum Transport Simulator

2.1 Introduction

As transistors were developed and the integrated circuit (I.C) grew as a major industry, there was an increased need to understand the working principle of the various components of an I.C so that the I.C design could be optimized in order to provide th e maximum performance/cost. TCAD can inexpensively provide insight while exploring alternative material and device geometry, hence accelerating innovation. There are three major components to simulating the performance of an I.C - 1) Process , 2) Device and finally 3) Circuit simulation. I'll briefly go through each of these TCAD (Transistor Computer Aided Design) tools in the subsequent sections while going into more detail about the device simulation which is involved in every study in this thesis.

2.2 TCAD Hierarchy

Process Simulation

Semiconductor process simulation is used to model the fabrication of a semiconductor device. This is usually the first step in the bottom-up approach to TCAD flow, where the transistor fabrication is first simulated before the device characteristics are extracted using a device simulator. Even though major advances have been made in device simulations, process simulation has usually lagged behind as sophisticated and super scaled devices were developed [63]. The fabrication of a transistor involves a number of steps called the process flow, each of which are then modeled by the process simulation. An example of a process flow can be seen in Fig. 2.1, where a stress-engineered FINFET [42] is modeled using Sentaurus Process. The various steps involved in a process flow can be ion implantation for doping, annealing for diffusion of dopant and their activation, etching, deposition and oxidation. The primary goal of process simulation is to look at relative trends to guide the fabrication of devices. Currently the major process simulators in the semiconductor industry is the Sentaurus Process from Synopsys and ATHENA/VICTORY Process from SILVACO. As the scaling continues, it has become crucial to look at imperfections such as random dopant fluctuations and edge dislocation stress originating from process flow and much recent work has gone into modeling it [106, 121]. Other challenges that exist in process modeling includes predicting active substitutional dopant and predicting the doping profile [35] and the interaction of damage from implantation with the dopant [112]. An increased use of atomistic simulation tools and ab initio methods is also being seen to model process flow with fewer parameters which could result in more physical models [1]. Once the device geometry and doping profiles are set using process simulation, we move to device simulation to characterize the electrical performance of the designed transistor.

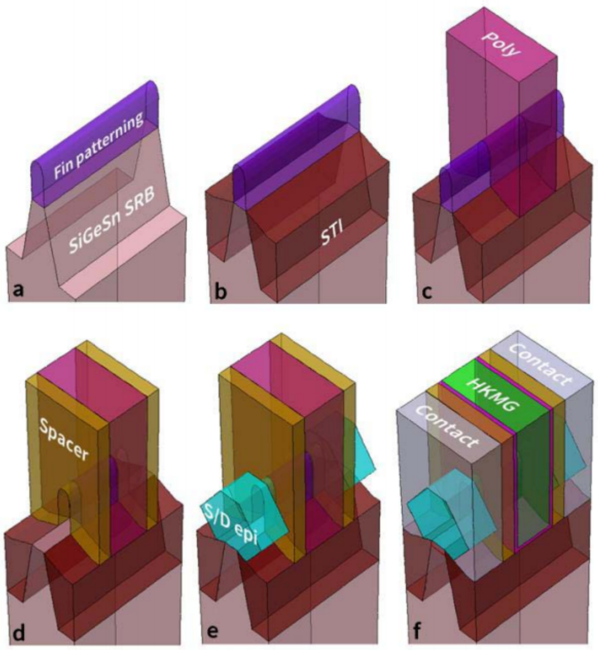


Figure 2.1: An example of a MOSFET process flow for a 7 nm stress-engineered FINFET which involves a) Fin patterning, b) STI (Shallow Trench Isolation) formation, c) Polygate definition, d) Spacer deposition and S/D patterning, e) epitaxial growth of doped S/D and finally f) Poly gate removal and high-K/metal gate formation followed by S/D contact formation. The above process flow was simulated using Sentaurus process simulation. Adapted from [42]

.

Device Simulation

Once a device is designed using a process simulation or a new device is conceptualized, extensive device simulation needs to be carried out to predict the performance and decide on design criteria. The need for device simulation has increased dramatically over the last couple of decades as fabrication of semiconductor devices has increased in complexity and it has become prohibitively expensive as the primary means to predict performance trends. Semiconductor physics was a major driver of condensed matter physics when the first transistor was invented [101]. A semi-classical approach to simulating device behavior was developed, namely the drift-diffusion equations were developed to describe charge transport[109, 4]. In this formalism, the electrons and holes were considered to be particles that move through a device driven by electric field and diffusive processes.

$$\frac{d\hbar\vec{k}}{dt} = -\nabla_r E_c(\vec{r}, t) + F_s(\vec{r}, t) \tag{2.1}$$

where $-\nabla_r E_c(\vec{r}, t)$ is the drift component due to electric field and $F_s(\vec{r}, t)$ is the diffusive part of the transport. $\hbar k$ gives the momentum of the particle while $E_c(\vec{r}, t)$ is the bottom of the conduction band at position $\vec{r}$ and time t. The above equation can be solved using Monte-Carlo methods [49] or by using a probability function which brings us to the Boltzmann Transport Equation.

$$\frac{\delta f}{\delta t} + \vec{v}.\nabla_r f - \frac{q\vec{E}}{\hbar}\nabla_k f = \hat{C}f \tag{2.2}$$

where $f(\vec{r}, k, t)$ is the probability distribution function of the electrons, $\vec{E}$ the electric field and $\vec{C}f$ describes the effects of scattering. f reduces to the Fermi distribution in equilibrium while the BTE has to be solved in non-equilibrium conditions to calculate the distribution function. Interestingly, the first moment of the distribution function in the Boltzmann equation gives the drift-diffusion equation [4]. The above equations were sufficient to model devices when the electric fields in the conducting channel were not very large. As the devices were scaled, higher electric fields (¿10^4V/cm) led to velocity overshoot leading to velocity of the electrons being higher than saturation velocity leading to the break down of drift equations. This led to the development of hydrodynamic equations to model kinetic energy dependent mobility instead of field dependent mobility. Both these models break down as devices are scaled to the ultimate limit where the electron correlation becomes significant enough that the wave nature of the electrons have to be considered leading to the development of quantum transport formalism. The predecessor to quantum transport was ballistic transport or the TOB (Top of the Barrier) model. In this case, the channel is small enough that no scattering event is assumed to happen as it travels from the source to drain. This theory works on the idea of calculating the average injection or thermal velocity of the charged particle and the charge density contribution from the source and drain respectively. The above quantities are usually estimated from the bandstructure of the material and the electrochemical potential at the contacts, and thus the wave description of the electrons and holes are introduced.

$$\begin{aligned} I &= v_{inj}(n_{top,S} - n_{top,D}) \\ n_{top,S(D)} &= \int dEDOS_{S(D)} f_{S(D)}(E) \\ f_{S(D)} &= \frac{1}{1 + e^{\left(\frac{E - E_{F_{S(D)}}}{k_B T}\right)}} \end{aligned} \tag{2.3}$$

Where DOS is the density of states at the top of the barrier and E_F is the Fermi potential at the source and drain. Eventually, a quantum description of charge transport was developed using the Non-Equilibrium Green's Function formalism (NEGF) to compute the electron and hole correlation across the whole device [4, 29, 28]. The top of the barrier equations described above then reduce to

$$\begin{aligned} I &= \frac{2e}{h} \int dET(E)(f_S(E) - f_D(E)) \\ n(x) &= \int dE(LDOS_S f_S(E) + LDOS_D f_D(E)) \end{aligned} \tag{2.4}$$

where $LDOS(E)$ is the local density of states and $T(E)$ is the transmission or mode density which is derived from the NEGF equations. I'll go into it in further detail when I describe the NEGF formalism. After the transfer characteristics are derived using either the quantum transport equations or the semi-classical equations, compact models can be used to describe the behavior of the transistor in a circuit to a considerable degree of accuracy for circuit simulation, thus avoiding the computationally intensive device simulation.

Circuit Simulation

Finally, circuit simulation is the top level in the heirarchy of semiconductor device modeling and is used to check the circuit designs while predicting the behavior of the designed circuit. The most popular program for doing the same is SPICE (Simulation Program with Integrated Circuit Emphasis) which has become the worldwide standard for circuit simulation[83]. SPICE was derived from the CANCER (Computer Analysis of Nonlinear Circuits, Excluding Radiation) program at the University of California, Berkeley[82]. SPICE and it's newer versions includes many semiconductor device models as well as analytical compact models which are derived from the Device simulation. The primary goal of circuit simulation is to accurately predict the behavior of a large I.C while keeping a small turnaround time between design and analysis. Some of the tasks done by circuit simulations include AC analysis, DC quiescent point analysis, noise analysis and transient analysis.

2.3 Quantum Transport and NEGF

Quantum Transport calculations is the major tool that has been used in this thesis as a means to evaluate the performance of novel materials at ultra-scaled dimensions to understand how effective they might be in replacing Silicon. For a conventional macroscopic conductor, the conductance is inversely proportional to it's length and directly proportional to its area.

$$G = \sigma A/L \tag{2.5}$$

This equation would suggest that as the length of a conductor is shortened, the conductance should tend to infinity. This is not true and there is a maximum value of conductance for one energy level in the device -

$$G_0 = q^2/h = 38.7E - 6S \tag{2.6}$$

For most systems a factor of 2 is multiplied to the above expression to account for the two spin states that can reside in a given energy level. Thus if there is a way to calculate the number of energy levels that'd contribute to current for a given electrostatic and non-equilibrium conditions, then we can calculate the total current flowing through the device. The essence of any quantum transport calculation is thus two fold as shown in Fig. 2.2.

1) Calculate the charge in the system by counting the number of energy levels in the system that are occupied according to the Fermi distribution. Self-consistency is achieved by solving it with the Poisson's equation to solve for the electrostatic potential across the device (Sec. 2.4). This is described by the loop in Fig. 2.2.

2) From the converged electrostatic state of the system calculate the total number of modes contributing to current flow in order to calculate the drive current of the transistor.

We'll do the above two calculations within the NEGF formalism which we'll discuss in the following section.

NEGF

The NEGF formalism to simulate transport was developed Keldysh in 1965 [52] although it started getting used by device simulation group as a standard means of understanding quantum transport in early 21$^{\text{st}}$ century [28, 60, 57]. In this formalism, we calculate the impulse response (Green's function) of the semiconducting channel of the device to scatterers from the reservoirs, where the effects of any scatterer or reservoir is included through the use of self-energy matrices. Fig. 2.3 gives a simple overview of the quantum transport problem.

A device is placed between two semi-infinite reservoirs, with different electrochemical potentials. The rate of electron flow between a reservoir and the channel is given by

$$q/\hbar\gamma_{1(2)}(f_{1(2)} - N) \tag{2.7}$$

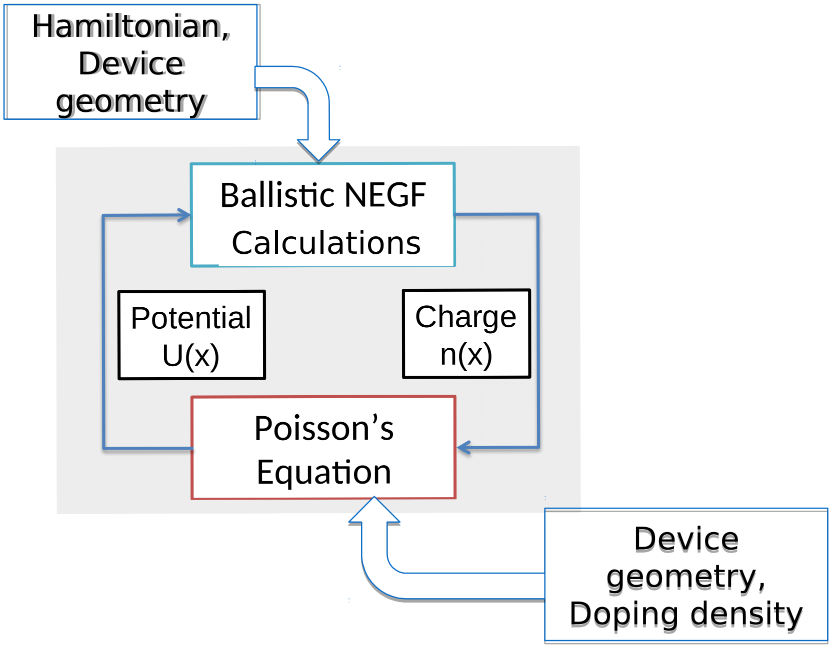


Figure 2.2: The process of a quantum transport calculation. There are two computational components of a quantum transport solver - the NEGF calculations and the Poisson’s equation. The NEGF calculations take into consideration the device geometry, material and other scattering mechanisms to compute the charge and current for a given electrostatic conditions. The Poisson’s equation involves solving a Laplacian taking into consideration the device geometry, dielectric properties of the materials and the charge calculated by the NEGF calculation to give an electrostatic potential. The above two calculations are done one after the other to achieve electrostatic self-consistency. Finally the current is calculated from NEGF equations using the converged potential.

.

The net rate of current flow is thus given by

$$I = q/\hbar \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} (f_1 - f_2) \tag{2.8}$$

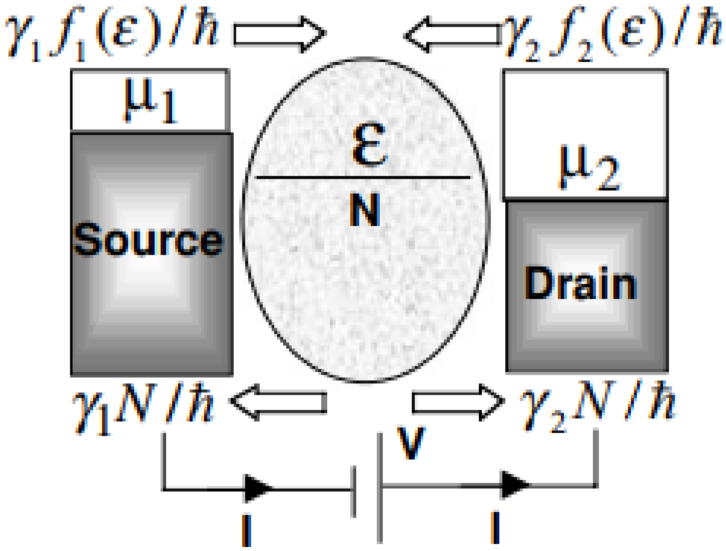


Figure 2.3: A toy example illustrating the fundamentals of non-equilibrium physics. A device is placed between two semi-infinite reservoirs, with different electrochemical potentials. Thus the rate of electrons flowing into the channel from either reservoir is different resulting in an non-equilibrium and thus current flow ensues. An even more general case exists when the coupling to the two contacts are different resulting in different flow rates to the reservoirs. Adapted from Quantum Transport: Atom to Transistor [29].

.

while the steady state charge N is given by

$$N = \frac{\gamma_1 f_1 + \gamma_2 f_2}{\gamma_1 + \gamma_2} \tag{2.9}$$

The above equations are valid in the case that the device has only a single energy level, while the result can be generalized to a multi-level device by using matrices instead of numbers to describe the device. In the above case γ is a measure of broadening of energy levels in the device when it is brought into contact with the reservoir. For a multi-level system the following equations describe the quantum transport:

$$\begin{aligned} G(E) &= (EI - H - \Sigma_1 - \Sigma_2)^{-1} \\ A &= i(G - G^\dagger) \end{aligned} \tag{2.10}$$

where $\Sigma_{1(2)}$ is the self-energy matrix corresponding to scattering from the left(right) contact, H is the Hamiltonian of the device section and I is an identity matrix of size equal to the Hamiltonian. The matrix $A(E)$ is the spectral function of the device and gives the density of states through the following equation:

$$D(E) = A(E)/2\pi \tag{2.11}$$

while G is the Green's function for the device. The broadening matrix for a multi-level system is given by

$$\Gamma_{1(2)}(E) = i(\Sigma_1 - \Sigma_2) \tag{2.12}$$

Any other scattering processes like optical phonon scattering, acoustic phonon scattering is included through self-energy matrices too. The electron and hole correlation function is calculated from the broadening matrix, Fermi function and the Green's function by the following:

$$\begin{aligned} G^n &= G\Gamma_1 G^\dagger f_1 + G\Gamma_2 G^\dagger f_2 \\ G^p &= G\Gamma_1 G^\dagger (1 - f_1) + G\Gamma_2 G^\dagger (1 - f_2) \end{aligned} \tag{2.13}$$

The electron and hole density is calculated from the above equation by summing over the energy levels.

$$\rho = \int \frac{1}{2\pi} G^n dE \tag{2.14}$$

Finally, the current is given by

$$I = q/h \int Trace(\Gamma_1 G \Gamma_2 G^\dagger)(f_1 - f_2) dE \tag{2.15}$$

In the current equation $Trace(\Gamma_1 G \Gamma_2 G^\dagger)$ is also called the transmission ($T(E)$) as it counts the number of modes that could contribute to current at a given energy level. This is the same density of modes that I described in the section describing quantum transport. $f_1 - f_2$) gives the probability that the modes at a given energy is filled up by only one of the contacts. In other words, the modes at a given energy would not contribute to current if it were populated at both source and drain. Thus, there is no current flow at equilibrium, when both contacts have similar probability of filling up any energy level.

The self-energy matrix $\Sigma_{1,2}$ is one of the major contributions of the NEGF formalism, as it has reduced the problem of an infinite contact to just one matrix of the same size of one layer of the Hamiltonian. NEGF works on the premise that you can break down the device into equivalent slices along the channel. Each of these slice could be represented by a matrix of size n, while there might be N_P number of slices along the channel. The self-energy matrix $\Sigma_{1,2}$ is reduced to a size of nXn instead of being infinity by using the property that the contacts are reservoirs at equilibrium with a constant potential on it. Thus, we compute self-energy using the surface Green's function as $\tau g \tau^\dagger$ with τ being the connection between the reservoir and the device and g the Green's function of the contact. Solving for the above iteratively could be tedious computationally and an effective layer approach as described in[99] is implemented to exponentially reduce compute time.

Finally, the most computationally intensive part of the calculation is calculating the Green's function of the entire device as it involves calculating the inverse of an $nXNP$ matrix. Thus, if rudimentary matrix inversion is applied, the above inversion scales as $O(n^3N_P^3)$. Two different algorithms are popular for the above calculation, the Recursive Green's Function (RGF) algorithm and the Nested Dissection (ND) algorithm. Both algorithms exploit the sparsity of the the matrix $(EI - H - \Sigma_1 - \Sigma_2)$. As each layer of the device interacts with only adjacent layers, the above matrix is block tri-diagonal. Also, because the broadening matrix γ has only $n{*}n$ non-zero elements, we don't need to calculate all elements of G to compute the correlation functions and the transmission function - only the diagonal blocks are needed[4]. Therefore, a Recursive Green's Function algorithm is used to calculate the appropriate entries of the matrix G [4, 41]. The algorithm works on the basis of the Dyson's equation which provides a computationally cheap way to find diagonal terms of the inverse of a block tri-diagonal matrix. The algorithm then computes the left-connected Green's function and then the right connected Green's function thus completing the inversion to compute the electron and hole correlation functions. The computational complexity of the above inversion thus reduces to $O(n^3N_P)$, thus making it particularly suitable for long channel devices.

The other approach for calculating the Green's function is the Nested Dissection method ([46]) which works by partitioning the matrix $(EI - H - \Sigma_1 - \Sigma_2)$ so that three matrices recursively, a left and right matrix (L,R) and a separator matrix (S) such that the left and right matrices don't have any interaction with each other. Thus, a tree is made by subsequently dividing the matrix until it cannot be divided anymore. The method can be optimized by setting the size of the smallest matrix to be partitioned. The algorithm then uses the LDL^T decomposition of a rearranged matrix to calculate the Green's function. An illustration of the algorithm is given in Fig. 2.4. Even though ND ($O(n^2N_P)$) is asymptotically better than the RGF, there is a significant computation overhead which makes the RGF calculation more efficient unless a system with large enough n is simulated. Even though both the algorithms were implemented by us, RGF was used for all the calculations presented in this paper which used NEGF calculations.

Hamiltonian

If we look at the matrix to be inverted $(EI - H - \Sigma_1 - \Sigma_2)$, it includes, the information about various scattering processes though self-energy matrices, the energy E and the Hamiltonian of the device H, which includes the most important physics that would determine transport characteristics. In this work we have used two ways to describe the Hamiltonian of the material, the effective mass Hamiltonian and the Tight-binding approach.

The effective mass approach is the simplest Hamiltonian used to describe a material. The Hamiltonian describing a certain valley is given by the following:

$$H(\vec{r}) = -\hbar^2/2\left[\frac{d}{dx}\left(\frac{1}{m_x}\frac{d}{dx}\right) + \frac{d}{dy}\left(\frac{1}{m_y}\frac{d}{dy}\right) + \frac{d}{dz}\left(\frac{1}{m_z}\frac{d}{dz}\right)\right] + eV(\vec{r}) \qquad (2.16)$$

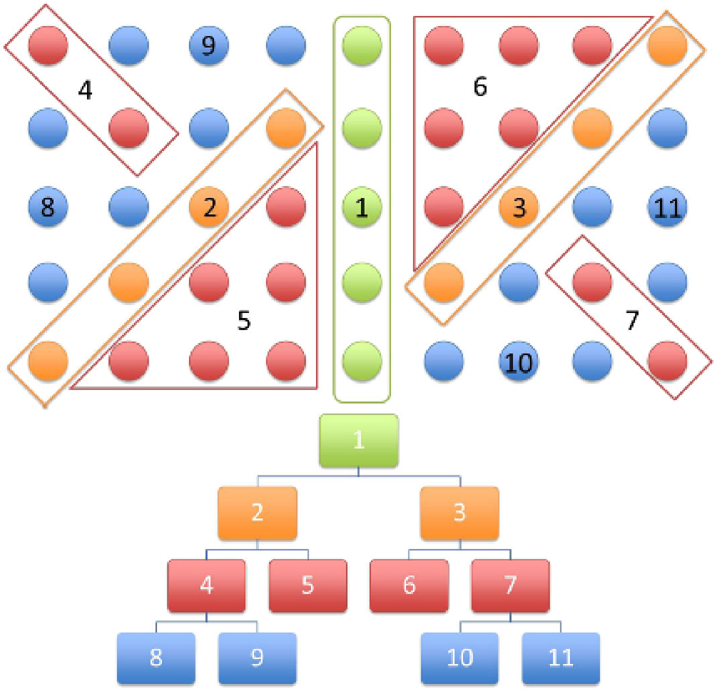


Figure 2.4: An example of the partition process using METIS to be used in the Nested Dissection approach. The first and last columns are dense matrices because of the self energy terms and thus cannot be partitioned into smaller parts. A tree is formed as shown in the figure to rearrange the matrix to apply the LDL^T decomposition. Adapted from [46]

.

where m_x, m_y and m_z are the effective masses along the given directions and $V(r)$ is the position dependent electrostatic potential. The Hamiltonian can then be discretized over the lattice of the material in order to be implemented numerically.

$$H_{m,m'} = -t_{(m_x,m_y,m_z)} I_s \tag{2.17}$$

where $t_{(m_x,m_y,m_z)} = \frac{\hbar^2}{2m_{(x,y,z)}a}$. The hopping term corresponding to the hopping in the correct direction is used in the above equation. If indices m and m' are adjacent in the x direction, then we use $-t_x$ for the Hamiltonian for that interaction. Sometimes, a sum of all interaction is negated and added to the on-site term so that the conduction band is at $0eV$.

The tight-binding approach is a more generalized version of the effective mass approach. This was developed in the early 20th century and formalized by Slater and Koster to be applied to crystals [104, 115, 43]. In this case, we consider a Linear Combination of Atomic Orbitals (LCAO) while considering all the the atomic orbitals to have zero overlap with each other. The size of the basis is given by the product of the number of atoms in a basis cell and the number of orbitals on each atom. The orbitals in its periodic form is then given by

$$\phi(r) = \Sigma_{i=[1,a],j=[1,m]} c_{i,j} \phi_{i,j}(r - b_i) \tag{2.18}$$

where i goes over the number of atoms in the unit cell (3 for MoS_2) and j runs through the number of orbitals per atom (9(sp_3d_5) * 2(spin) for MoS_2). b_i is the atomic location of each atom in the unit cell. The above ansatz is made periodic using the similarity transforms of the corresponding crystal structure and then using Bloch's equation, we get.

$$\Psi(r) = \Sigma e^{ikR_n} \phi(r - R_n) \tag{2.19}$$

where R_n runs over all the lattice vectors in the crystal lattice. Now using the orthogonality of two different atomic orbitals and the Schroedinger's equation, we get

$$H\Psi(r) = E\Psi(r) \tag{2.20}$$

Taking an inner product with $\phi_{i,j}(r - b_j)$, we get

$$\begin{aligned} \int \phi^*_{i,j}(r - b_j) H \Sigma e^{ikR_n} \phi(r - R_n) &= Ec_{i,j} \\ \Sigma_{i',j'} \left[\int \phi^*_{i,j}(r - b_j) H \Sigma e^{ikR_n} \phi_{i',j'}(r - b_{j'} - R_n) \right] c_{i',j'} &= Ec_{i,j} \end{aligned} \tag{2.21}$$

The secular equation above is used to construct the Hamiltonian for the material. As we calculated above the basis size for the TB Hamiltonian of MoS_2 or other TMDs is 54 as used in Chap. 4. On the other hand, if we use the effective mass approach as used in Chap. 3, the basis size is 2 (including the K point and Σ_{min} points). But in that case the degeneracy of each point has to be considered.

Another important concept in treating the Hamiltonian is the mode-space approach. Say, the above Hamiltonian, can be written as the following:

$$H = E_0 + H_1 e^{ik_x a} + H_1^\dagger e^{-ik_x a} + H_2 e^{ik_y a} + H_2^\dagger e^{-ik_y a} + H_3 e^{ik_x a + ik_y a} + H_3^\dagger e^{-ik_x a - ik_y a} \tag{2.22}$$

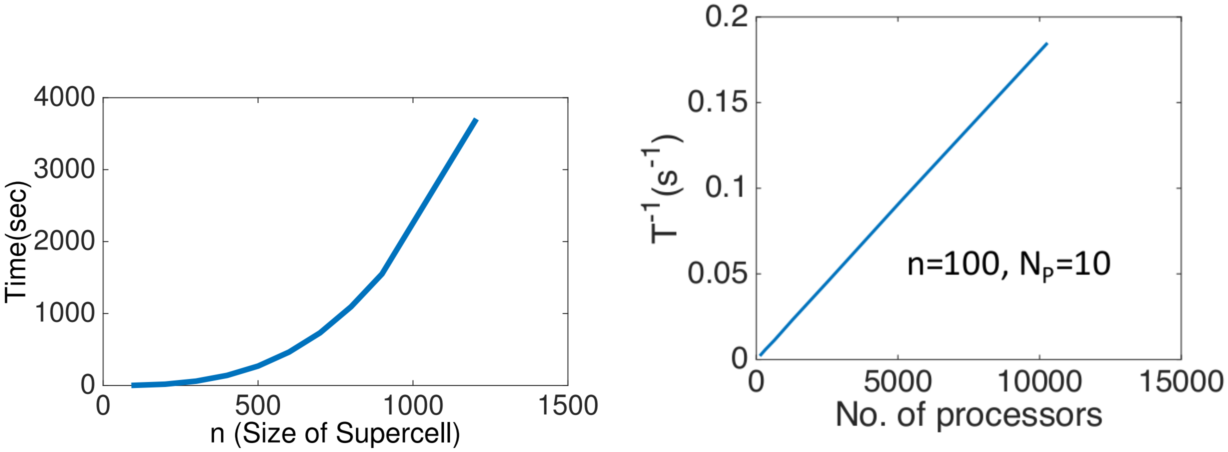
The above equation can be reduced to a k_y dependent Hamiltonian

$$H(k_y) = (E_0 + H_2 e^{ik_y a} + H_2^\dagger e^{-ik_y a}) + (H_1 + H_3 e^{ik_y a}) e^{ik_x a} + +(H_1 + H_3 e^{ik_y a})^\dagger e^{-ik_x a} \tag{2.23}$$

So, $(E_0 + H_2 e^{ik_y a} + H_2^\dagger e^{-ik_y a})$ is the on-site Hamiltonian and $(H_1 + H_3 e^{ik_y a})$ is the Hopping term. This allows us to break down the problem to solve for different energy levels as well as different transverse modes which helps in parallelizing the Quantum Transport Calculation.

Parallelization

Even though the calculation involved for each energy level is very involved, NEGF equations have a significant opportunity for parallelism to reduce the computation time. As the Hamiltonian is a function of energy and transverse momentum, while the DOS and mode density are also a function of the same, the problem is embarrassingly parallel. A given compute node can calculate the physical quantities for a set of energy levels and transverse momentum while they can be summed over at one node for the rest of the self-consistent calculation. We used MPI (Message Passing Interface) to parallelize over energy. The code was run at NERSC (National Energy Research Scientific Computing Center) on the Hopper and Edison machines and also in the in-house clusters (The in-house cluster had 256 CPU cores distributed between 32 nodes). Because inverting the block tridiagonal matrix and computing the Green's function is so computationally intensive, the communication is almost never a bottleneck and the code scales very well (Fig. 2.5).



(a)

(b)

Figure 2.5: Scaling studies on the implemented Quantum Transport Calculator. a) shows the scaling of RGF for a problem with fixed number of particles in the channel ($N_P = 10$) with n ranging from 10 to 1200. The compute nodes at Edison start running out of memory after that. The figure shows the cubic dependence of the algorithm on the basis size. b) shows the scaling of the algorithm as the number of processors are increased for a problem with a given number of energy points. The slope of the plot when normalized is very close to 1 showing the scalability.

.

2.4 Poisson's Equation

The second half of the transport calculation is solving for the Poisson's equation which is used to calculate the electrostatics for a given charge distribution. More often than not, a lot of time may be spent on looking at the NEGF equations and the Hamiltonian, while the transistor/device might be electrostatics dominated and lead to most of the interesting physics that's seen in the results. For the Poisson's equation, we solve the Laplacian

$$\vec{\nabla}.\vec{D} = -\frac{\rho}{\epsilon} \tag{2.24}$$

which can also be written as

$$\vec{\nabla}.\epsilon\nabla U = q(n - p + N_A^- - N_D^+) \tag{2.25}$$

where $\vec{D}$ is the displacement field, ρ is the charge density, ϵ is the dielectric constant, n the free electron density, p the free hole density,N_A^- the negatively charged ionized acceptors and N_D^+ the positively charged ionized donors. In order to apply it numerically we discretize it over a 2-D rectangular grid for most of the simulations done in this work.

$$\begin{aligned} \epsilon(i,j) * \left(\frac{U(i+1,j) + U(i-1,j) - 2U(i,j)}{a_x^2} + \frac{U(i,j+1) + U(i,j-1) - 2U(i,j)}{a_y^2}\right) \\ + \left(\frac{(\epsilon(i+1,j) - \epsilon(i-1,j))(U(i+1,j) - U(i-1,j))}{4a_x^2}\right) + \left(\frac{(\epsilon(i,j+1) - \epsilon(i,j-1))(U(i,j+1) - U(i,j-1))}{4a_y^2}\right) \\ = q(n(i,j) - p(i,j) + N_A^-(i,j) - N_D^+(i,j)) \end{aligned} \tag{2.26}$$

where i and j gives the index along the x and y direction respectively and a_x and a_y is the discretization size along the two axes. Proper care has to be taken to use the correct boundary condition for the Poisson's equation. A Neumann boundary condition is used when field lines go to zero, i.e, the potential at two consecutive points perpendicular to the direction is the same [95]. This is used in the case of the boundary between the insulator and air, and when doped contacts are used. Dirichlet boundary conditions, on the other hand are used when the potential at a certain point is fixed. This usually happens when there is a contact to a metal at any point, such as the gate and if metal contacts are used at the source and drain electrodes. For most purposes, doing a simple matrix inversion is computationally simple enough to compute the potential at each point.

Convergence Strategies

Once the solution from the Poisson's equation is calculated, the result has to be mixed with the older guess to predict a new guess for the electrostatics of the device. One approach is to do a linear mix

$$U_{new} = U_{old} + \alpha * (U_{sol} - U_{old}) \tag{2.27}$$

where α is the mixing parameter and U_{sol} is the solution from the Poisson solver. α might need to be changed as the simulation progresses. Usually the device is simulated for one bias point and then progressively solved for small increments of bias voltages. For each change in bias voltage, the potential might need to be coarsely adjusted first and then fine-tuned as the solution gets closer to convergence. The mixing parameter might need to be large for the first few iterations and successively reduced in magnitude. Another approach is to use quasi-Fermi approximations to accelerate convergence [95]. Here an inner loop is used to transform the charge density to quasi-Fermi potential using a semi-classical DOS based approximation.

Another approach to the same problem is to calculate the new guess based on past few iterations instead of only one. This helps in convergence especially in the ON state when a lot of charge is accumulated in the channel and the simulation can runaway and become unstable. Pulay mixing or DIIS (Direct Inversion in the Iterative Subspace) can be used for the same purpose though it can be unstable [59]. We used an extension of the linear mixing where we mixed different weights of previous 4 or 5 iterations. One interesting thing to be noted is that if the calculations are done for the same S-D bias voltage and different gate voltage, the mixing in the doped regions can be taken to be close to zero as the potential at the contacts do not move much for different gate bias. Thus most of the mixing can be done only in the channel region which could help accelerate convergence, especially in the Laplacian dominated regime (OFF state) by using larger mixing parameter selectively. The same can be done when only the drain bias is changed while the gate bias is kept constant. In this case, the mixing parameter is maximized in the drain region.

2.5 Conclusion

In this chapter, we discussed the working and theory behind the quantum transport simulator that is used in this study. The theory is used extensively in Chap. 3,4 and 6. In Chap. 3, an effective mass model is used to treat four different TMD materials for different number of layers to evaluate its performance trade-offs. In chap. 4, we discuss transistors based on MoS_2 in detail using a TB Hamiltonian and look at the effects on its performance due to screening. In chap. 6 we look at how to generate 1-D spin channels in the heterojunction of two different TMDs by using an effective mass Hamiltonian. In the rest of the chapters, the ideas of quantum transport is still applied, even though NEGF is not used for the current calculations. Having developed this tool, we'll now look at different TMDs and their potential application in low power electronics.

Chapter 3

Dependence of Intrinsic Performance of Short Channel TMD Transistors on Materials and Number of Layers

3.1 Introduction

In this chapter, we'll discuss how good 2-D materials, more specifically transition metal dichalcogenides could be in low power electronic applications while using an effective-mass approach to treat the Hamiltonian. Weak inter-layer interactions in TMDs allow thin flakes to be exfoliated from their bulk crystals, thus facilitating the fabrication of layered devices. Layer-dependent properties of TMDs such as transition from an indirect to direct band gap system on moving from bulk to monolayer $\mathrm{MoS_2}$, combined with an increase in the gap from 1.3 eV to 1.8 eV, have attracted attention in optoelectronic applications, where unlike graphene based layered FETs, which suffer from high leakage currents because of zero band gap, TMDs have gained interest as potential candidates for future low-power, flexible and transparent electronic devices [74, 125, 119]. Structurally confined TMD-monolayers offer excellent electrostatic control of the channel, resulting in better scalability. However, because of reduced dimensionality, the number of modes available for transport might be limited. While this limitation could be mitigated using multi-layer TMDs to deliver more current, it comes at the cost of reduced electrostatic control. This trade-off between electrostatic control and number of modes poses an interesting question: What is the optimum number of layers to be used for ultra-scaled TMD transistors? Leakage currents from direct source-to-drain tunneling could also become an important concern at these ultimately scaled limits, which would determine their viability in low power applications. Here we will present a comparative study of four TMDs starting from first-principles material-physics and subsequent self-consistent ballistic quantum transport simulations that addresses this question within the context of sub-10 nm channel length transistors. Our results show that relatively large effective mass along with a large band gap in TMDs inhibits source-to-drain tunneling even at extremely small gate lengths (5 nm), thus demonstrating that TMD materials can be scaled down with excellent ON/OFF ratios. The degradation in switching capabilities with increasing number of layers emphasize the extreme requirements on electrostatics that will become important at ultra-scaled nodes.

3.2 Simulation Approach

The schematic of the simulated 5nm gate length TMD devices including device dimensions and other parameters is shown in Fig. 3.1. To study the ultimate scaling limits, a gate length of 5 nm is used with an underlap of 2 nm on both sides of the gate to minimize fringe capacitance. We compute the electronic structure of few-layered and bulk MoS_2, $MoSe_2$, WS_2, and WSe_2 using the Vienna Ab initio Simulation Package (VASP) with GGA-PBE pseudopotentials [59, 91] (Fig. 3.2). The dielectric tensor was computed using the linear response routines in VASP. The geometry optimization and self-consistent calculations are performed with a 991 k-point mesh and a plane wave energy cutoff of 500 eV. The band gaps of TMDs with varying number of layers are shown in Fig. 3.3a.

Notably it is important to account for the effect of a satellite valley at the Σ_{min} point of the BZ (conduction band minimum (CBM) for bulk), in addition to the conduction band minima at K (CBM for monolayers), as can be seen from Fig. 3.3b. The electron effective masses (m*) in both longitudinal and transverse directions are extracted at K and Σ_{min} Fig. 3.4. The two-fold degeneracy at K and 6-fold degeneracy at Σ_{min} is accounted for by setting the effective masses to values such that the correct density of states is reproduced. A conduction band description is only considered in these devices because sizable band gaps in TMDs result in a significant barrier to hole injection.

A tight-binding (TB) Hamiltonian is constructed using these effective masses with both coupling between the valleys and that between the layers set to zero, the latter due to weak

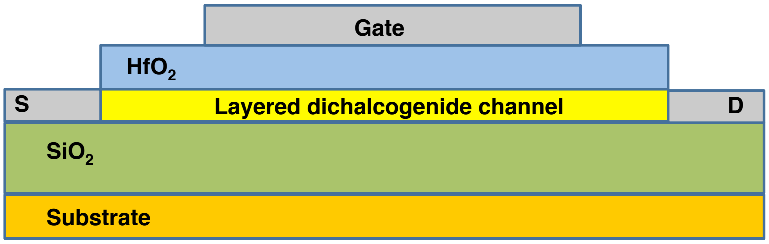


Figure 3.1: Schematic of the simulated layered TMD channel transistors. The gate length is 5 nm. There exists a very small underlap on both sides of the gate. The number of layers of TMDs is varied from 1 to 5, resulting in a body thickness in the range of 0.63 nm. The source and drain contacts are Ohmic (no Schottky barrier). No workfunction difference exists between gate electrode and the channel. EOT (Effective Oxide Thickness) is 0.5 nm. Interactions with the substrate are neglected.

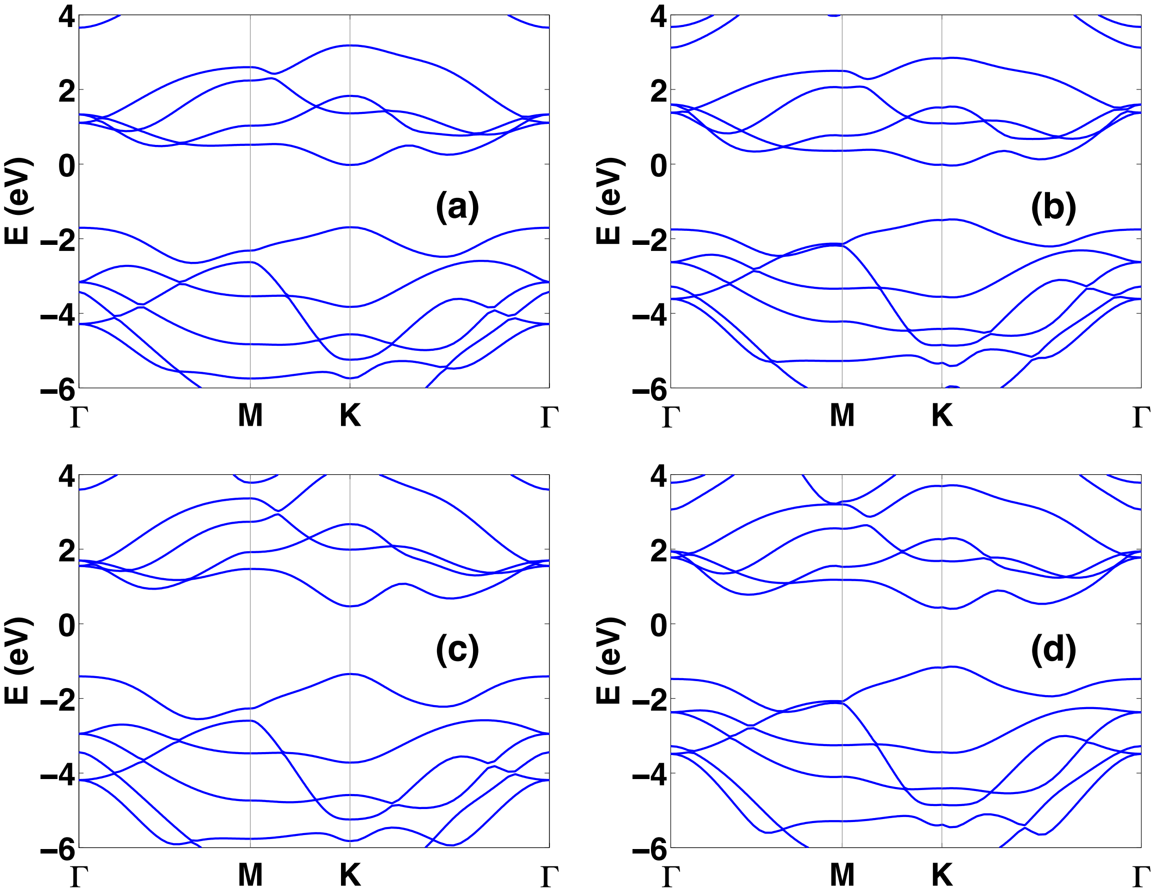


Figure 3.2: Electronic structure of monolayers of (a) MoS2, (b) MoSe2, (c) WS2, and (d) WSe2 computed using density functional theory (DFT) along high symmetry lines. Transverse and longitudinal effective masses are calculated for the K and Σ_{min} (along the K-Γ direction) valleys, which correspond to the first two conduction band minima in these materials. The band gaps are respectively 1.67, 1.44, 1.77, and 1.54 eV. Band structure is also calculated for 2-5 layers and bulk (not shown herein).

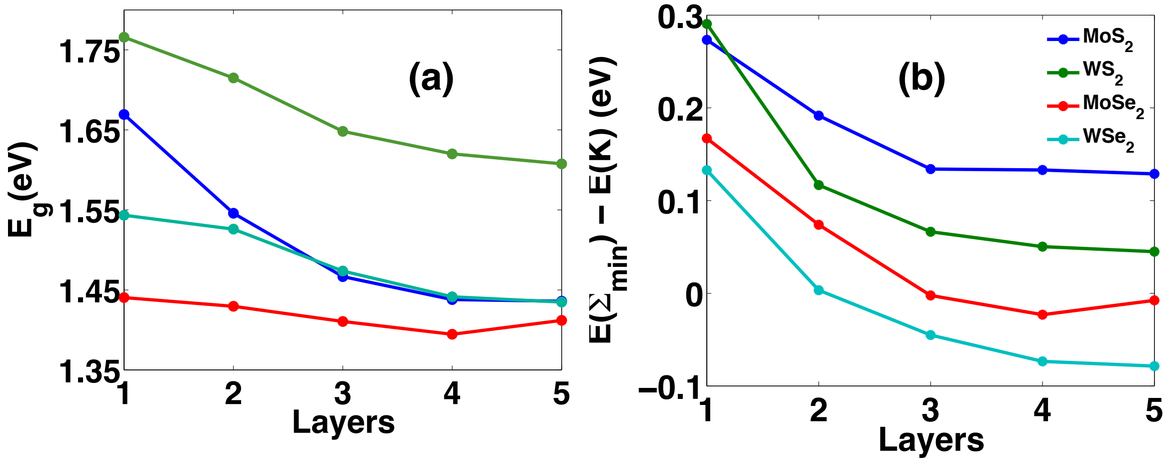


Figure 3.3: Variation of (a) band gap energy E_g and (b) energy difference between the K and Σ_{min} valleys with number of layers in TMDs. Our interest in investigating n-type carrier transport, coupled with the existence of large band gap in these systems motivates us to include only the conduction bands in our simulations.

inter-layer interactions in TMDs. A periodic boundary condition is used along the transverse direction. Finally, ballistic electronic transport behavior is investigated with zero Schottky barrier at the contacts in order to assess the maximum drive-current capabilities of these transistors. Dirichlet boundary conditions are used at the metal contacts. Transport equations are solved within the non-equilibrium Greens function (NEGF) formalism to calculate charge while self-consistently solving for electrostatics using Poissons equation. The current is calculated from the converged potential profile.

3.3 Results

Transfer Characteristics

The calculated $I_{DS} - V_{GS}$ characteristics for various TMD FETs corresponding to 1, 2, 3 and 5 layers are shown in Fig. 3.5. The subthreshold swings (SS) are 68, 68, 70 and

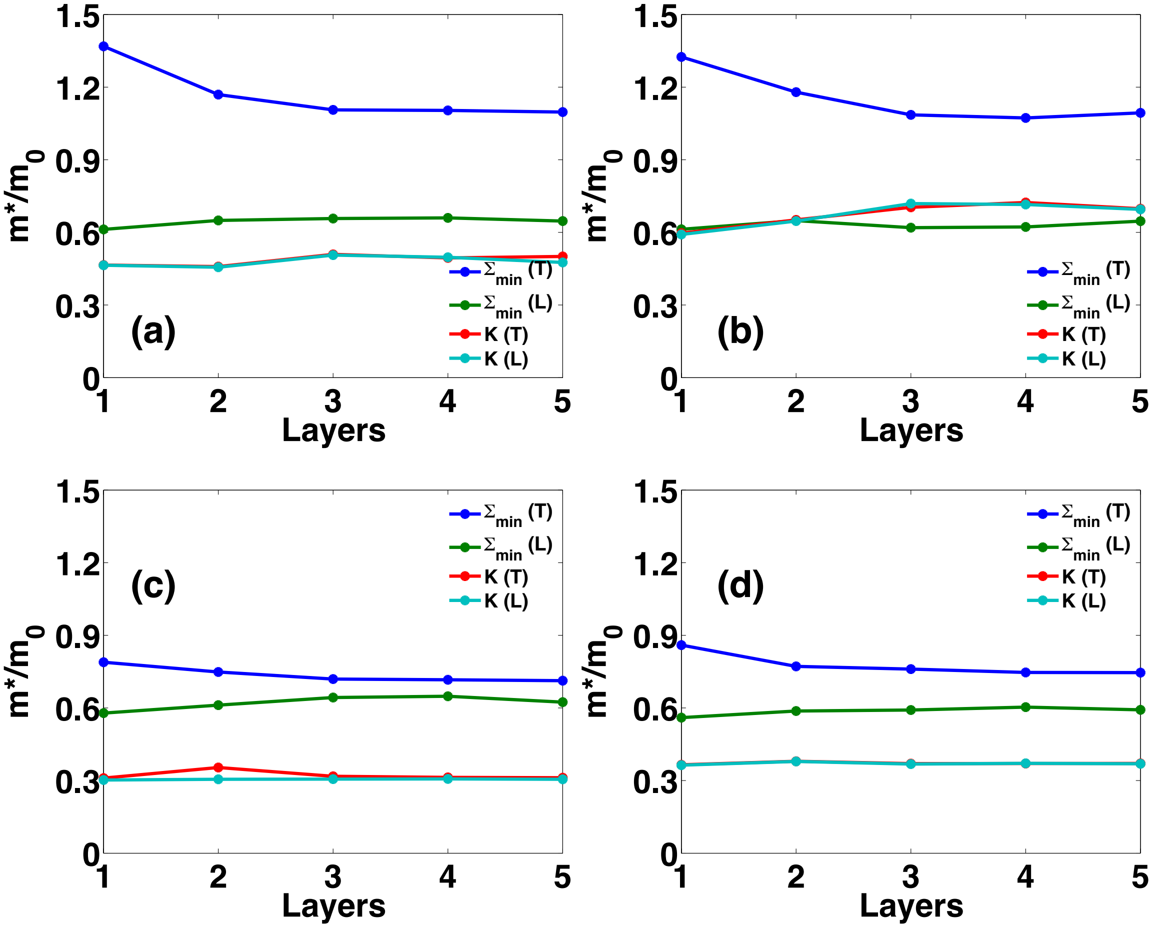


Figure 3.4: Variation of effective mass (m*) in the longitudinal (transport) and transverse directions with increasing number of layers in case of (a) MoS_2, (b) $MoSe_2$, (c) WS_2, and (d) WSe_2 extracted from density functional theory at K and Σ_{min} points. As is evident from Figs. 3.2 and 3.3 , the energy difference between the valleys therein is, in most cases, very low (less than 150 mV). Therefore, we incorporate both of them in our transport simulations. We note that our DFT calculations show the transverse effective masses at Σ_{min} in MoS_2 and $MoSe_2$ to be significantly larger than the rest, thereby contributing significantly to the total current in the bias ranges of our consideration.

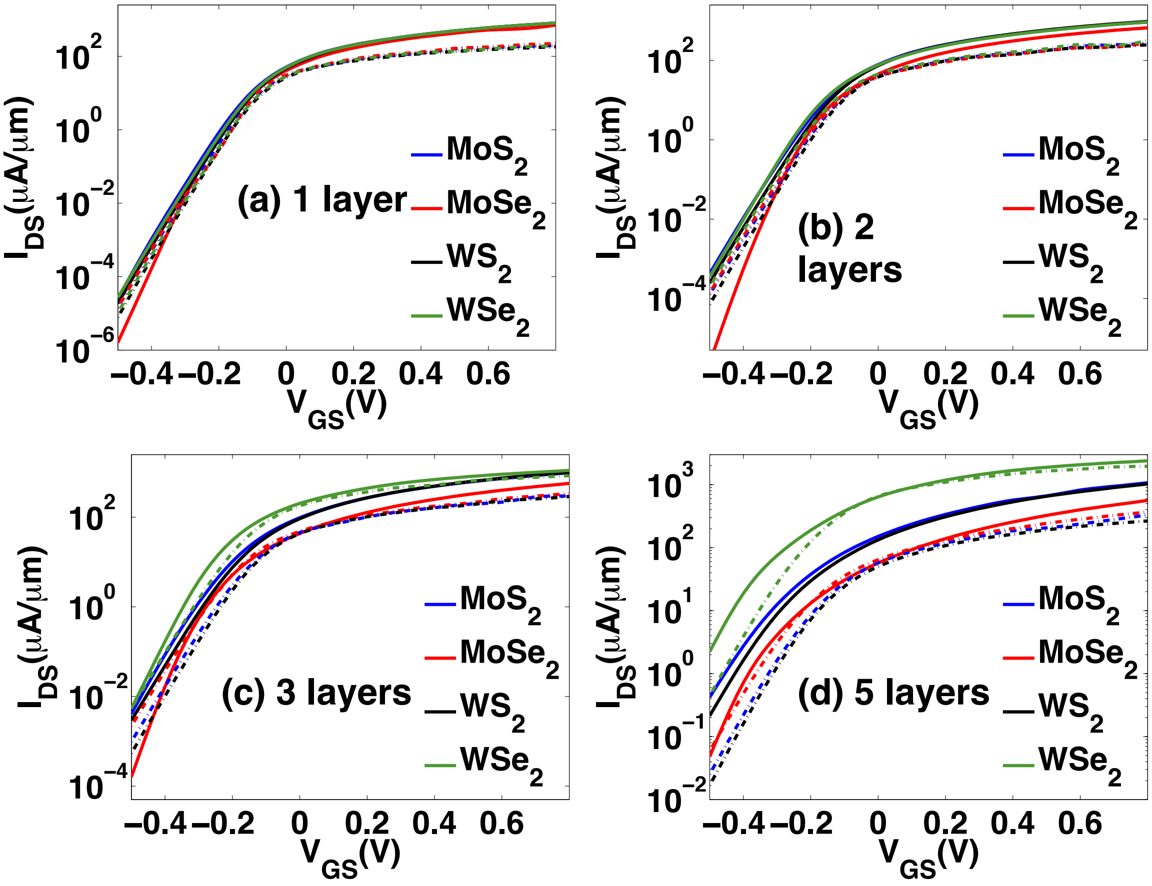


Figure 3.5: Calculated $I_{DS} - V_{GS}$ characteristics corresponding to two different drain voltages $V_{DS} = 0.05$ (dashed) and 0.5 V (solid lines) for the cases of various (a) monolayer, (b) bilayer, (c) 3-layer, and (d) 5-layer TMD transistors. Monolayer transistors exhibit excellent scaling behavior due to excellent electrostatic integrity. The transfer characteristics are progressively degraded with an increasing number of layers in all material-systems; i.e., the subthreshold swing and drain-induced barrier lowering (DIBL) both increase. The degradation is particularly severe in WSe$_2$ devices, while in MoS$_2$ FETs it is the least.

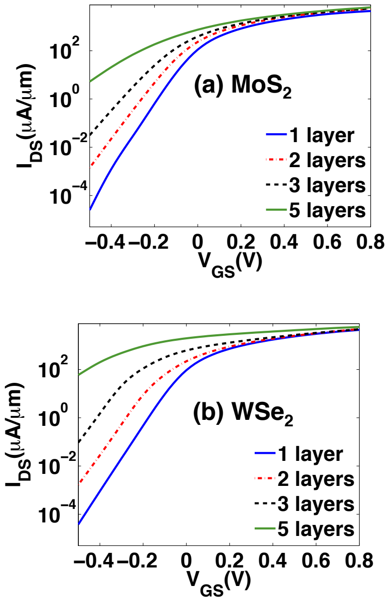


Figure 3.6: Calculated $\mathrm{I_{DS} - V_{GS}}$ characteristics of (a) MoS$_2$ and (b) WSe$_2$ transistors for varying thicknesses, showing the better scalability of the former. The larger effective mass of WSe$_2$ leads to a higher quantum capacitance resulting in a degraded subthreshold behavior.

73 mV/decade for monolayer MoS$_2$, MoSe$_2$, WS$_2$ and WSe$_2$ transistors respectively. The OFF state is severely degraded in WSe$_2$ FETs due to the dominant contribution from the heavier Σ_{min} valley, which is more easily accessible in WSe$_2$ than in other materials. The switching behavior degrades considerably with increasing number of layers due to diminished gate control. There is an increased difficulty in turning off the layers farthest from the gate due to dielectric screening from the layers closer to the gate, thus resulting in higher OFF currents. This is evident from Fig. 3.6; e.g., the SS in MoS$_2$ and WSe$_2$ devices increases to 86 and 92 mV/decade respectively for 3 layers. The effect of dielectric screening can also be seen in the ON state as the maximum achievable current doesnt increase significantly with increase in thickness even though there exists a greater number of transmission modes.

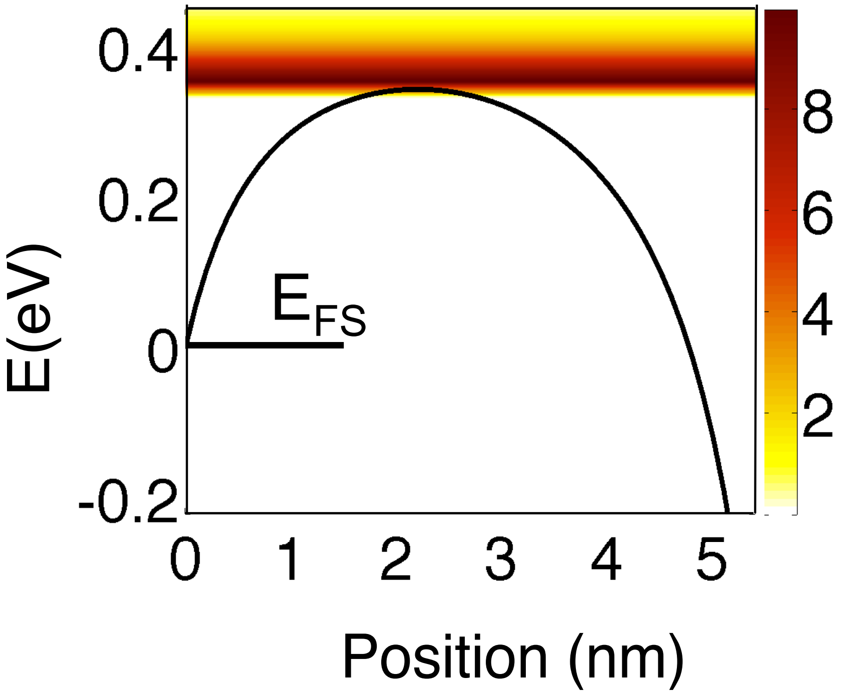


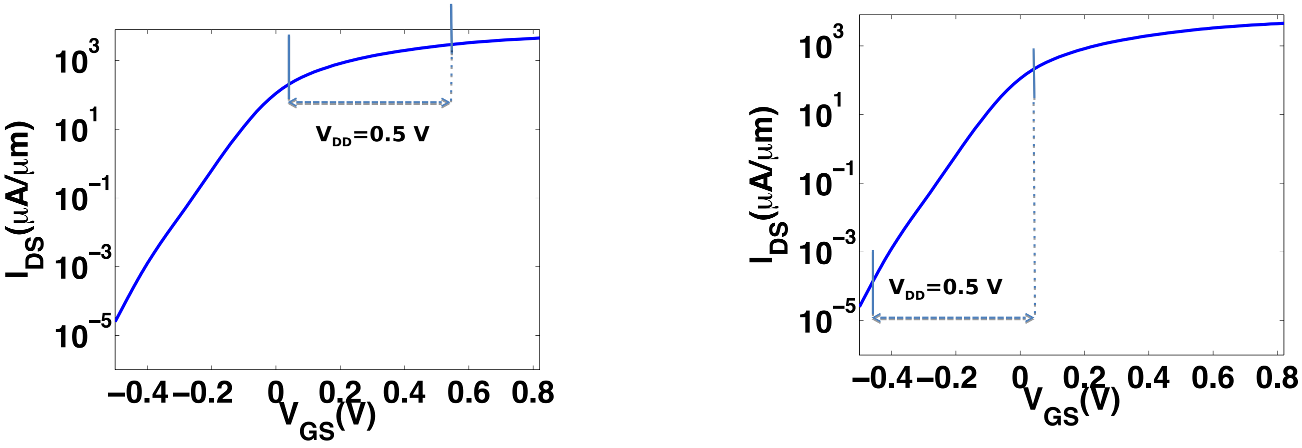
Figure 3.7: Current spectrum as a function of energy across the channel at V_{GS}= -0.5 V, V_{DS} = 0.5 V for monolayer MoS_2. Most of the current flows above the barrier, thus demonstrating inherent resistance to tunneling in TMDs.

Direct Source-Drain Tunneling

While previous studies have demonstrated direct source-to-drain tunneling to severely degrade OFF state leakage in sub-10 nm narrow bandgap transistors [72], we show that TMD FETs owing to their higher effective masses inhibit tunneling. Fig. 3.7 shows the energy-resolved current spectrum in the OFF state, wherein the majority of current is shown to flow above the barrier, thereby providing excellent scalability at these length scales. Band-to-band tunneling is limited in these devices because of high bangaps of TMDs resulting in a large barrier to hole injection. Thus, the performance of TMD FETs are still limited by thermionic injection at these gate lengths.

ON/OFF Ratio

The thickness dependence of the trade-off between the ON-current and ON/OFF ratio at V_{DS} = 0.5 V is plotted in Fig. 3.9 for all four materials considered in this study. The illustration of how the above was done is given in Fig. 3.8. Given the supply voltage of 0.5 V, the ON current can be changed to obtain different ON/OFF ratios. A balance between the achievable ON and OFF current can be achieved by sliding the window of



(a) (b)

Figure 3.8: Illustration of the study of ON/OFF ratio as a function of ON current for a given transfer characteristics. Given the supply voltage of 0.5 V, the ON current can be changed to obtain different ON/OFF ratios. If the ON current is taken into overdrive (a), then the OFF current might not be low enough to achieve the desirable ON/OFF ratio. On the other hand if the OFF current is driven too far into the subthreshold region, we might achieve the desirable OFF current, but fail to provide a good ON current. The trade-offs for different TMD materials and layer thickness is given in Fig. 3.9

supply voltage around to optimize the device. The trade-offs for different TMD materials and layer thickness is given in Fig. 3.9. With an increase in number of layers, we observe that (a) the maximum achievable ON/OFF ratio reduces, and (b) the ON-current for a reasonable ON/OFF ratio decreases, both consequences of reduced control of gate and the consequent degraded subthreshold behavior. Monolayer TMDs show the best scalability with the largest ON/OFF ratios, achieving ON current levels of $450\mu A/\mu m$ for an ON/OFF ratio of 10^5, while the bilayer devices deliver about half this value. These results emphasize the significant loss in gate control on adding only a second layer to the conducting channel.

Output Characteristics

The output ($I_{DS} - V_{DS}$) characteristics shown in Fig. 3.10 exhibit a linear dependence of saturation current on V_{GS}. The electrostatic integrity in TMDs results in well-saturated output characteristics with very small output conductance (g_0) 600 and 2000 $\mu S/\mu m$ at V_{GS} = 0.6 V respectively for mono- and 5-layer MoS_2 devices. The higher output conductance for 5-layer devices can be explained by an increased DIBL as drain control increases for thicker

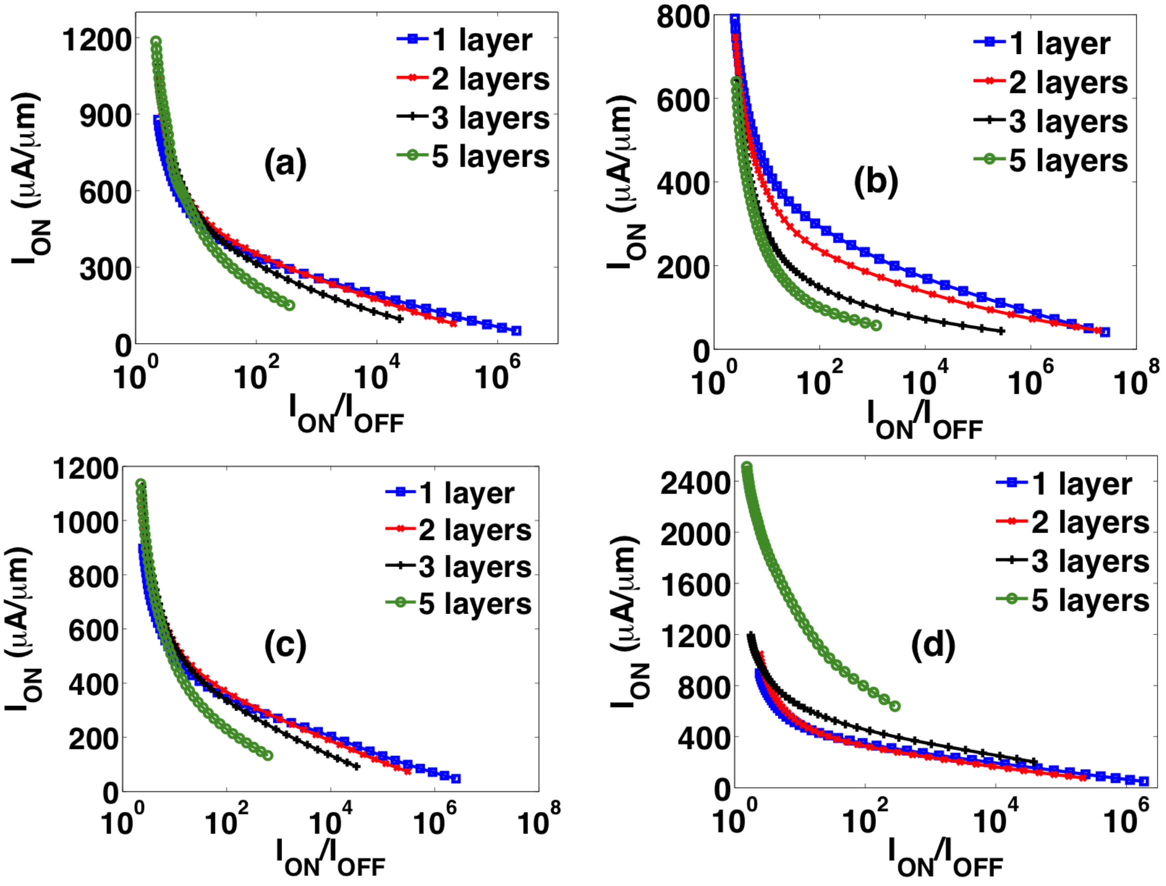


Figure 3.9: ON-state current as a function of the ON/OFF ratio for (a) MoS_2, (b) $MoSe_2$, (c) WS_2 and (d) WSe_2 transistors of varying channel thickness for $V_{DS} = 0.5$ V. The ON current for a given ON/OFF ratio reduces with increasing number of layers because of reduced gate control.

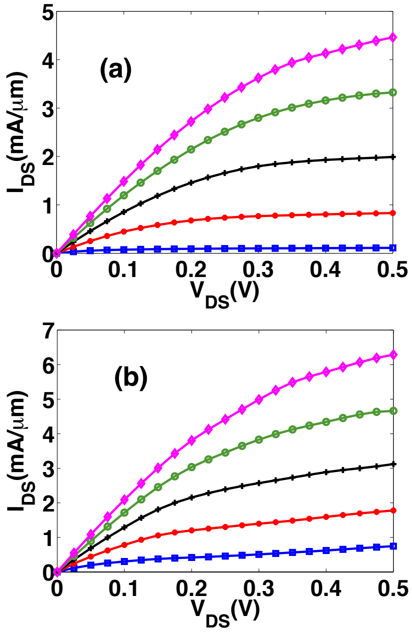


Figure 3.10: $I_{DS} - V_{DS}$ characteristics corresponding to various gate voltages from V_{DS} =0.0 V to 0.8 V in steps of 0.2V for (a) monolayer and (b) 5-layer MoS_2 transistors. The output resistance is lower for thicker channels and for higher gate voltages

channels. The output conductance also increases for higher gate voltages as V_{DS} determines the number of states contributing to transport at these levels of current.

3.4 Conclusion

Using first-principles electronic structure calculations and self-consistent ballistic NEGF simulations, we have investigated 5 nm channel length FETs based on TMD materials. Two significant conclusions can be drawn from our results: (i) TMD materials could be scaled down to 5 nm with excellent ON/OFF (10^6), subthreshold swing (65 mV/decade) and g0 ($150\mu S/\mu m$) thanks to their large bandgaps and high effective masses that inhibit direct source-drain tunneling; (ii) monolayer is a critical requirement at these dimensions: no increase in current can be obtained by increasing the number of layers as the loss of gate control from increased body thickness significantly affects the swing, reducing the ON current for reasonably high ON/OFF.

While it is possible to achieve the ITRS (International Technology Roadmap for Semiconductors) LOP (Low Operating Power) specs in terms of ON/OFF ($> 10^5$) at $V_{DS} = 0.5$ V for monolayer devices, the ON current, at that ON/OFF, is lower than the required value ($700\mu A/\mu m$), even though we considered operation in a purely ballistic regime. This highlights the difficulties in operating below 0.5 V at 5 nm gate lengths. The next chapter will go into details for one TMD material, MoS_2 and look at the transport properties using a rigorous tight-binding Hamiltonian.

Chapter 4

Screening in Ultra Short (5nm) Channel MoS_2 Transistors: A Full Band Quantum Transport Study

4.1 Introduction

In the last chapter, we looked at transport properties and drive current capabilities of four different TMDS for different number of layers. Even though we looked at both the K and Σ_{min} valleys to get a quantitative understanding of what extent of gate control can actually be achieved, a more rigorous model of the bandstructure is required to look at the effect of applied voltages on the charge density so that the current levels can be appropriately captured. Here, we present, a full band, self-consistent, quantum transport study of electrostatic screening in 5 nm channel length MoS_2 transistors with doped contacts as a function of number of layers with both single gate and double gate geometry. Our results show that for such ultra-short channel lengths (i) the layer closest to the gate can effectively screen out the gate potential due to a large density of states, and as a result the gate cannot effectively control more than one layer, (ii) for a monolayer, a significant short channel effect can still be observed and (iii) because of (i) and (ii), only a double gate geometry for a monolayer device provides reasonable gate control (subthreshold swing $\sim$ 84 mV/decade). Surprisingly, these numbers are not better and rather comparable to what could be achieved with a surround gate Si nanowire of small dimensions ($\sim$ 3 nm diameter)[72].

4.2 Approach

Bandstructure

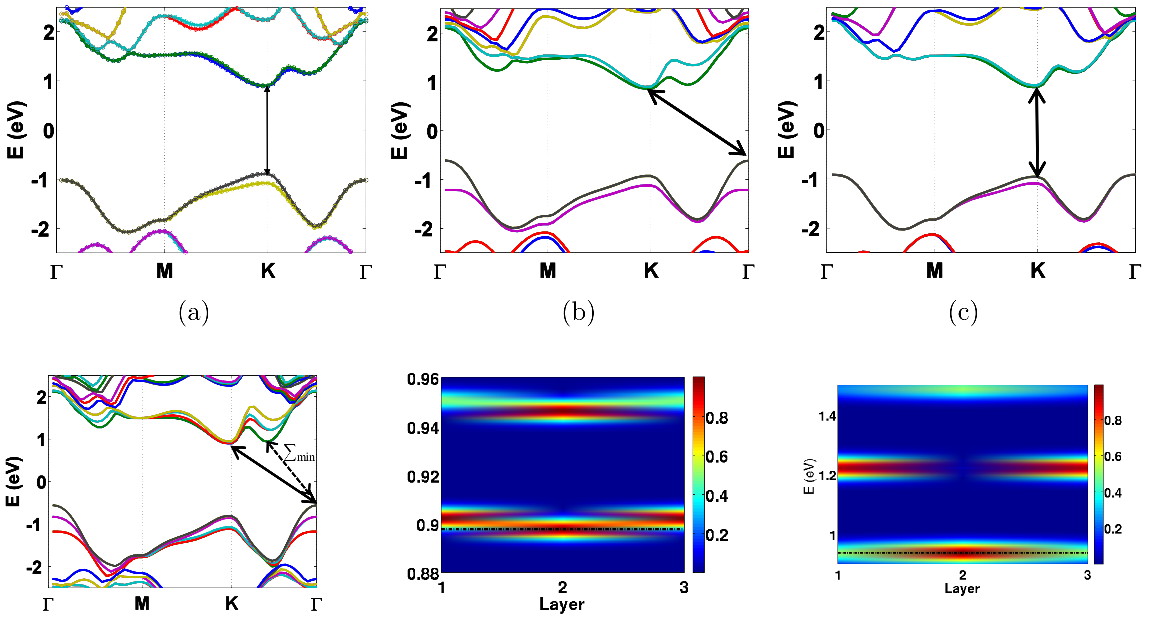
The electronic structure calculations (Fig. 4.1(a-d)) of MoS_2 were performed by fitting orthogonal tight-binding parameters (TB) to density functional theory (DFT) calculations.

The parameterization scheme used in this work follows a similar technique as described in [127], but with improvements that allow for directly including the deviations of band gaps and effective masses into the minimized cost function [84, 104, 15, 20]. The obtained tight-binding parameters are listed in Table I, and the corresponding fitting results of band gaps and effective masses are listed in tables II and III, respectively. The electronic states near the top of valence bands and the bottom of conduction bands are mainly contributed from Mo d-orbitals and S p-orbitals, mixing with Mo s-orbitals [75]. The energy positions of the states are determined through complicated interactions between those orbitals and many other states in the Hilbert space with higher energies. In order to reproduce the band structure obtained using the first-principles method with high precision, we have also included Mo p-orbitals, S s- and d-orbitals in our tight-binding model, which are used to include the influence of the many other states with higher energies, in an effective way. Therefore, the parameters related to those orbitals may lose their original physical meanings, and should be considered as pure mathematical parameters. The inter-layer interactions were included in the bandstructure calculation. This leads to an indirect band-gap for few-layered MoS$_2$ as shown in Fig. 4.1(a-b). The band-gaps from the TB parameters were calculated to be 1.8 eV, 1.48 eV and 1.46 eV for monolayer, bilayer and 3-layer MoS$_2$ respectively. The bandstructure matches well with previous theoretical and experimental results [31]. Monolayer MoS$_2$ has a direct band gap at the K-point. The valence band maximum shifts to the Γ point for few-layered devices, while it shifts back from the Γ point to K point if the inter-layer interactions are removed, showing their significance especially in p-type transport. The conduction band minimum also shifts from the K point to Σ_{min} (local minimum along $K-\Gamma$) point as the number of layers is increased from 3. In this study, we will only investigate the electronic properties of 1-3 layered transistors.

One important observation for multi-layer (e.g., the 3-layer) structures is the fact that increased surface energy leads to a higher projected density of eigenstates in the inner layer (Fig. 4.1 (e-f)). The layer-wise density was calculated by normalizing the eigenvector corresponding to a certain energy and momentum and summing over the probability of all the orbitals belonging to a layer. The probability density of eigenstates at the bottom of the conduction band at both the K and Σ_{min} points shows a considerable confinement of electrons to the inner layer, thus resulting in higher charge densities in those layers. There is reasonably large density of states in the surface layers 100 meV above the band minimum at the Σ_{min} point while it is significant 20 meV from the minimum at the K-point. Thus the effect will be more significant for multi-layered devices, as Σ_{min} becomes the conduction band minimum. A higher current could be expected in the middle layer in the OFF state for 3-layered transistors, if the electric field can penetrate through the top most layer.

Device Simulation

The schematic of the simulated device structures along with the device parameters are shown in Fig. 4.2. We use both single-gated (SG) and double-gated (DG) devices in this study. The oxide thickness for each gate stack has an effective oxide thickness of 0.5 nm. Highly doped



(a) (b) (c)

(d) (e) (f)

Figure 4.1: Electronic structure computed from tight-binding models along high symmetry lines for (a) monolayer MoS$_2$ (Electronic structure from DFT calculations marked with circles), for bilayer MoS$_2$ calculated using TB models with the interaction between the two layers turned (b) on and (c) off and for (d) 3-layer MoS$_2$. The band gap of monolayer MoS$_2$ is 1.8 eV with the offset between the K (conduction band minimum) and $\Sigma_{\min}$ (minimum energy point along the K-Γ direction) valleys equal to 0.2503 eV. The valence band maximum for monolayer MoS$_2$ shifts from the K point to the Γ point when the interaction is turned on resulting in an indirect band-gap. The band gap of bilayer MoS$_2$ is equal to 1.48 eV, while that of 3-layer MoS$_2$ is 1.46 eV. The conduction band minimum shifts from K to $\Sigma_{\min}$ as the number of layers is increased from 3. The layer-wise projected density of states for the eigenstates at the bottom of conduction band at (e) K and (f) $\Sigma_{\min}$ points for 3-layer MoS$_2$ show a higher confinement of electrons in the middle layer. The bottom of the conduction band is shown by the dotted black line. A lower temperature was used at the K point as opposed to room temperature used at the $\Sigma_{\min}$ point to show the confinement effects at the K point as the eigenstates are closer to each other. The effect is thus less significant at the K point compared to the $\Sigma_{\min}$ point.

Table 4.1: Tight-binding parameters for MoS$_2$ using orthogonal model with sp^{3}d^5 orbitals, nearest-neighbor interactions, and spin-orbit coupling, in the unit of eV.

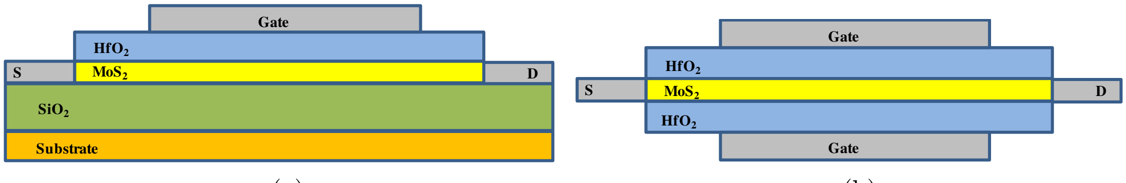
	Single Layer	Double Layer	Bulk
On-site energy			
s(S)	17.9023	17.4692	17.2848
p(S)	-2.4009	-2.6415	-2.2950
d(S)	75.2885	75.1980	74.6303
s(Mo)	9.9447	9.7433	10.1452
p(Mo)	36.6985	37.4747	36.2812
d(Mo)	4.1225	4.3256	4.2039
Spin-orbit splitting λ_{SO}			
p(S)	0.05388	0.2446	0.3126
p(Mo)	0.9235	0.8246	1.5907
Slater-Koster energy integral (intra-layer)			
s(S)s(S)σ	-0.8590	-0.5104	-0.4557
s(S)p(S)σ	-0.2142	-0.1193	-0.2661
p(S)p(S)σ	0.8715	0.9152	0.9406
p(S)p(S)π	-0.2449	-0.2604	-0.3175
s(S)d(S)σ	3.1818	3.3973	3.6853
p(S)d(S)σ	0.1138	-0.4924	0.2515
p(S)d(S)π	-0.4476	0.2920	-0.3709
d(S)d(S)σ	3.7203	-2.7489	3.6162
d(S)d(S)π	-2.5901	-2.7489	-1.3972
d(S)d(S)δ	-1.1719	-1.0508	-1.3690
s(Mo)s(Mo)σ	-1.5166	-1.4157	-1.4603
s(Mo)p(Mo)σ	0.4991	0.5353	0.1376
p(Mo)p(Mo)σ	-3.8198	-4.4557	-5.3723
p(Mo)p(Mo)π	4.5562	4.6003	4.3870
s(Mo)d(Mo)σ	0.007971	0.06601	0.3100
p(Mo)d(Mo)σ	1.3306	1.0464	1.3263
p(Mo)d(Mo)π	-0.4728	-0.3587	-0.3780
d(Mo)d(Mo)σ	0.9590	1.0819	0.9623
d(Mo)d(Mo)π	-0.452	-0.3991	-0.4319
d(Mo)d(Mo)δ	0.5143	0.4971	0.4623
s(S)s(Mo)σ	-0.1246	0.4253	0.1484
s(S)p(Mo)σ	3.9553	4.3327	3.7256
p(S)p(Mo)σ	1.2385	0.9410	1.1730
p(S)p(Mo)π	-0.2589	-0.3840	-0.4290
s(S)d(Mo)σ	1.6798	1.2016	1.6079
p(S)d(Mo)σ	-2.8710	-2.7683	-2.9008
p(S)d(Mo)π	0.8901	0.8137	0.9168
d(S)d(Mo)σ	4.8937	5.7088	5.0221
d(S)d(Mo)π	-9.3391	-9.3064	-9.2758
d(S)d(Mo)δ	1.2478	1.1624	1.6762
s(S)d(Mo)σ	1.6798	1.2016	1.6079
s(Mo)p(S)σ	1.1862	1.0713	1.0930
s(Mo)d(S)σ	10.4024	9.5661	9.9100
p(Mo)d(S)σ	16.3744	16.4443	16.2916
p(Mo)d(S)π	-16.6761	-16.7952	-16.4873
Slater-Koster energy integral (inter-layer)			
s(S)s(S)σ		0.3665	-0.1649
s(S)p(S)σ		-0.7006	-0.03491
p(S)p(S)σ		0.4188	0.3206
p(S)p(S)π		0.07841	0.06415
s(S)d(S)σ		-0.09494	0.5781
p(S)d(S)σ		0.8274	1.0903
p(S)d(S)π		-0.6468	-0.6043
d(S)d(S)σ		-0.1055	-0.4620
d(S)d(S)π		-1.2847	-0.7753
d(S)d(S)δ		-0.5428	-0.9156

Table 4.2: Band gap energies obtained by DFT-HSE [127] and our TB model. The fifth column is the deviation between the HSE and the TB values. All the energies are in the unit of eV. Subscripts v and c stand for valence band and conduction band, respectively. The splitting of the valence band maximum at K point is given by $\mathrm{K_{v1}}$ (top) and $\mathrm{K_{v2}}$ (bottom), whereas Σ is the midpoint of the line joining the Γ and the K points.

Structure	Transitions	Band-gap energies (eV)		
		HSE (target)	TB (fitted)	Deviation%
Monolayer	$\mathrm{K_{v1}}$ to $\mathrm{K_c}$	1.7857	1.7857	0.00
	$\mathrm{K_{v2}}$ to $\mathrm{K_c}$	1.9742	1.9742	0.00
	Γ_v to $\mathrm{K_c}$	1.9457	1.9123	-1.72
	Γ_v to Σc	2.2252	2.1613	-2.87
Bilayer	Γ_v to $\mathrm{K_c}$	1.4801	1.4749	-0.36
	Γ_v to Σ_c	1.6178	1.5532	-3.99
	$\mathrm{K_{v1}}$ to $\mathrm{K_c}$	1.7787	1.7894	0.60
	$\mathrm{K_{v2}}$ to $\mathrm{K_c}$	1.9802	1.9829	0.14
Bulk	Γ_v to Σ_c	1.3280	1.3280	0.00
	Γ_v to $\mathrm{K_c}$	1.3661	1.3543	-0.86
	$\mathrm{K_{v1}}$ to $\mathrm{K_c}$	1.7751	1.6755	-5.61
	$\mathrm{K_{v2}}$ to $\mathrm{K_c}$	1.9985	2.0411	2.13

Table 4.3: Values of effective masses at various band edges in the unit of free electron mass ($\mathrm{m_0}$) calculated using the HSE method [127] and our tight-binding model. The subscripts l and t refer to the masses calculated at the point along the longitudinal and the transverse directions of the line connecting the Γ point and that point, respectively.

		Electron mass ($\mathrm{m_0}$)			Hole mass ($\mathrm{m_0}$)		
Structure	Point	HSE (target)	TB (fitted)	Deviation (%)	HSE (target)	TB (fitted)	Deviation (%)
Monolayer	$\mathrm{K_l}$	0.4065	0.4072	0.16	0.4852	0.4855	0.06
	$\mathrm{K_t}$	0.4035	0.4031	-0.10	0.4804	0.4802	-0.05
Bilayer	Γ				1.0387	1.0387	0.00
	$\mathrm{K_l}$	0.4302	0.4275	-0.63	0.4851	0.4853	0.03
	$\mathrm{K_t}$	0.4227	0.4260	0.79	0.4810	0.4853	-0.21
Bulk	Γ				0.7849	0.7849	0.00
	Σ_l	0.5737	0.5737	0.00			
	Σ_t	0.8186	0.8186	0.00			



(a) (b)

Figure 4.2: Schematic of the simulated layered MoS$_2$ transistors with (a) Single-gate and (b) Double-gate. The gate length is 5 nm. The number of layers of TMDs is varied from 1 to 3, resulting in a body thickness in the range of 0.6-1.8 nm. The source and drain contacts are Ohmic (no Schottky barrier) with the gray region being the doped semiconductor. The effect of metal contacts (which would be deposited on either side of the channel) to the channel is ignored as we want to look at the best-case performance of the transistor. EOT is 0.5 nm.

[36] semiconductor contacts, with a doping concentration of 3×10^{13} cm^{-2} (large doping concentration also reduces the size of the problem to be solved as the length of source and drain regions can be shortened) are used in the source and drain regions. The workfunction difference between the gate and channel is assumed to be zero. The width of each device was assumed to be large enough so that a mode space summation could be used along that direction, while a real space representation was used along the direction of transport [113]. The TB parameters fitted from the DFT calculations were used to formulate the full-band Hamiltonian for each of the considered devices. The charge was calculated within the Non-Equilibrium Greens Function (NEGF) formalism [29, 4]. The calculated charge was then used by a finite difference Poisson solver with appropriate boundary conditions to calculate the potential corresponding to the charge [95]. Dirichlet boundary conditions are used at the gate contacts, while Neumann boundary conditions are assumed at the electrostatic domain boundary for doped contacts so that the electric potential profile floats to ensure charge neutrality at the boundaries. This was then used by the transport solver to achieve a self-consistent solution for each bias point. The transmission $T(E)$ as a function of energy was then calculated using the converged potential profile along the channel, while summing over all the transverse modes. Scattering effects could be considered to be minimal at these channel lengths. The valence band could be ignored in most of the calculations, because of the large band gap of MoS$_2$ and hence a lack of band-to-band tunneling. The total current was then calculated by summing the transmission over the energy grid by weighting it with the difference in Fermi distribution at the source and drain.

$$I = \sum_E dE\, T(E) \left(f\left(E - \mu_S\right) - f\left(E - \mu_D\right)\right), \tag{4.1}$$

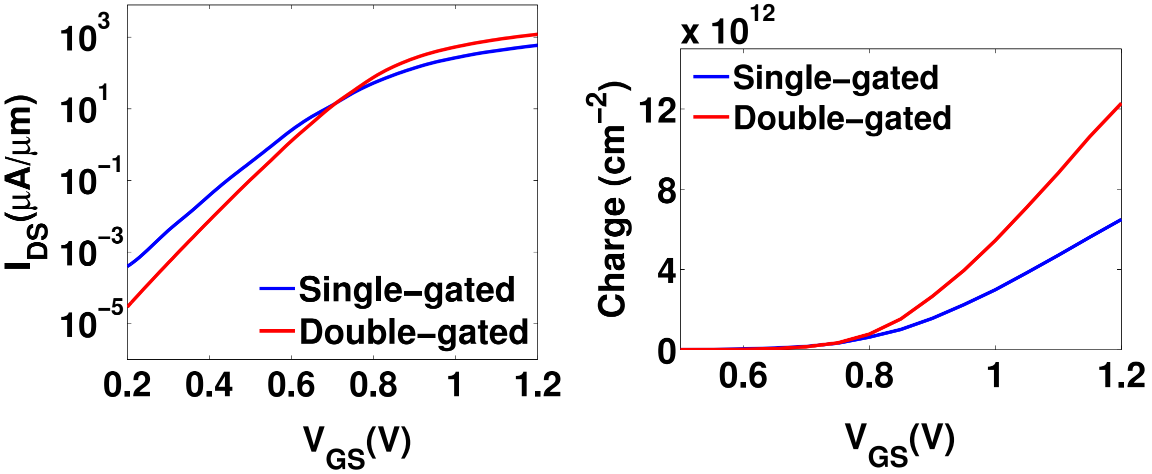
where $f(E)$ is the Fermi-Dirac distribution, while μ_S and μ_D are the chemical potentials at the source and drain respectively.

4.3 Results and Discussion

The calculated $I_{DS} - V_{GS}$ characteristics for single and double gated MoS$_2$ for low drain voltage ($V_{DS} = 0.05$ V) are shown in Fig. 4.3. The DG device (SS $\sim$ 84 mV/decade) shows better performance than the SG device (SS $\sim$ 102 mV/decade) even in the case of a monolayer demonstrating significant short channel effects. The charge accumulated at the top of the barrier is almost twice for the DG device compared to the SG device showing a considerably higher gate control. There is a need of further scaled gate oxides at these gate lengths to achieve reasonable performance characteristics. Fig. 4.4a compares $I_{DS}-V_{GS}$ characteristics for double-gated monolayer devices with fixed top oxide thickness of 0.5 nm and varying bottom oxide thickness. The subthreshold behavior degrades with increased bottom oxide thickness and tends to resemble the single-gated device. The $I_{DS} - V_{GS}$ characteristics for monolayer single-gated MoS$_2$ shown in 4.4b demonstrates the increased drain control as the contact doping is increased leading to degraded subthreshold behavior. The $I_{DS} - V_{DS}$ characteristics are shown in Fig. 4.5a for single-gated monolayer MoS$_2$ at different gate biases. Negative differential resistance can be observed for high drain voltages (V_{DS} greater than 0.45 V) because of limited bandwidth of the first few bands. The current spectrum as a function of energy as shown in Fig. 4.5b for $V_{DS} = 0.40$ V and 0.45 V lie on top of each other, while that of $V_{DS} = 0.50$ V is lower. This effect has been observed before and is a manifestation of ballistic transport, which may not be observable in experiments because of electron-phonon scattering [108, 17]. Electron-phonon coupling introduces transmission channels among different transverse modes, thus eliminating the gap in current transmission.

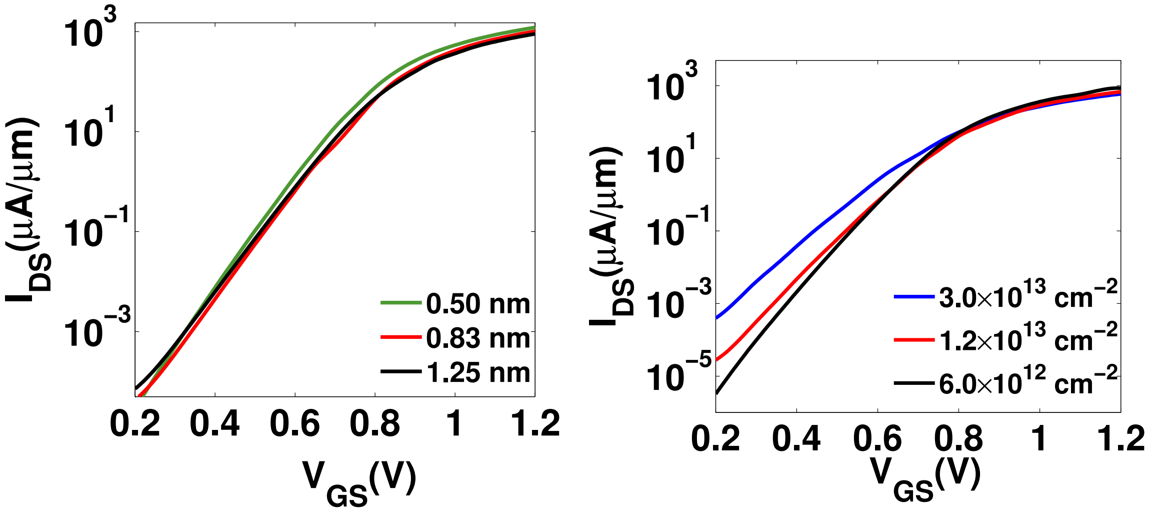
Going on to bilayer, Fig. 4.6a shows a comparison of $I_{DS} - V_{GS}$ characteristics of SG vs DG bilayer MoS$_2$ at low drain voltage. Bilayer MoS$_2$ shows much weaker gate control compared to monolayer MoS$_2$; SS $\sim$ 109 and 140 mV/decade for DG and SG respectively. The charge density at the top of the barrier in Fig. 4.6b shows the screening of the 2$^{\text{nd}}$ layer from the gate by the top layer in the ON state. The current contribution from the 2$^{\text{nd}}$ layer is therefore significantly less than that of the top layer. The subthreshold swing goes down to 90 mV/decade for a doping concentration of 6×10^{12} cm^{-2} for Fig. 4.7 shows the loss in gate control of the bottom layer in both the ON and OFF states. The barrier to current flow is higher in the OFF state for the top layer resulting in marginally lower current, while it is significantly lower in the ON state thus screening the bottom layer from the gate. These results underline the significant screening effect by charge accumulated at the layer closest to the gate and therefore poor gate control of any additional layers. Because of these small screening lengths, the DG bilayer behaves like two SG monolayer transistors resulting in similar transfer characteristics to the same.

To test the consistency of these results, we have further investigated a 3-layer device. The higher gate control of DG compared to SG 3-layer transistors can be seen in Fig. 4.8a, with lower OFF and higher ON current in the DG device. Fig. 4.8b shows layer-wise currents for the SG 3-layer device, where the bottom layer contributes the highest current in the OFF state, while the top layer conducts most of the current in the ON state, showing that



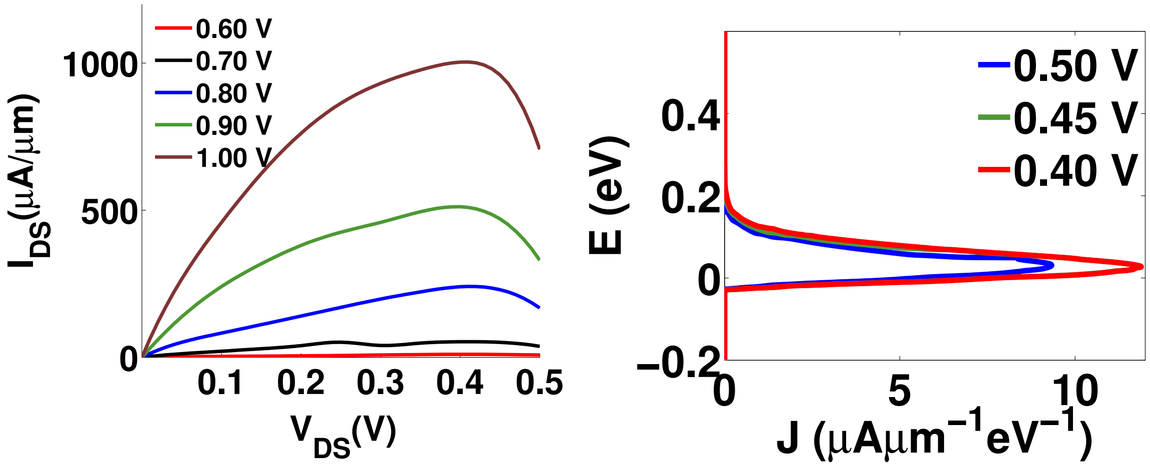
(a) (b)

Figure 4.3: (a) Calculated $I_{DS} - V_{GS}$ characteristics and (b) Charge density at the top of the barrier for single-gated and double-gated monolayer MoS$_2$ transistors at $V_{DS} = 0.05$ V. The double-gated transistor has significantly better electrostatic control with better subthreshold swing and higher on-currents. The charge density is about two times higher for double-gated transistors.



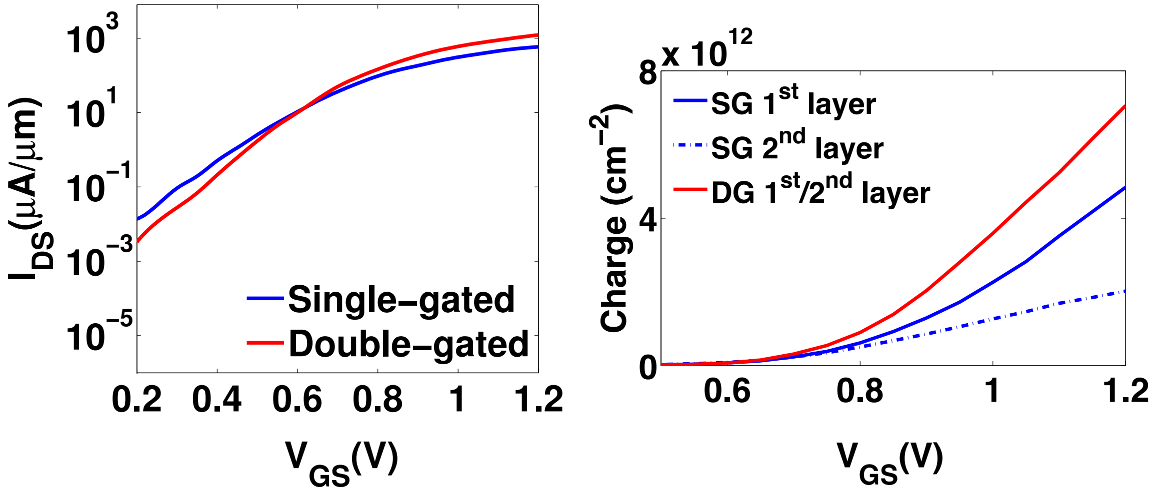
(a) (b)

Figure 4.4: Calculated $I_{DS} - V_{GS}$ characteristics of (a) double-gated monolayer MoS$_2$ transistors with top oxide thickness of 0.5 nm and varying bottom oxide thickness and (b) single-gated monolayer transistors with varying source-drain doping concentration at $V_{DS} = 0.05$ V. The gate control goes down with increased oxide thickness. The drain control increases with higher doping concentration leading to worse electrostatic control



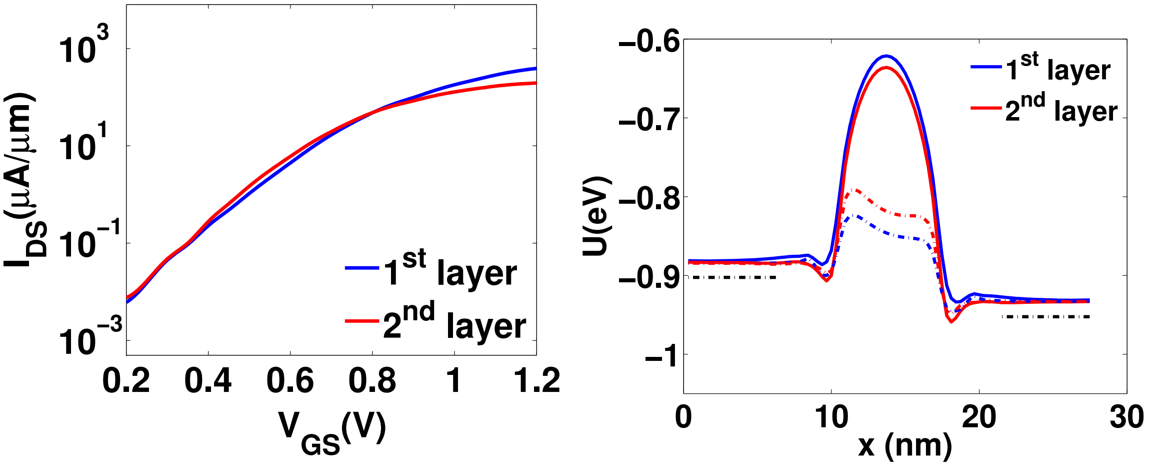
(a) (b)

Figure 4.5: (a) Calculated $I_{DS} - V_{DS}$ characteristics of monolayer MoS$_2$ transistors and (b) current spectrum as a function of energy. Negative differential resistance can be observed in the transistors at high drain voltages due to limited bandwidth. The effect is confirmed in the current spectrum as transmission reduces beyond a drain voltage of 0.45 V.



(a) (b)

Figure 4.6: (a) Calculated $I_{DS} - V_{GS}$ characteristics and (b) Charge density at the top of the barrier for single-gated and double-gated (each layer) bilayer MoS2 transistors at $V_{DS} = 0.05$ V. The double-gated transistor shows better performance owing to greater gate control. The charge on the 2^{nd} layer is significantly lower than on the 1^{st} layer showing a small effective screening length.



(a) (b)

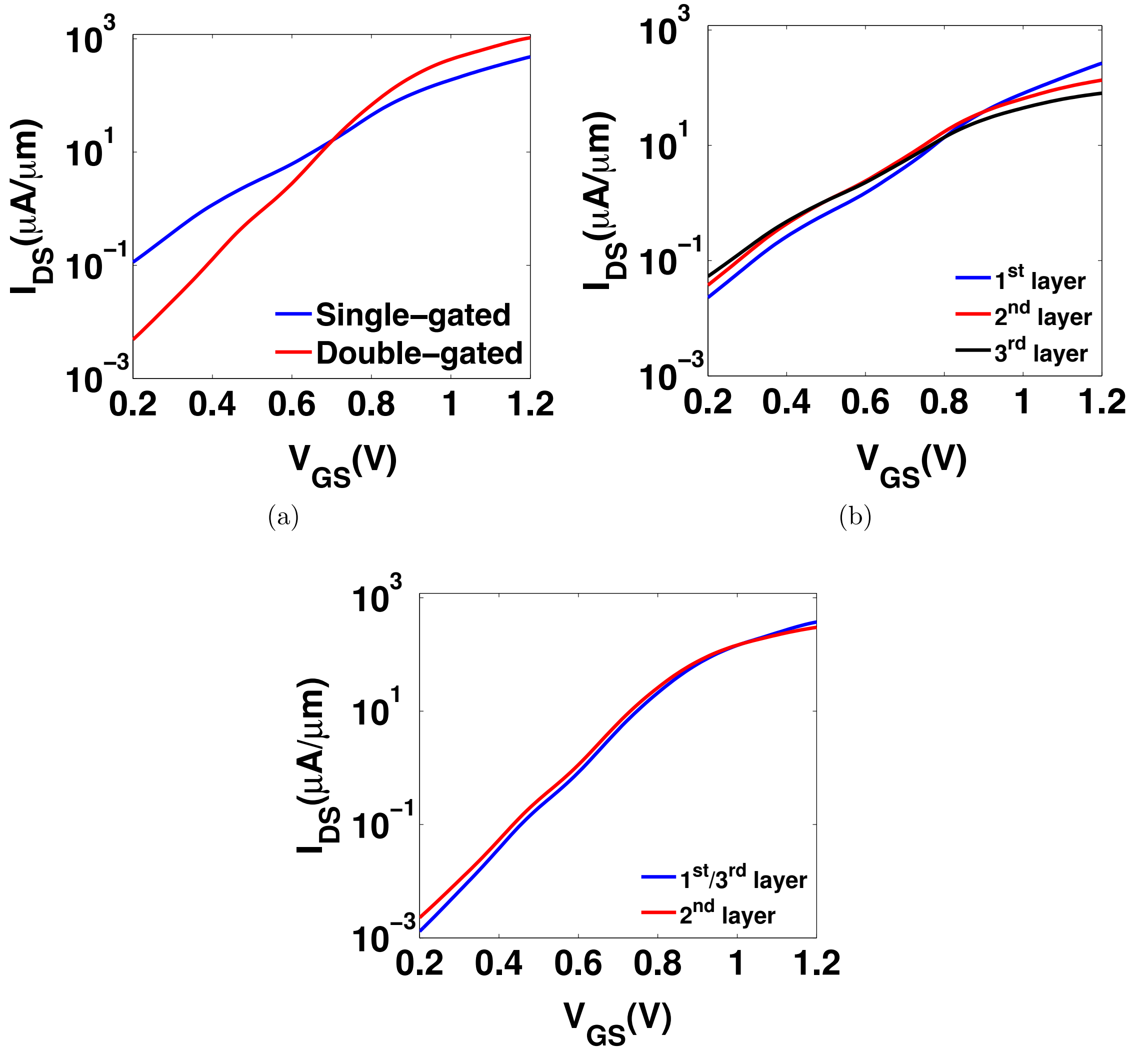
Figure 4.7: (a) Layer-wise current for single-gate bilayer MoS2 at V_{DS}=0.05 V, (b) potential profile along the channel at $V_{GS} = 0.40$ V (solid lines) and V_{GS}=1.10 V (dashed lines). The 1^{st} layer has higher current for higher gate voltages while showing marginally lower OFF current. The potential barrier is lower for the top layer at high gate voltages resulting in higher current.

the change in thermal barrier to current flow is the highest for the top layer. The DG 3-layer device shows similar characteristics with the middle layer carrying lowest current in the ON state, while some confinement effects can be seen in the OFF state (Fig. 4.8c). The middle layer carries more current than the surface layers in the OFF state, i.e. for $V_{GS} = 0.8$ V. This effect, which is a consequence of higher DOS near the band edge in the middle layer (Fig. 4.1(e-f) is further illustrated in Fig. 4.9. Fig. 4.9a shows the potential profile along the channel for the DG 3-layer device at $V_{GS} = 0.75$ V. The conduction band is further below the Fermi level for the surface layers compared to the middle layer, because of higher concentration of electrons towards the center of the channel. The barrier height of the electrons from the Fermi level is the same for all the layers at this bias point as screening effects are not significant. The current density as a function of energy,

$$J(E) = T(E)\left(f(E-\mu_S) - f(E-\mu_D)\right), \tag{4.2}$$

is shown in Fig. 4.9b at the same bias point. As expected, higher current is drawn from the middle layer because of increased surface energy for the outer layers. Both the surface layers have similar characteristics because of symmetry in the simulated device.

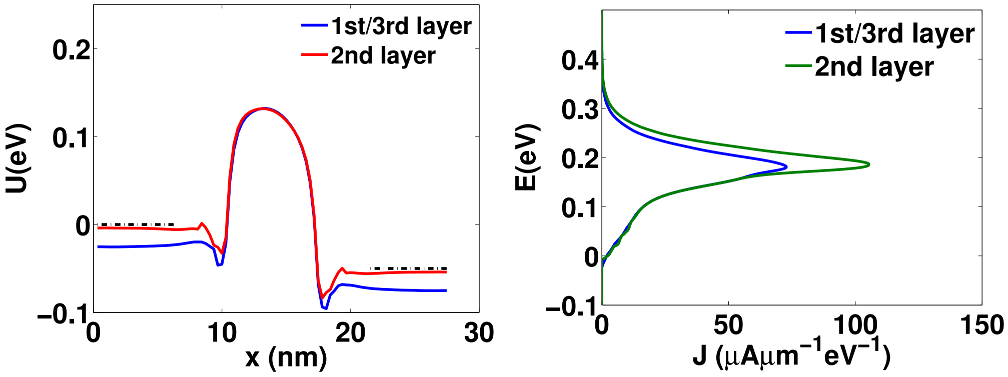
The $I_{DS} - V_{GS}$ characteristics at $V_{DS} = 0.50$ V for all devices considered are shown in Fig. 4.10a, while the ON current for corresponding ON/OFF is shown in Fig. 4.10b. The transfer characteristics follow similar trends as the previous results with DG monolayer showing the best ON and OFF currents. The SG monolayer device shows a DIBL of 70 mV/V while the bilayer device shows a DIBL of 300 mV/V, showing significant drain control at these



(a) (b)

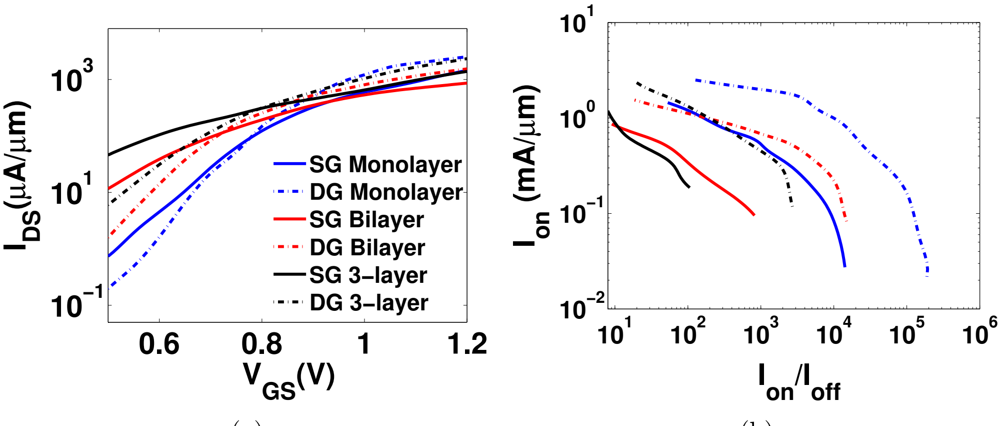
(c)

Figure 4.8: (a) Calculated $I_{DS} - V_{GS}$ characteristics of single-gated and double-gate 3-layer MoS$_2$ transistors at $V_{DS} = 0.05$ V and layer-wise current for (b) single-gate and (c) double-gate 3-layer transistor. The double-gated transistor shows better performance with higher ON currents and lower OFF currents. The 3$^{\text{rd}}$ layer provides the maximum current in the OFF state due to lower gate control while it has the lowest current in the ON state due to screening from the top layers single-gate devices. The middle layer carries maximum current in OFF state owing to higher confinement of electrons while it is screened in the ON-state for double-gate transistors.



(a) (b)

Figure 4.9: (a) Electric potential profile along the channel for double-gate 3-layer devices and (b) Current density as a function of energy at $V_{GS} = 0.75$ V (OFF state). The middle layer carries greater current compared to the surface layer due to higher confinement of eigenstates. The conduction band is further below the Fermi level for the surface layers owing to the same reason.



(a) (b)

Figure 4.10: (a) Calculated $I_{DS} - V_{GS}$ characteristics of single and double-gate monolayer, bilayer and 3-layer MoS$_2$ transistors at $V_{DS} = 0.50$ V and (b) ON current for the above transistors as a function of ON/OFF ratios at $V_{DD} = 0.50$ V. The double-gated monolayer transistor shows the best ON currents for a given ON/OFF, achieving the highest ON current and lowest OFF current. The trends for the 3 layer devices move towards worse ON/OFF ratios

scales. The DIBL for the DG monolayer device is 30 mV/V showing that better performance characteristics could be achieved through EOT scaling. The ON/OFF ratios shown in Fig. 4.10b correspond to a supply voltage of 0.50 V, while the bias window for gate voltage is moved along the transfer characteristics. The above could be achieved in practice by engineering the workfunction of the gate metal [111]. None of the SG devices were able to achieve an ON/OFF of 10^5, the minimum ratio needed to be considered as a viable alternative for low operating power transistors. Only DG monolayer device achieved an ON/OFF of 10^5, but at an ON current of 90 $\mu A/\mu m$. The trends for SG monolayer transistor is similar to the DG bilayer transistors with the bilayer device having higher current levels. The 3-layer devices show worse ON/OFF ratios because of increased screening effects.

4.4 Conclusions

To summarize, layered MoS$_2$ transistors with doped contacts and 5 nm channel length were studied using a full band, self-consistent quantum transport model within the ballistic NEGF formalism. One of the interesting observations from this full band study is the fact that in a multi-layer structure, the middle layers give the lowest energy states and would therefore fill up first. This effect is evident in our calculations for small charge levels when the gate electric field can still penetrate through the top layers. As for the current-voltage behavior, the ballistic approximation is relevant in the view that the gate length is only 5 nm. In addition, together with a doped contact, the ballistic approximation provides the best-case scenario for these devices. It is observed that the ballistic ON current for these devices for a V_{DD} swing of 0.5 V is not competitive with what can be otherwise obtained from Si, III-V or carbon nanotube devices at this channel length. This is not surprising because the injection velocity for MoS$_2$ is small due to its large effective mass. It also found that it is not possible to boost the ON current up by increasing the number of layers at this channel length because the gate electric field (for $\sim$ 0.5 nm EOT) is almost completely screened out by the layer nearest to the gate. As a result layers underneath cannot be effectively controlled by the gate, leading to significant reduction in subthreshold swing. In fact, even for a single monolayer, the short channel effect is prominent and only a double gate geometry can provide a reasonable subthreshold swing ($\sim$ 84 mV/decade). Surprisingly, this is comparable to (and not better than) what has been predicted to be achievable with surround gate small diameter ($\sim$3nm) Si nanowire transistors [72]. On the other hand, a double gate geometry for a monolayer structure may prove to be very challenging to fabricate. However, one particular aspect stands out: MoS$_2$ transistors could provide a 10^5 ON/OFF ratio even at 5 nm channel length, albeit at a small ON current level, which is not possible at all in Si or III-V due to much stronger direct source to drain tunneling. This indicates potential use in very low power applications where performance is not a critical need. In this study, we used doped contacts instead of Schottky contacts and realized that changing the doping concentration of the contacts could affect the ON current significantly, which shows the importance of contacts in 2-D materials. In the next chapter, we'll try to understand how doping affects the contact resistance in these materials.

Chapter 5

Intrinsic Limits to Contact Resistivity in Transition Metal Dichalcogenides

5.1 Introduction

In the third chapter we discussed the transistor performance of four different transition metal dichalcogenides using an effective mass Hamiltonian while using Schottky contacts. This leads to significant contact resistance because of low tunneling probability through the contact. In the fourth chapter we looked at MoS_2 in more detail while using doped source-drain contact. There we found that increasing the S-D doping leads to an increased ON current even though it could lead to increased short-channel effects such as DIBL (Drain Induced Barrier Lowering) or higher SS. This brings us to a very interesting question for few-layer devices, namely their contact resistance. The physical thinness, that provides great electrostatic control, could also lead to significant issues in terms of series resistance. Indeed it is customary to use epitaxial growth in the source and drain regions of ultra-thin body MOSFET to raise the source and drain in order to minimize contact resistance. Large field effect mobilities of up to 200 cm^2/Vs have been observed in TMDs as better contacts and appropriate encapsulation for these materials were developed [89, 24, 117]. There is a recognition of this issue and much recent work has focused on potential solutions [32, 23]. In that context, we have explored the band structure limited contact resistance for four different 2-D semiconductor materials. This is the minimum possible contact resistance that is achievable in these materials and is found to be a strong function of the S-D doping. We find that the minimum contact resistance can indeed satisfy the requirement set forward by ITRS road-map.

5.2 Simulation Approach

Bandstructure

The band structure for four different TMDs MoS_2, $MoSe_2$, WS_2 and WSe_2 were calculated using the DFT package Quantum ESPRESSO. The self-consistent field calculation was done using 8x8x1 k-point mesh using experimental lattice parameters [25][79] and PBE pseudopentials. The band structures are as shown in Fig. 5.1 with good qualitative match with previous studies. As expected all the TMDs have a direct band gap at the K point with a satellite minimum at the Σ_{min} point (between K and Γ points). The minimum at Σ_{min} is marked by a red circle for each of these materials. The distance between the conduction band minimum at K and Σ_{min} is greater for disulfides compared to diselenides. The value of this offset between the two minimum significantly affects the device performance of these materials as discussed in chap. 3 and it affects the contact resistance similarly.

Contact Resistivity Calculation

In order to calculate the contact resistivity, the number of transmission modes at a given energy needs to be computed as the resistivity per contact is the following [56][73]

$$\frac{1}{\rho_c} = \frac{2e^2}{h}\int M(E)\frac{df}{dE}dE \tag{5.1}$$

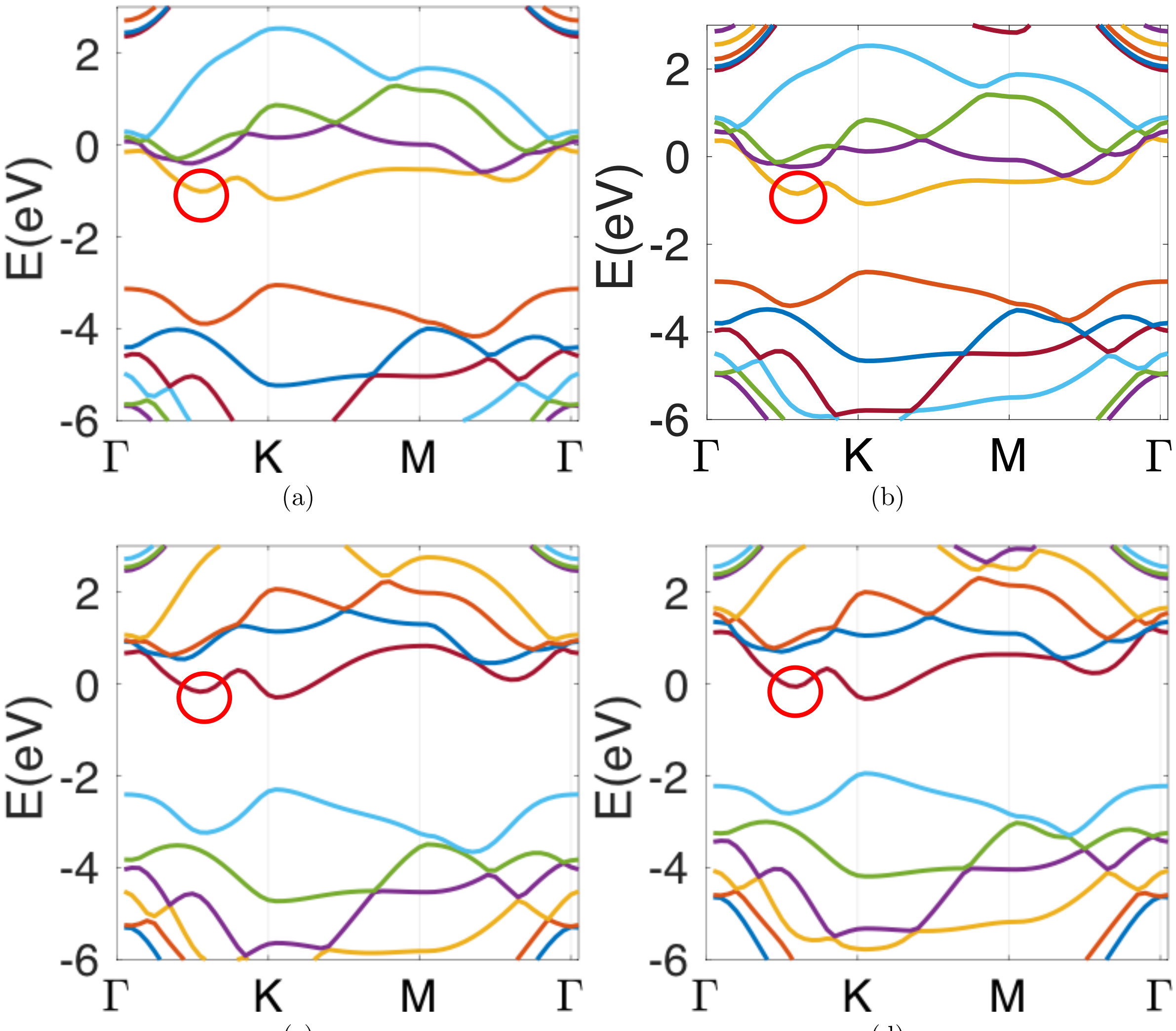
Where $M(E)$ is the density of modes (DOM) at the given energy, and f is the Fermi-Dirac distribution. The mode density can be calculated either through a full transmission calculation using NEGF, or it can be calculated via counting methods which are computationally simpler. Eqn. 5.1 essentially samples the number of available modes at the Fermi energy because df/dE behaves as a delta function around the electrochemical potential. In fact df/dE is often used in the computational community to provide a simple estimate of a delta function of width k_BT centered around the electrochemical potential:

$$\begin{aligned} \frac{df}{dE} &= \frac{\beta e^{\beta(E-\mu)}}{(e^{\beta(E-\mu)}+1)^2} \\ \beta &= \frac{1}{k_BT} \end{aligned} \tag{5.2}$$

The energy dispersion was calculated on a uniform grid using a non-self-consistent field calculation in Quantum ESPRESSO over a quarter of the first Brillouin zone in order to calculate the DOM.

5.3 Results

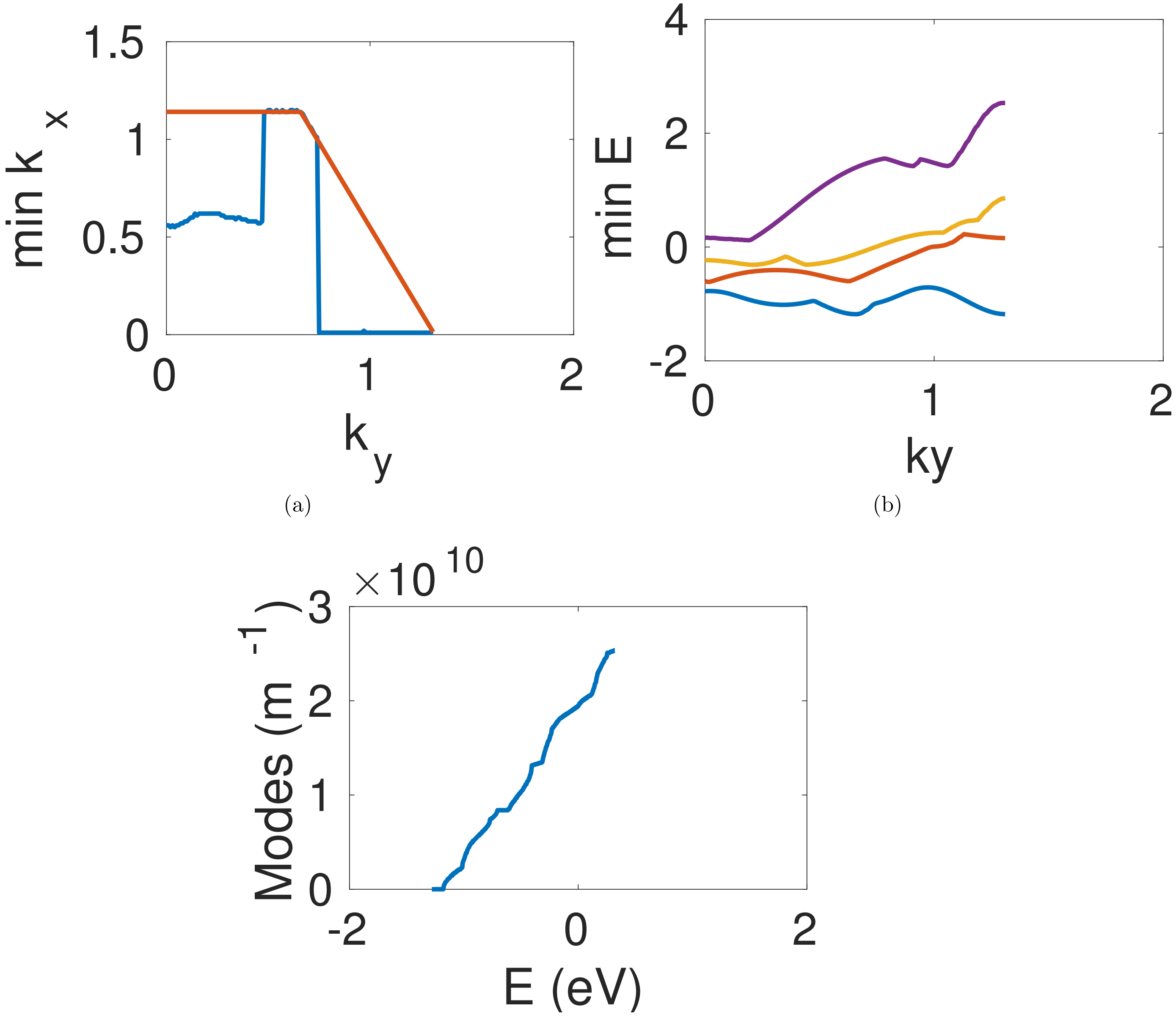
The mode density as a function of energy is shown in Fig. 5.3. As it can be seen for all the monolayer TMDs, the mode density does not keep increasing deep into the conduction band



(a) (b)

(c) (d)

Figure 5.1: Bandstructure calculations for a) MoS_2 , b) $MoSe_2$, c) WS_2 and d) WSe_2 using Quantum espresso. All of the materials have their conduction band minimum at the K point with a secondary minimum at the Σ_{min} point. The Molybdenum TMDs have lower effective mass at the K point compared with the Tungsten TMDs, which in turn affects the number of transmission modes and thus the transport properties of these materials.The difference in the second minimum at the Σ_{min} point can also be observed from the plot. This affects where the second jump in transmission modes as a function of energy can be seen (Fig. 5.3). The difference between the 1st and 2nd band is about 0.6 eV or higher for all the TMDs, showing the significant importance of the first band and its bandwidth on transport properties.



(a)

(b)

(c)

Figure 5.2: a) k_x corresponding to minimum energy (blue) and the edge of the Brillouin zone (red) for a given transverse momentum along the, b) first 4 minimum energy levels for a given transverse momentum along the Brillouin zone and c) Mode-density as a function of energy calculated using eqn. 5.3. The minimum energy contour can be seen to pass through the K point. The above approach fails to capture the limited bandwidth of the TMD materials as the first band is only 0.6 eV wide leading to a reduction in mode density. The mode density calculated using eqn. 5.3 keeps increasing as a function of energy as opposed to Fig. 5.3.

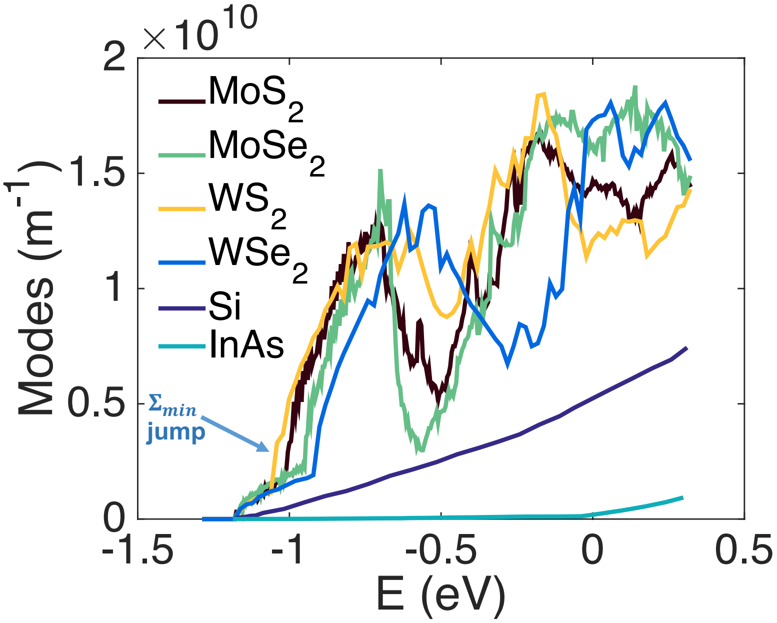


Figure 5.3: Transmission modes as a function of energy for the considered TMDs as well as silicon and InAs. The data for silicon and InAs has been extracted from [73] and reproduced in the above figure. The number of modes in silicon and InAs were calculated from [15] for one unit cell. $M(E)$ increases more sharply compared to Si and InAs owing to large effective masses in the conduction band. The second minimum at Σ_{min} (marked with the arrow) results in the second jump in transmission modes while the location of the point is affected by the gap between the conduction band at Σ_{min} and K. As expected, the Mo TMDs have higher modes just above the conduction owing to larger effective mass, while it becomes unclear at higher energies due to varying positions of Σ_{min}. Disulfides have a lower Σ_{min} point leading to a larger number of available transmission modes deep in the conduction band.

unlike silicon or a typical III-V material. This is because the band width of the first band is limited as discussed in [18]. So, the equation used in [73]

$$M(E) = \frac{1}{2\pi} \int_{BZ} \Sigma_{n^+} \Theta(E - E_{k_\perp}) dk_\perp \tag{5.3}$$

,where $\Theta(x)$ is the unit step function, cannot be used to calculate the DOM at a given energy. The minimum energy for a given transverse mode in the first quadrant of the hexagonal Brillouin Zone is given by Fig. 5.2a. We can see the Σ_{min} and the K point in the plot as $k_\perp$ is changed along the Brillouin zone in Fig. 5.2b. The corresponding mode density would be given by Fig. 5.2c which does not show the characteristic bandwidth limitation of TMD materials. Instead, the number of positive band crossings in the band dispersion are counted as a function of energy to calculate the DOM. We look at only the

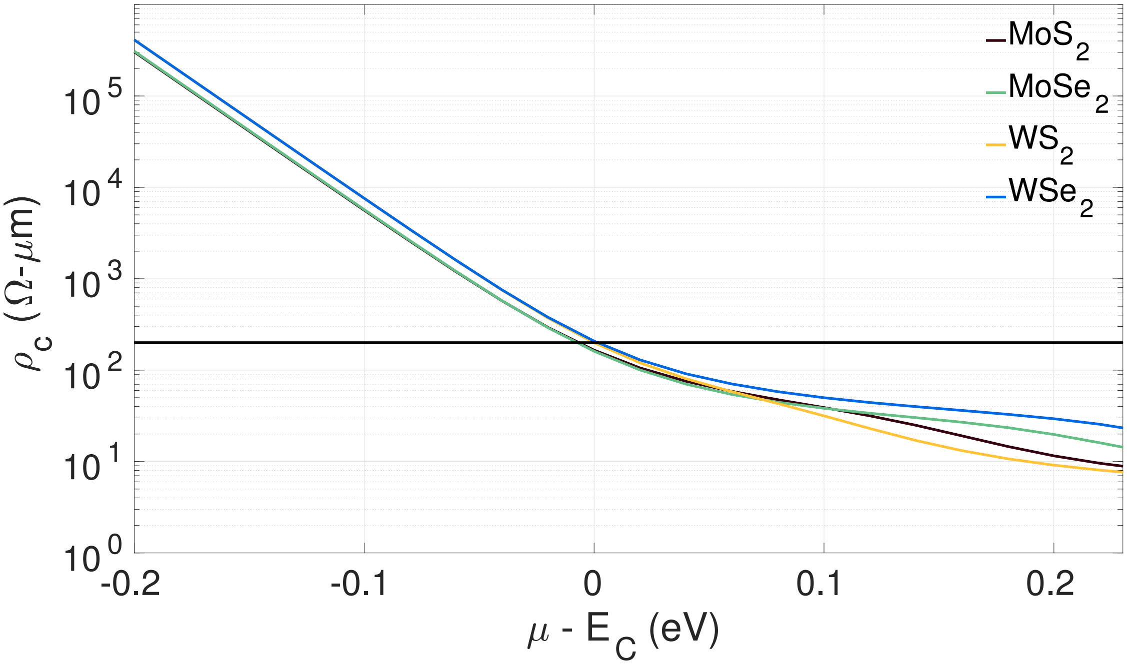


Figure 5.4: Contact resistance as a function of Fermi level for the considered TMDs, Si and InAs. The resistivity of the TMDs reduces as the Fermi level goes deep into the conduction band. The resistivity of the disulfides reduces the fastest compared to the diselenides because of closer second minimum. The resistivity for MoS_2 and $MoSe_2$ start following the same trend while they diverge for higher doping (higher Fermi levels) due to secondary minimums at the Σ_{min} point. The same effect can be seen for WS_2 and WSe_2.

positive band crossings as only they would contribute a positive velocity thus resulting in flow of current. As the mode density drops for higher energy, the resistivity can be expected to show a sudden increase beyond certain large charge densities. The resistivity as a function of electrochemical potential is shown in Fig. 5.4. The resistivity increases for higher chemical potential, but that regime is not easily attainable as it corresponds to unrealistic charge densities ($> 10^{15} cm^{-2}$). The resistivity as a function of charge density is shown in Fig. 5.5. As expected the resistivity decreases with an increase in charge density, but the relationship between the two is not very clear. For parabolic bands, it can be shown that the resistivity is inversely proportional to the square root of charge density for degenerately doped 2-D semiconductors. This is because $M(E) \propto \sqrt{E - E_c}$ and $n(E) \propto (E - E_c)$ for degenerately doped semiconductors. Fig. 5.6 shows the ratio of DOM effective mass and DOS (density of states) effective mass according to the equation [73]

$$\rho_c = \frac{h}{4q^2} * \sqrt{\frac{\pi}{2}} \sqrt{\frac{m_{DOS}}{m_{DOM}}} \tag{5.4}$$

which is valid for degenerately doped 2-D semiconductors with parabolic band structure.

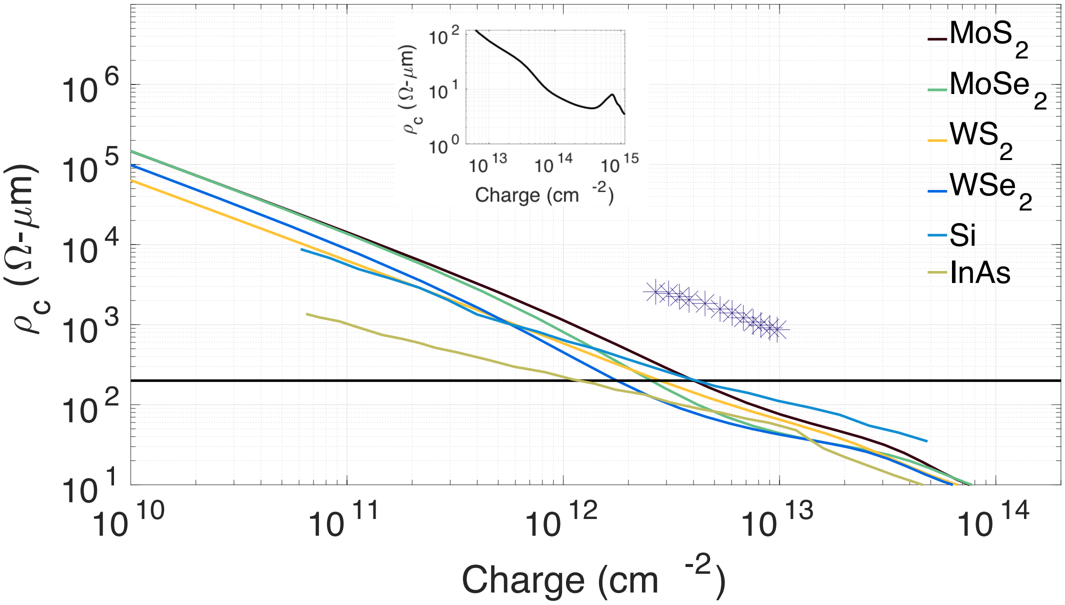


Figure 5.5: Contact resistance as a function of charge density for the considered TMDs, Si and InAs. This follows the same trends as Fig. 5.4 because the charge density increases as the electrochemical potential is pushed deeper into the conduction band. The 2-D charge density has been calculated using a single unit cell. Inset shows the resistance for large charge densities of MoS$_2$, where the reduced mode density (Fig. 5.3) results in higher contact resistance. The 2012 ITRS specs for low operating power devices is given by the solid black line at 200 $\Omega - \mu m$ in both plots. InAs has a large resistance as a function of electrochemical potential because of low DOS, but the resistance as a function of charge density lies in a similar band for all the considered materials.[73]. The experimental values for Gold contacts to MoS$_2$ [32] is shown with markers and is found to very closely follow the theoretical minimum (only higher by one order of magnitude). Low resistance contacts to WS$_2$ and WSe$_2$ has also been experimentally observed[24, 53].

We know that $m_{DOS} = \Sigma_{g_v}\sqrt{m_y m_z}$ and $m_{DOM} = [\Sigma_{g_v}\sqrt{m_y}]^2$, where m_y and m_z are the transverse and transport effective mass respectively, while g_v is the number of degenerate valleys contributing to transport. The ratio saturates to around 1 for high charge densities showing low mass anisotropy, but the ratio diverges at low charge densities because of the small offset between the band minimum at K and Σ_{min} point. The second band minimum results in a rapid increase in density of modes compared to charge density. This is because the charge density depends on DOS only through an integral over energy.

$$n = \int DOS(E) f dE \tag{5.5}$$

The contact resistance on the other hand depends on the value of mode density at the electrochemical potential and thus changes more drastically with change in the Fermi level.

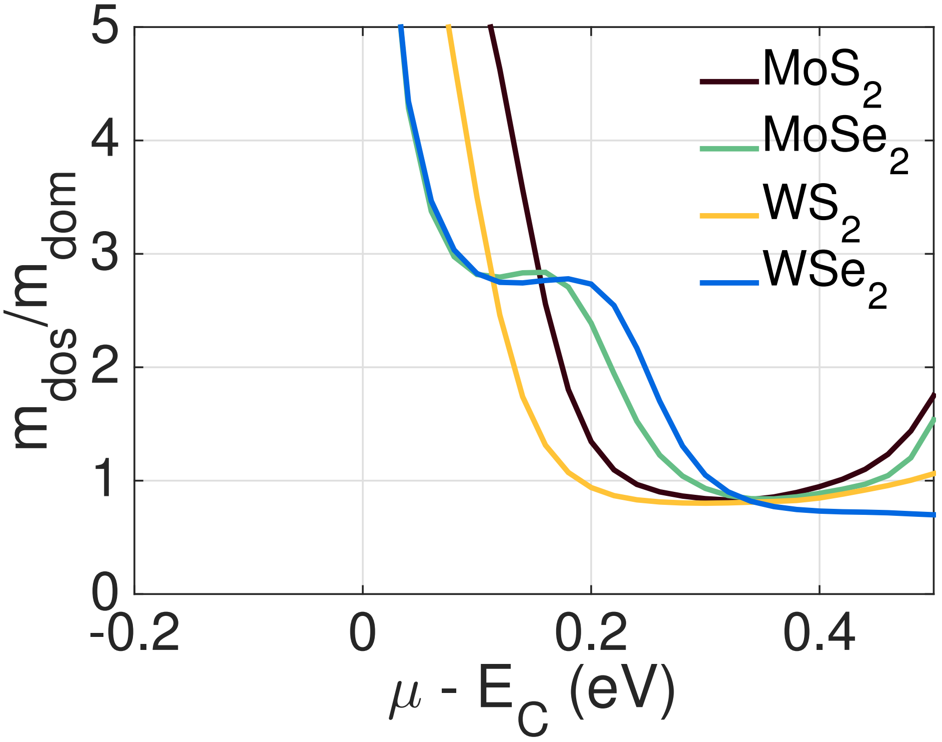


Figure 5.6: Ratio of the DOS effective mass to the DOM effective mass for all the considered materials as a function of the electrochemical potential. The ratio determines how strongly the contact resistivity is correlated with the charge density. The lower the ratio, the lower the contact resistivity for a given charge density. For the dichalcogenides, the ratio converges to a value slightly less than 1 for high charge densities showing lower mass anisotropy compared to Silicon [73]. The ratio diverges at lower electrochemical potential due to close proximity of the two band minimums at K and Σ_{min}.

Eqn. 5.4 is thus valid when only one parabolic band significantly contributes to both the density of modes and charge density.

Fig. 5.5 shows that contact resistivity of 200 $\Omega\mu m$ as specified by 2012 ITRS for ultimately scaled devices is attainable in TMDs at charge densities of $2e12 - 5e12$ cm^{-2} with MoS_2 requiring the largest charge density for the required contact resistance. As doping by ion implantation is not effective for the confined structure of TMDs, other chemical methods such as surface charge transfer doping and substitutional doping have been developed [36, 124]. Even though significant work has been done in both, charge transfer methods suffer from thermal and air instability while it is tough to form abrupt profiles with substitutional doping[24]. The layered devices suffer from a drawback compared to bulk devices due to electrical isolation between the layers owing to weak interlayer Van der Waals interaction. Thus, the overall contact resistance cannot be improved beyond the ballistic limit by increasing the transfer length as it is otherwise possible in planar intrinisically 3D devices using

$$R_c = \rho_c / L_T \tag{5.6}$$

,where R_c is the contact resistance and L_T the transfer length. Prohibitive resistance issues could arise as SOI structures are further thinned essentially reducing them to few unit cells of the bulk structure, thus reducing the effective contact length and hence increasing contact resistance [107]. We can also notice from Fig. 5.5 that the best contact resistivity achieved with MoS_2 experimentally [32] is only about one order of magnitude higher than the ballistic limit that we determined showing that the state of the art FETs have close to the minimum possible resistance. Resistivity of 0.3 $k\Omega\mu m$ and 0.93 $k\Omega\mu m$ has also been observed for 2D-2D contacts with WSe_2 and metal contacts to LiF doped WS_2 which is close to the discussed intrinsic limit [24, 53].

5.4 Conclusion

In conclusion, we have studied the intrinsic limit to contact resistance of four different 2-D semiconductor materials. We find that, due to significant non-parabolic nature of the bands, conventionally known equations for contact resistance for two dimensions cannot be used and a numerical calculation is necessary. It was shown that contact resistance requirement set forth by ITRS is achievable, but only at high doping densities. One of the drawbacks of 2-D semiconductors is the anisotropic transport in the vertical direction. As a result, the standard practice of raised source/drain in ultra thin body Si MOSFETs is not expected to be as efficient for 2-D materials. Therefore, finding ways to dope the source and drain to very high levels ($> 1e20$ cm^{-3}) will be critical. In the last three chapters, we looked at the applicability of TMDs in conventional transistor design. For the next two chapters, we'll explore two interesting spin and magnetism based applications which are unique to TMDs because of their two-dimensional nature.

Chapter 6

1D Spin Channel in 2D Transition Metal Dichalcogenide Hetero-structures

6.1 Introduction

As discussed in the previous three chapters, two dimensional (2D) transition metal dichalcogenides (TMD) have attracted significant interest in the electronics and opto-electronics community in the recent years [119, 74]. The potential to obtain excellent electrostatic confinement along with a compensated and therefore clean surface and a large band gap make them a prime choice to replace Silicon at the ultra-scaled nodes [93, 77, 108]. In addition to electronics and optoelectronics, TMD materials also enjoy large spin orbit coupling, which could enable spintronic applications for low power logic and memory [11, 122, 87, 97, 86]. In this chapter, we show that it should be possible to obtain one dimensional spin polarized channels at the heterojunction of two TMD materials by a current flow without the need for any external magnetic field. The combined effect of two unique properties of the TMD materials, namely their 2D structure and large spin orbit coupling, produce these one dimensional spin polarized channels, which are otherwise very difficult to realize in other material systems.

6.2 Working Principle

The underlying physics of 1D spin channels is based on the recognition that in a material with spin orbit coupling, electrons flowing in a specific direction could feel an effective magnetic field if an electric field exists in a direction other than the direction of the current flow. We take advantage of the band-offset at a lateral heterojunction to obtain this electric field (in the y-direction as shown in fig. 6.1[111]. If now a current is applied along the junction (x-direction in fig. 6.1), the electrons will feel an effective magnetic field in the z-direction

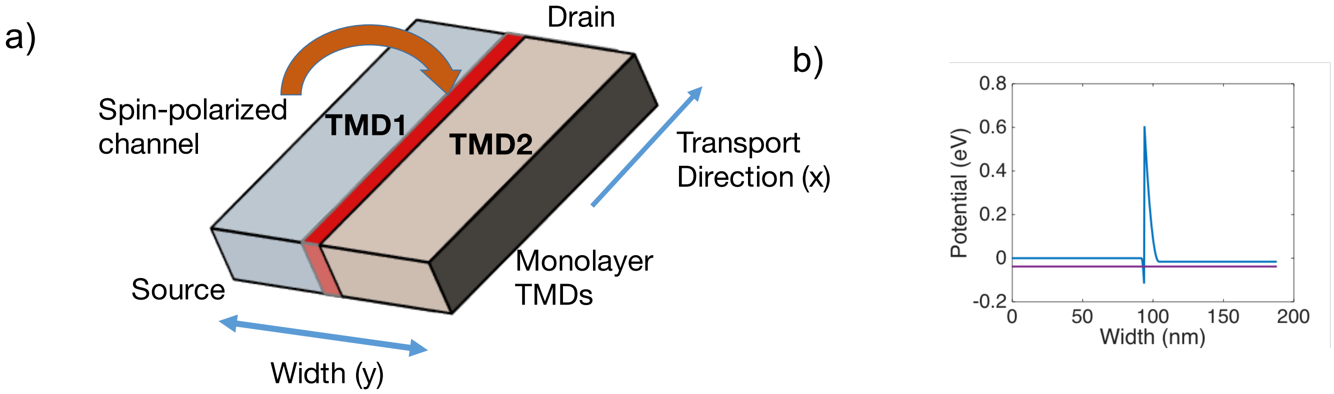


Figure 6.1: a) Schematic of the proposed device and b) conduction band along the transverse direction. Two monolayer TMD materials are deposited next to each other to form a lateral heterostructure, resulting in steep potential profiles at the junction for highly doped TMDs. The high transverse electric field creates a spin channel at the junction due to spin-orbit coupling because of a pseudo magnetic field generated in the z-direction.

and get polarized. Recent demonstration of lateral heterojunctions indicate that it should be possible to experimentally realize such junctions[39]. For our study, we investigated MoS_2/WSe_2 lateral heterojunction, where a large band offset of 0.76 eV [21] exists. When a junction is made from highly doped TMDs on either side, a quantum well appears at the MoS_2 side of the junction as the conduction band dips below the electrochemical potential (fig. 6.1b). The channel material was treated using effective mass models[76] considering only the K-point (the conduction band minimum for monolayer TMDs) in the calculation. An external voltage of 0.2 V was applied across the device in the x-direction (fig. 6.1), while the electrostatic potential was assumed to be uniform in the x-direction. As stated earlier, the band offset provides a structural electric field in the y direction ($\vec{E} = \nabla E_c/q$), which, when combined with the momentum of electrons flowing in the x-direction, leads to an effective magnetic field due to spin orbit coupling:

$$\vec{B_z} \sim \vec{E_y} X \vec{k_x} \tag{6.1}$$

where $\vec{B_z}$,$\vec{E_y}$ and $\vec{k_x}$ stand for the magnetic field, electric field and momentum. TMDs having high molecular weights have large spin-orbit coupling co-efficient because of greater electrostatic forces in the molecule. We assumed a doping concentration of about $1 * 10^{13} cm^{-2}$ to achieve steep electric fields [36]. The quantum transport was then studied within the Non-equilibrium Greens Function (NEGF) [4, 95] formalism, where the S-O coupling is included in the formalism using first order Taylor series expansion as shown in [85]. The device was assumed to have semi-infinite leads with the heterostructure extending beyond the simulation domain to avoid any spurious result from reflections. The spin-orbit coupling parameters were used from previous studies [58], where the Rashba spin orbit coupling parameter has been calculated. MoS_2 has lower splitting compared to that of WSe_2 by an order of magnitude (0.20 vs 0.033 eA^{-2}) [58] (Fig. 6.2) because of a smaller molecular mass. We calculated spin polarization in the basis of S_Z as the magnetic field and the splitting is in the z-axis.

	MoS_2	WS_2	$MoSe_2$	WSe_2
$\|\lambda_{BR}\|$ [eVÅ]	0.033 E_z	0.13 E_z	0.055 E_z	0.18 E_z

Figure 6.2: Estimates of Bychkov-Rashba coupling parameters in different transition metal dichalcogenide materials. The authors used first principle calculations to estimate the Bychkov-Rashba spin-orbit coupling parameters by deriving an effective Hamiltonian allowing them to calculate the effect of magnetic field on the above materials. Adapted from [58]

6.3 Simulation Approach

As discussed above, a first order approximation of the spin dependent Hamiltonian is included in NEGF using the Taylor series expansion. The Hamiltonian including spin for Rashba effect is given by:

$$H = \frac{p_x^2}{2m_L} + \frac{p_y^2}{2m_T} + \frac{\alpha_1}{\hbar}(p_y\sigma_x - p_x\sigma_y) + V_{conf}(x,y) \tag{6.2}$$

where $\frac{\alpha_1}{\hbar}(p_y\sigma_x - p_x\sigma_y)$ is the spin splitting from spin-orbit coupling. $\sigma_{x,y,z}$ are the Paul spin matrices, $p_{x,y,z}$ the momentum and V_{conf} the confining potential. In this case, there is an electric field perpendicular to the plane of the device. In order to use the above equation in NEGF we use the basis of local orbitals by separating out the Laplacian and the Rashba component of the Hamiltonian.

$$H_{(m,\sigma),(m',\sigma')} = -t_{(m_x,m_y)}I_s \mp it_{SO}\sigma_{(y,x)} \tag{6.3}$$

where $t_{(m_x,m_y)} = \frac{\hbar^2}{2m_{(L,T)}a}$, while $t_{SO} = \alpha_1/2a$. For our case, the equations are a bit different as the electric field is in plane. There's only one term in the spin orbit coupling.

$$H = \frac{p_x^2}{2m_L} + \frac{p_y^2}{2m_T} + \alpha|\vec{E}_y|(k_x\sigma_z) + V_{conf}(x,y) \tag{6.4}$$
$$\text{where } k_x = p_x/\hbar$$

The above equation's SOC component can be simplified as the following:

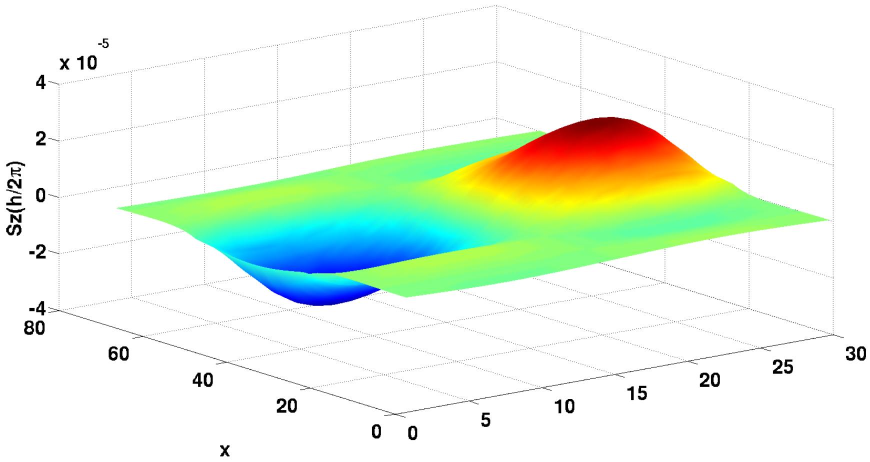


Figure 6.3: Spin density across the plane of a material of $t_0 = 1eV$ and a SOC of $t_SO = 0.002eV$ turned on in the middle of the device (x = 20 - 60) to observe the SHE due to the Rashba field. This result is similar to the one observed [85] where spins polarized out of plane but with opposite orientation flow to either side of the device due to a transverse spin-orbit force. This is not the same as the effect described in this work where the spin is accumulated only at the junction and there exists no transverse force except for diffusion.

$$\begin{aligned} H_{SO} &= \alpha|\vec{E_y}|(k_x\sigma_z) \\ &= \frac{\alpha}{2ia}|\vec{E_y}|(e^{ik_xa} - e^{-ik_xa}) \end{aligned} \tag{6.5}$$

The hopping term thus becomes equal to

$$H_{(m,\sigma),(m',\sigma')} = -t_{(m_x,m_y)}I_s + \frac{\alpha}{2ia}|\vec{E_y}|(e^{ik_xa}) \tag{6.6}$$

We did one calculation with the equation described in Eqn. 6.3 to see the familiar spin hall effect (SHE) as shown in Fig. 6.3. Here we simulate a 2D material with $t_0 = 1eV$ and $t_{SO} = 0.002eV$ of SOC turned on in the middle of the device. It can be seem from the figure that spins perpendicular to the plane of the device are generated and they flow to opposite edges of the device. The reason for the same is the existence of a transverse spin orbit force which can be calculated via the following:

$$\begin{aligned}\vec{v} &= [r, H_{SO}]/i\hbar \\ where\ H_{SO} &= \alpha|\vec{E_y}|(\sigma_y p_x - \sigma_x p_y) \\ \vec{v_y} &= i\alpha\sigma_x\hat{y} \\ \vec{v_x} &= -i\alpha\sigma_y\hat{x} \\ \vec{F_y} &= k[v_y, H_SO] \\ \vec{F_y} &= \alpha^2 p_x\sigma_z\end{aligned} \tag{6.7}$$

The device proposed in this work doesn't have any transverse force as described in subsequent sections and thus the spins are localized to the junction.

6.4 Results and Discussion

Fig. 6.4a shows the free charge density as a function of width and position in the channel from quantum transport (NEGF) calculations. As the bands are assumed to be flat along the direction of transport, the charge density is constant along the longitudinal (transport) direction, whereas the charge density does vary along the width (transverse) direction with the charge density going to zero at the extremities. The free charge density at other parts of the device follows the conduction band with the charge density going up for $\mathrm{MoS_2}$ near the junction at the quantum well, while going down for $\mathrm{WSe_2}$ as the conduction band moves farther away from the Fermi level. The spin accumulation appears at the junction as it can be seen in Fig. 6.4a, while, away from it, it decays quickly due to lack of electric field and the corresponding pseudo-magnetic field. The spin polarization as a percentage of the total free carriers can be seen in Fig. 6.4b, where a maximum polarization of 0.1% is observed for this device. The presence of a quantum well at the junction increases the effective spin density as the transverse electric field is only prominent there adding another degree of confinement to the system. Further structures can be designed which confine the 1-D spin density to regions with the largest charge density using quantum wells while having negligible charge density elsewhere leading to easier spin detection. The total spin density (fig. 6.4a (blue)) is given by the product of the spin polarization (fig. 6.4b) and the total carrier density (fig. 6.4a (red)), resulting in the profile seen at the junction, where the spin density is low in the $\mathrm{WSe_2}$ side of the device due to low carrier density. It is worth noting that we are working with conduction band electrons. As a result, the valence band splitting due to spin orbit coupling does not affect our results. The splitting in the conduction band is very small and therefore it has been ignored.

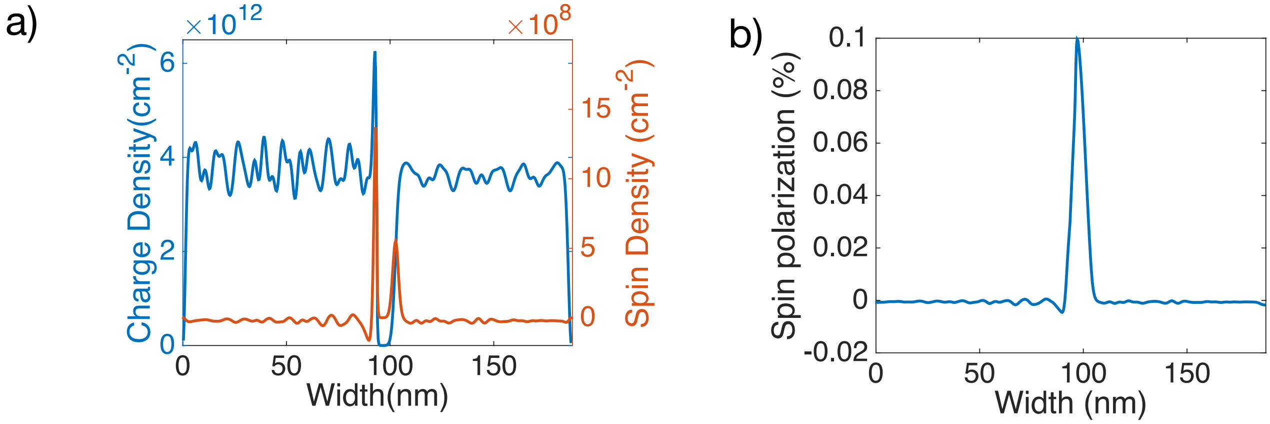


Figure 6.4: a) Charge (blue) and spin (red) density and b) Spin polarization as a function of width in the heterostructure. Charge density goes to zero at the boundaries because of aperiodic boundary conditions, while it follows the conduction band at other positions along the width. Spin polarization is observed at the junction because of transverse electric fields across the junction in combination with current flow due to non-equilibrium conditions along the junction, while the generated spin density is minimal far from the junction (due to lack of transverse electric fields). The one-dimensionality of the spin generation becomes more apparent when we consider the spin polarization as a percentage in fig. 6.1b, showing the peak at the hetero-junction. The absolute density of spin in (a) is a product of spin polarization in (b) and charge density in (a). The peaks and valley seen in a are because of the change in charge density along the width of the channel in conjunction with the spin polarization.

6.5 One-Dimensionality

Coming back to the spin channel, the generated spins are localized to where they are generated because of an absence of transverse spin-dependent force in contrast to the spin Hall effect. In case of the spin Hall effect, the transverse fields on the charged particles is spin dependent and thus the spins are separated and flow to the lateral edges of the device [85]. In this case because of the 2-D confinement, there is no transverse force on the charged particles from the pseudo magnetic field as can be seen from the following :

$$\begin{aligned} \vec{v} &= [r, H_{SO}]/i\hbar \\ where\ H_{SO} &= \alpha|\vec{E_y}|\sigma_z p_x \\ \vec{v_x} &= i\alpha\sigma_z\hat{x} \\ \vec{F_x} &= k[v_x, H_SO] \\ \vec{F_x} &= k[i\alpha\sigma_z, \alpha|\vec{E_y}|\sigma_z p_x] = 0 \end{aligned} \tag{6.8}$$

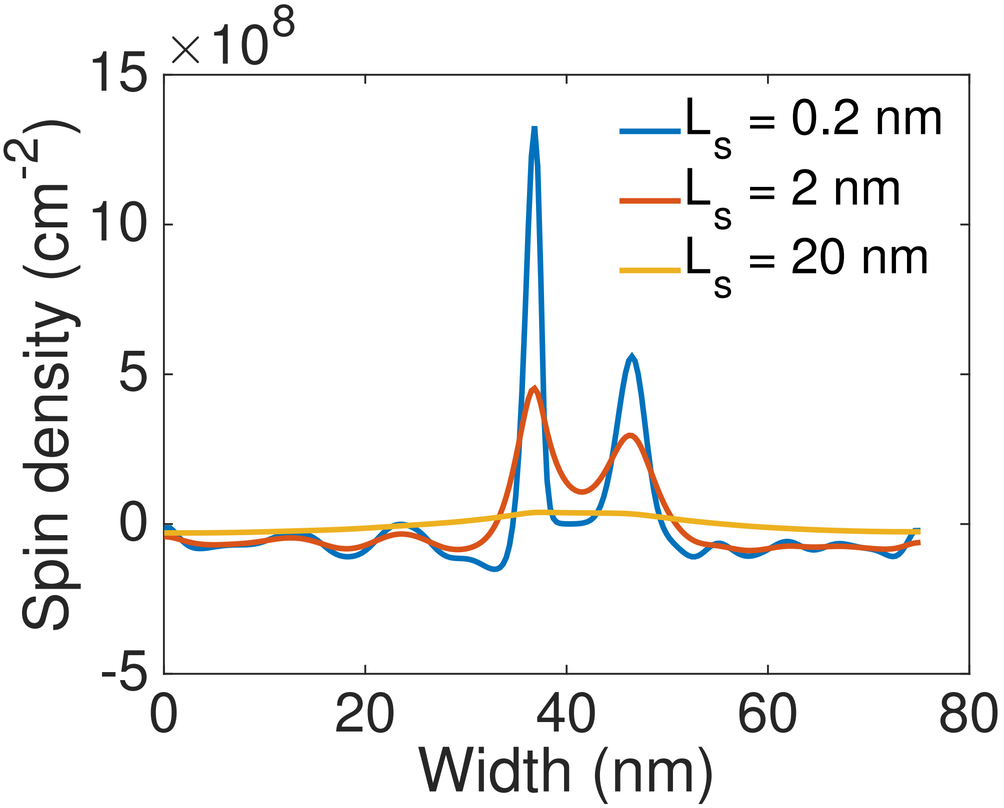


Figure 6.5: Spin polarization density for a spin-diffusion length of ballistic (blue), 2 nm (red), 20 nm (orange) and 50 nm (purple). The one-dimensional property of the spin polarization reduces with increased Ls, resulting in polarized spins getting diffused along the width of the channel. If the diffusion length is larger or comparable to the width of the device, the one-dimensionality of the spin channel is lost.

This is a critical distinction that stems from the 2D nature of the material system. The spins, however, could flow out of the junction via spin diffusion processes, resulting in a dependence of the observability effect on the spin diffusion length. With an estimated diffusion length of 20 nm, which is approximately the spin-diffusion length of MoS_2 at high doping densities [118], the change in the spin polarization is shown in fig. 6.5. The spin density in the valley at the junction in fig. 6.4a(blue) increases while the peak value of the spin density reduces as the generated spins diffuse away from the channel.

The results for different spin-diffusion length is also shown the same figure. The diffusion of polarized spin density, n_s is governed by the standard diffusion equation at steady state [110]:

$$\begin{aligned} Dd^2n_s/dy^2 - n_s/\tau_s = n_0/\tau_s, \\ D\tau_s = L_s^2 \end{aligned} \tag{6.9}$$

where D is the diffusion constant, τ_s is the spin relaxation time and n˙0 is the spin density without any spin relaxation. With increase in L˙s, the spin diffuses out more and for a large enough Ls, the one-dimensionality of the spin channel is lost. This leads to a somewhat counterintuitive situation, i.e, if there is no spin relaxation, then a pure 1D channel ensues; but in presence of finite spin relaxation, one would need to have a reasonably strong relaxation so that the spin polarization dies out very quickly in the transverse direction, leading to a pseudo-1D channel; if the diffusion length is large compared to the width of the device, the spin density along the width will be highly correlated leading to a uniform transversal distribution of spins via diffusion.

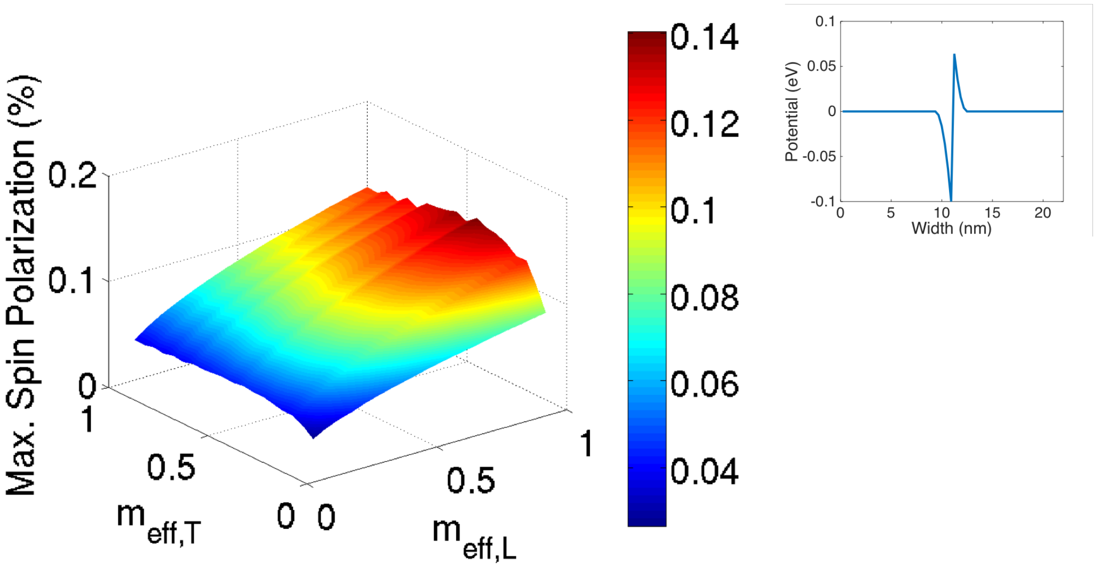


Figure 6.6: Maximum spin polarization across the device as a function of effective mass in transverse(width) and longitudinal(transport) direction. The potential profile along the width is as shown in the inset, with a band offset of 0.2 eV across the junction. The effective mass along each direction is the same throughout the device. The spin polarization increases monotonically with increase in effective mass in longitudinal direction, while it increases with transverse effective mass for lower masses saturating for higher effective masses with small ripples. The above effect arises from the modulation of average transport momentum on changing the anisotropic effective mass, while the reason for the same is discussed in fig. 6.7-6.6

6.6 Material Parameter Optimization

We further explored the effect of anisotropic effective mass on the above effect as it significantly influences electronic transport. The peak polarization observed in the device is plotted in fig. 6.6 as a function of longitudinal and transverse effective mass keeping other simulation parameters to be the same (potential profile along the width shown in inset). The polarization increases monotonically with increase in longitudinal effective mass. For transverse effective mass, however, it initially increases for the low mass values but then saturates for larger values. The reason for this can be understood as follows: The effect of spin-orbit coupling depends significantly on the longitudinal momentum of the charged particles contributing to transport (6.1). Larger momentum in the direction of transport leads to a more significant spin-splitting. For a parabolic bandstructure,

$$\begin{aligned} E &= E_c + \frac{\hbar^2 k^2}{2m} \\ k &= \sqrt{\frac{2m(E-E_c)}{\hbar^2}} \end{aligned} \tag{6.10}$$

Thus, for a given energy slice a material with larger mass has a larger average momentum. Fig. 6.7a shows the comparison of the bandstructure along the transport direction for two different materials having the same transverse effective mass of 0.3 and different longitudinal effective mass of 0.7 and 1.0 respectively. The degree of quantization depends on the transverse effective mass and thus the bands are equally spaced for both devices. However, the average momentum for each energy level is higher for the material with higher mass (see fig. 6.7b). The momentum as a function of energy can be seen in fig. 6.7c, which demonstrates the same effect at every energy which is reflected on the spin polarization as a function of energy in fig. 6.7d. Hence, the polarization increases monotonically with longitudinal effective mass. Fig. 6.8 analyzes the similar properties for two materials with the same longitudinal mass of 1.0 and different transverse mass of 0.1 and 0.3 respectively. The spacing between the bands now changes because of difference in quantization of the bands (as seen in fig. 6.8a). If any cross-section is taken for the ellipsoidal cone of the band dispersion at a given energy, the average momentum will be directly proportional to the product $m_T m_L$. For low effective masses, the quantization is not as important as the average momentum for a single band, and thus the momentum is higher for higher m_T as it is shown in fig. 6.8b. Fig. 6.8c-d show the dependence of momentum and polarization on energy respectively and thus show the increase of both quantities with m_T. The momentum as a function of energy decreases as new bands are introduced because the momentum of the states at the bottom of each band is zero. At higher levels of m_T, the effect of quantization can be assumed to be similar as that of a pulse width modulation, where the momentum drops whenever a new band is introduced, while the rate of introduction of new bands is inversely proportional to m_T. Thus the increase in average momentum on an ellipse is compensated by the higher frequency of the pulse width modulation as the effective mass is increased. Therefore, the effective momentum and consequently the spin polarization of the charged carriers becomes fairly independent of the effective mass in these regimes.

Fig. 6.9 shows the variation of spin polarization as a function of temperature. The decrease in polarization with increased temperature happens due to the larger window of energy that the effect is averaged over, resulting in contribution from states having lower momentum and thus lower splitting. Note that, in principle, a material with large spin orbit coupling could become completely spin polarized at very low temperature if the Fermi level can be placed in between the spin-split bands. However, this is not expected for current induced spin polarization at any temperature due to the fact that the electrons have zero momentum at the bottom of the band and as a result no effective magnetic field can be generated.

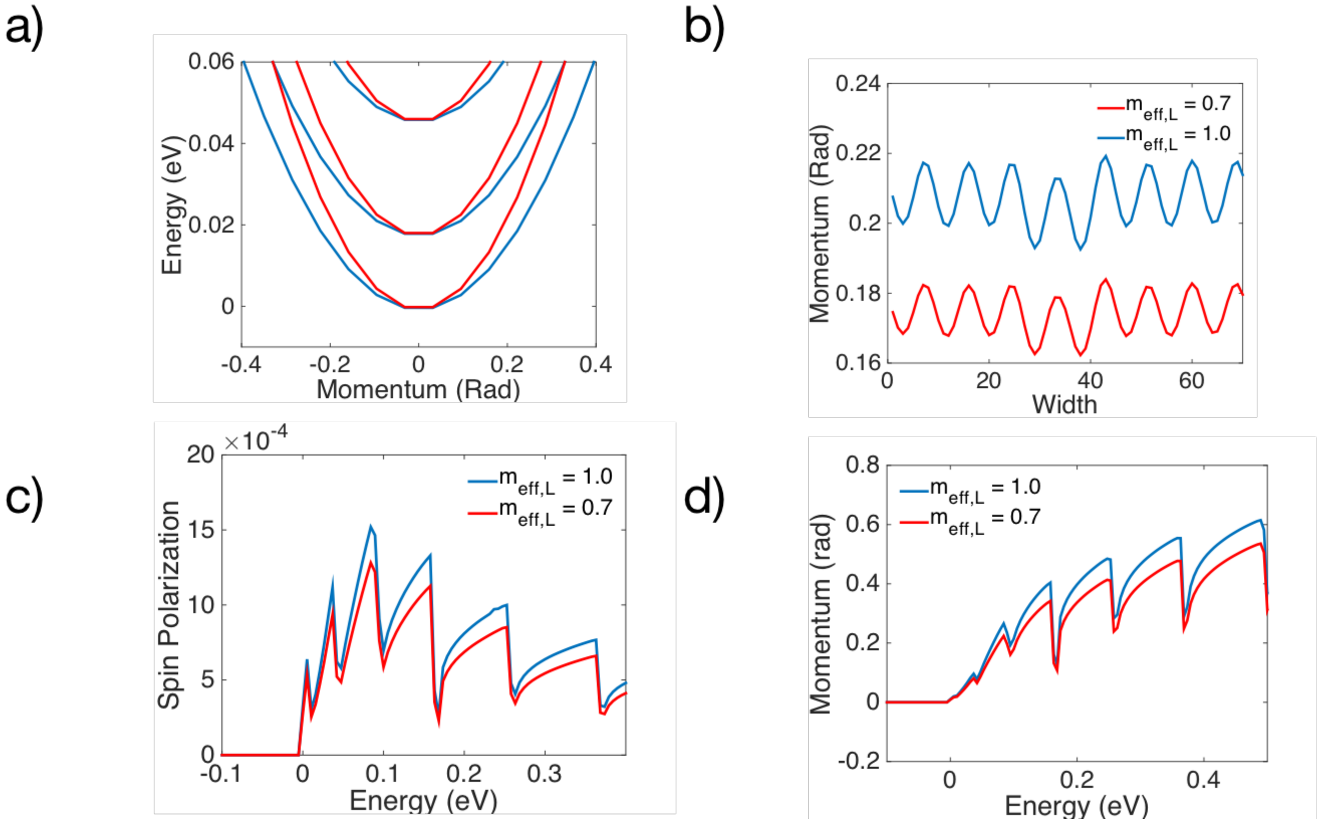


Figure 6.7: (a) Bandstructure as a function of momentum, (b) average momentum along the width, (c) spin polarization and (d) momentum as a function of energy for a transverse effective mass of 0.3 and two different longitudinal effective mass of 0.7 (red) and 1.0 (blue). Higher effective mass leads to larger momentum for a given energy as can be seen in a, leading to a larger average momentum along the direction of transport in b leading to stronger spin-splitting. (c)-(d) shows how the spin polarization behaves as a function of energy and the strong correlation between average momentum and spin splitting at a given energy.

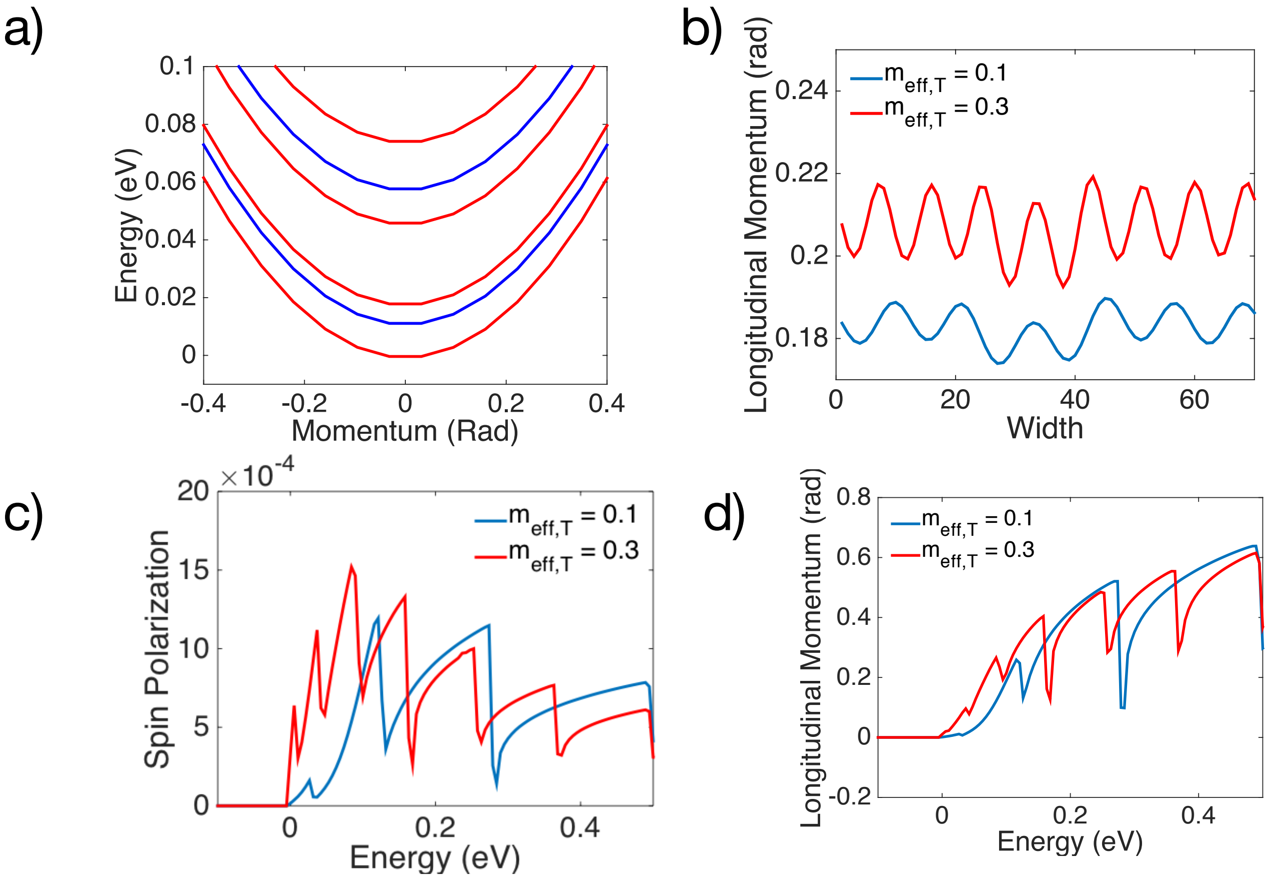


Figure 6.8: (a) Bandstructure as a function of momentum, (b) average momentum along the width, (c) spin polarization and (d) momentum as a function of energy for a longitudinal effective mass of 1.0 and different transverse effective mass of 0.1 (red) and 0.3 (blue). Higher effective mass leads to a denser quantization of bands, which could lead to lower average momentum at a given energy (because of zero momentum at band minimum). But at low effective mass regimes single band physics becomes dominant resulting in higher average momentum for larger mass. The two effects cancel out for higher effective mass (>0.4) (fig. 6.6a).

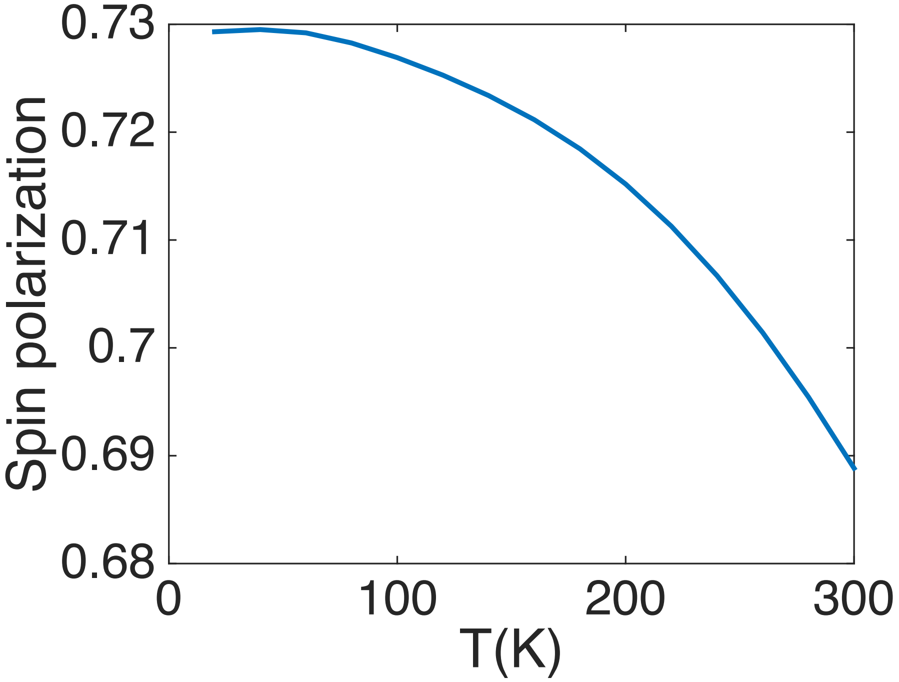


Figure 6.9: (a) Spin polarization as a function of temperature at a coupling co-efficient of 1.0 for a transverse effective mass of 0.3 and longitudinal effective mass of 1.0. There is negligible change due to temperature because of a lack of spin-splitting for zero momentum at the conduction band minimum.

6.7 Spin Polarization in a $\mathrm{n}^+ - \mathrm{n}$ junction

We have also considered spin polarization in a $\mathrm{n}^+ - \mathrm{n}$ junction as shown in Fig. 6.10. We studied this effect for WSe_2, which has a high SO coupling co-efficient of 0.18 eA^{-2}. The transverse electric field was realized using a doping profile along the width forming an n+/n junction as shown in Fig. 6.10a . The potential profile along the width is assumed to be as shown in Fig. 6.10c, with a peak electric field of 0.1 V/nm. A small Fermi gap of 0.1 eV is opened between the source and drain sides resulting in current flow between the two electrodes. A magnetic field is experienced by the charged carriers due to coupling of the current and electric field, (where and stand for the magnetic field, electric field in transverse direction and momentum along transport axis). The spin polarization resulting from the above magnetic field in WSe_2 is shown in Fig. 6.10d, while the total free carrier concentration is shown in Fig. 6.10c. As the transverse electric field goes to zero far from the junction, the effective polarization also goes to zero. A peak polarizability of 0.1 % can be seen for this device too. The polarizability can be increased by using other doping profiles for steeper electric fields. The results here are different from the previous results in that the charge is not confined to a quantum well and thus the spin density has a distinct peak as

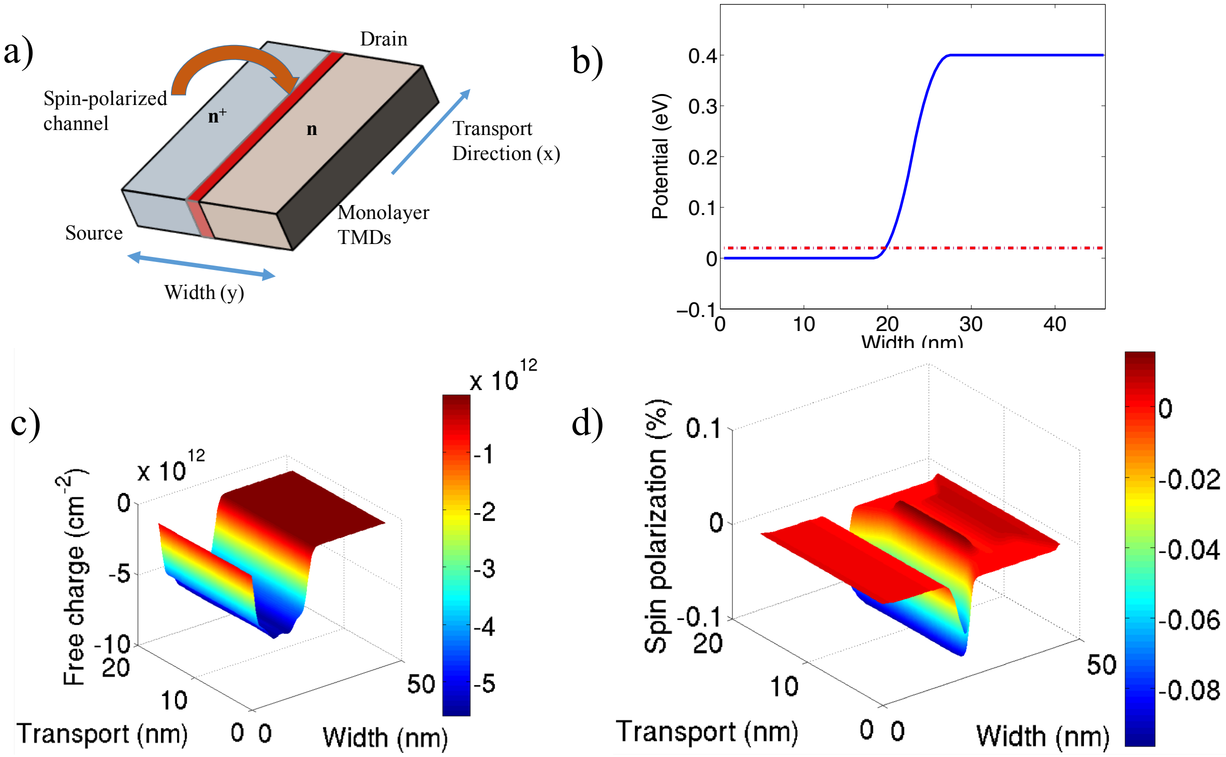


Figure 6.10: (a) shows the schematic of the simulated $n^+ - n$ device with the $n^+ - n$ junction is used to create a potential profile along the transverse direction as shown in (b). The spin polarized channel is supposed to appear in the junction of the two differently doped region where the transverse electric field exists. (c) shows the free carrier charge density along the device with higher concentration of free carriers in the n+ region. (d) shows the spin polarized channel along the junction, while there is minimal polarization far from it. Note the absence of a valley in the spin polarization in this case as opposed to the heterostructure because of a more uniform charge density distribution.

seen in Fig. 6.10d as opposed to Fig. 6.4.

6.8 Conclusion

To summarize, we have shown that a one dimensional channel of polarized spins can be produced in a lateral heterostructure of two dimensional materials without the need of any external magnetic field. While this effect is similar in nature to the well-known spin Hall effect, the two dimensional structure imposes a symmetry restriction, leading to a zero transverse force on the spins, essentially confining them in a 1D channel. We find that in a typical heterostructure of MoS_2/WSe_2 spin polarization as high as 0.1 % could be obtained at room temperature. This can be further enhanced if certain geometries are considered to increase the detectability of the above effect. For example, the electrochemical potential could be lowered so that the only region with charge density is in the quantum well in Fig. 6.1. This would reduce the valley that we see in the spin density in Fig. 6.4 and generate a spin density profile which peaks at the junction. As ongoing work on lateral heterostructures of two dimensional materials advances, such 1D spin channels could become experimentally realizable.

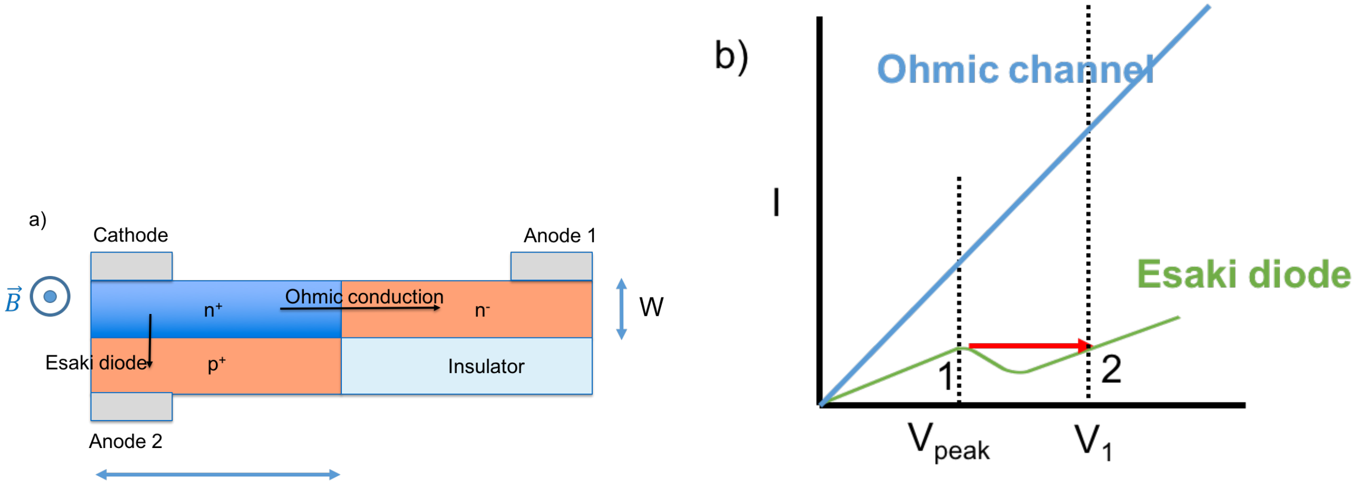
Chapter 7

On the Possibility of Observing Magnetoresistance in Esaki Diodes

7.1 Introduction

In the previous four chapters, we discussed how confined structures could have potential applications as transistors as well as in spintronics. Here, we discuss another interesting application of a confined device to achieve magnetoresistance in the absence of magnetic materials. Magnetoresistive devices have revolutionized the storage industry by providing a low power alternative to transistor based technologies [19, 129]. Recently various spin physics based devices such as STT-RAM (Spin−Torque Transfer Random−Access Memory) are being used to switch magnetic devices for magnetoresistance based memory applications [67, 128]. The most commonly used technology for detecting magnetization is to use a magnetic tunneling junction (MTJ) where the current across the tunneling junction is high (low) if the two ferromagnets on either side of an insulating layer are parallel or anti-parallel in their magnetization. Recently, MTJs have shown tunneling magnetoresistance (TMR) of up to 200% using MgO instead of amorphous insulating layers, but most commercial devices are limited by a TMR of up to 10 % [129]. As the packing density intensifies and theres an ever increasing need for higher storage capacity, there is a need to develop a device to improve the magnetoresistance read heads. Here we propose a device which uses Lorentz forces and direct band-to-band tunneling in an Esaki diode to provide MR up to 500%.

The physics of Esaki diode has been known for over half a century when it was discovered by L. Esaki [33, 109]. The Esaki diode has three regions of operation as shown in fig. 7.1b the tunneling regime, negative differential resistance and the diode like operation. Our theory relies on the fact that there are two available voltages for the peak current, V_{peak} and V_1. The proposed device schematic is shown in fig. 7.1a. Two conducting channels, one ohmic and one Esaki diode are fabricated orthogonal to each other. The Esaki diode is biased slightly below V_{peak} while the conducting channel is designed to provide a constant current flux. Now as a magnetic field is applied perpendicular to the device, current from



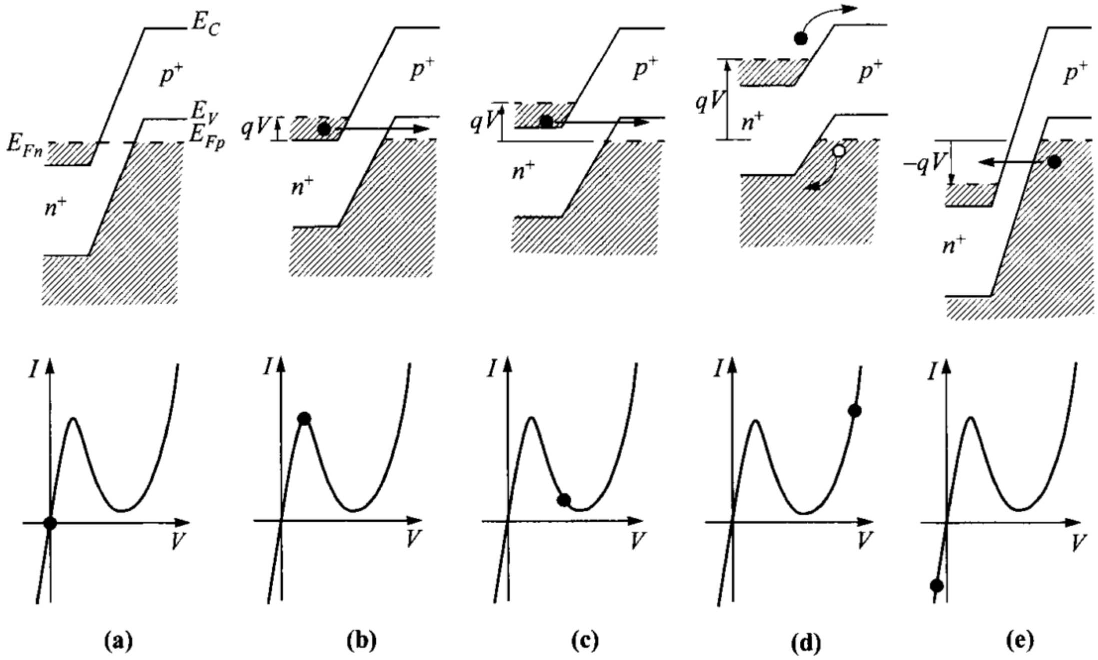
(a) (b)

Figure 7.1: a) Schematic for the proposed device and b) example I-V characteristics to show the effect. The proposed device includes two conducting channels (an Esaki diode and an ohmic contact) orthogonal to each other. Esaki diode shows NDR in its I-V characteristics with a peak voltage depending on the material and doping densities. The idea of MR revolves around shifting the peak voltage of the Esaki diode via applying a current flux due to Lorentz force on the ohmic current. The Esaki diode is initially biased at the peak voltage while A magnetic field is applied perpendicular to the schematic which pushes current from the ohmic channel into the Esaki diode forcing the voltage to jump to a higher value for a given current (V_{peak} to V_1).

the ohmic channel bends and the flux is incident on the tunneling junction or it flows away from the tunneling junction depending on the orientation of the magnetic field [116]. In this example, an out of plane magnetic field has current flux incident on the junction and vice versa. As the current flux at the junction increases, the current through the junction goes up while the magnitude depends on the tunneling probability of the designed Esaki diode. The only way for the Esaki diodes current to increase from its peak tunneling current is by jumping to its diode region of operation, which has significantly higher resistance. If the magnetic field points into the plane of the device, the current flux into the junction reduces and hence the increase in diode resistance will not be observed. This provides the desired the magnetoresistance. A device with higher MR can be designed by increasing V_1- V_{peak}, thus increasing the difference in static resistance at the two bias points. The next section explains the simulation approach for modeling the hetero Esaki diodes while the results section describes the design consideration for the above device.

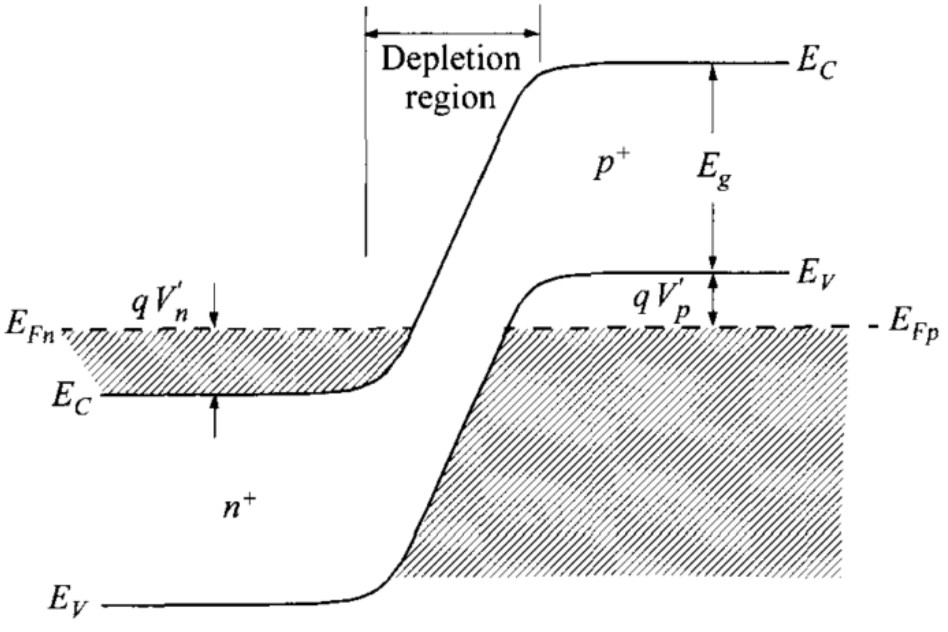
7.2 Modeling Approach

The Esaki diode is modeled using the methodology described in [111], while the material parameters for the TMDs are derived from [94][8]. The esaki diode has three regions of operation and we'll discuss how to model the regions separately in the following sub-sections. The different regimes of operations can be seen in fig. 7.2. The tunneling and diode regions of operations of the Esaki diode are modeled separately while they are brought together using empirical models.



(a)

Figure 7.2: Schematic of band diagram at a given applied voltage and corresponding position of bias in the IV characteristics at a) thermal equilibrium, b) Small forward bias leading to tunneling current, c) Larger forward bias leading to lower tunneling current (valley), d) Onset of diode regime and e) at reverse bias. The total current at bias point b) is calculated using WKB (WentzelKramersBrillouin) approximation while the current at d) is calculated using the standard diode equation. The current at c) is calculated by using an empirical formula fitted with experimental data. Most observed experimental results do not show a peak-to-valley ratio beyond 2. Adapted from Physics of Semiconductor Devices, S.M Sze and Kwok K. Ng.



(a)

Figure 7.3: Band diagram of a degenerately doped p-n junction diode at zero bias. V'_p and V'_n are the degeneracies in the n and p side respectively. Increasing V'_p and V'_n is possible via doping and it consequently increases the peak current and delays the onset of diode regime. This is because the minority carrier concentration reduces on both sides when doped to large densities. Adapted from Physics of Semiconductor Devices, S.M Sze and Kwok K. Ng.

Tunneling Regime

The tunneling regime exists in a conventional diode if both n and p sides of the diode are doped degenerately. The well-known Zener diode also works in a similar design space as illustrated in fig. 7.2e. The value of V'_p and V'_n (Fig. 7.5a)is calculated using the inverse Fermi function:

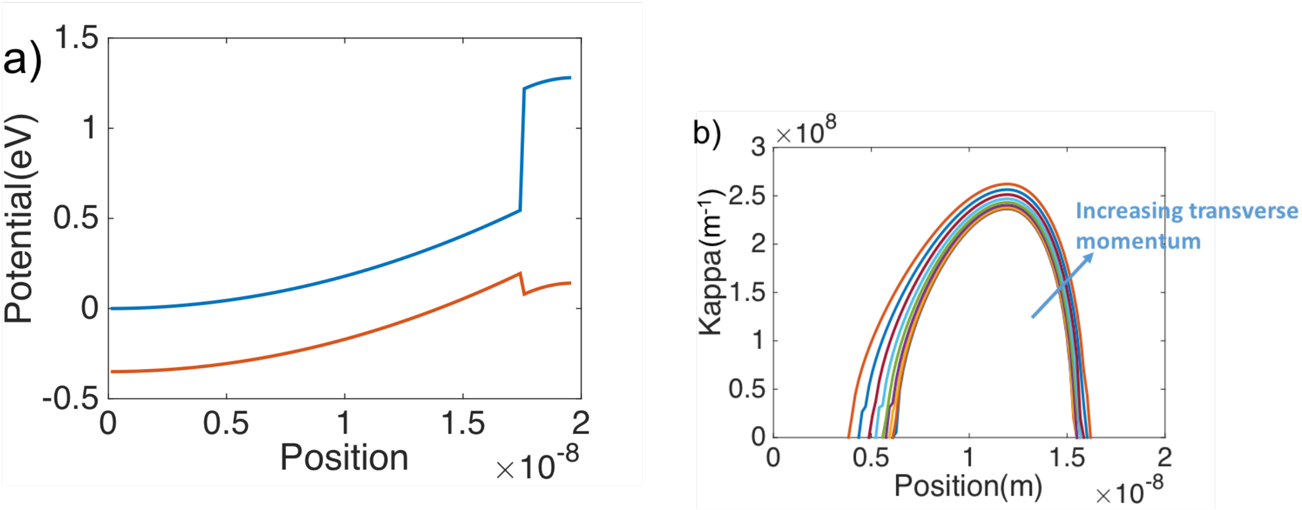
$$V'_n(p) = k_B T * F^{-1}(\frac{\sqrt{\pi} * n(p)}{2N_{c(v),InAs}}) \tag{7.1}$$

The built-in potential at equilibrium is then given by:

$$V_{bi} = V_n + V_p + E_{g,\mathrm{In_xGa_{1-x}As}} \tag{7.2}$$

The depletion width and electric field is calculated from an analytical equation by solving the 1-D Poisson's equation [111]. An example for the same can be seen in fig. 7.4a, where the potential profile for a InAs/$\mathrm{In_{0.2}Ga_{0.8}As}$ is calculated for the described doping concentration. As the doping concentration is increased $V'_{n(p)}$ increases resulting in a larger electric field in the junction. The tunneling regime can be understood from fig. 7.2a and b, where when non-equilibrium is introduced in the junction via a small forward potential, the electrons tunnel from the n-side to the p-side leading to a forward-bias current. The value of the current keeps increasing with bias voltage until the states in source and drain have the highest overlap. The ways to maximize this on-current has been discussed in past literature [3]. We use a brute-force approach because of low computational demand by calculating the current for different bias voltages until the current starts to decrease because of a lack of states to tunnel from the n-side and into the p-side of the junction. The current up to V_{peak} is calculated within the WKB approximation, where in the Kanes formula is used to calculate the transmission for a given energy and transverse momentum [38].

$$T(E, k_{\perp}) = e^{\left(-2\int \kappa(x,k_{\perp})dx\right)} \tag{7.3}$$



(a)

(b)

Figure 7.4: a) Band diagram as a function of position for a InAs/$In_{0.2}Ga_{0.8}As$ heterojunction and b)imaginary momentum as a function of position across the junction. The potential is calculated by solving the Poissons equation across the junction, while assuming uniform n-type doping of 0.03e20cm^{-3} for InAs and p-type doping of 0.2e20 cm-3 for $In_{0.2}Ga_{0.8}As$. The complex momentum as a function of position is calculated using Kanes formula [38]. The momentum increases with transverse momentum as the effective bandwidth increases thus reducing tunneling.

where κ is the complex momentum as a function of position and transverse momentum for a given energy slice. The above equation can be written as a function of energy range in the band gap as well as shown in [38]. The complex momentum is calculated using Kanes formula [51] for which effective mass parameters are used for both materials of the device.

$$\kappa(E) = \sqrt{\frac{2m_h}{\hbar^2}E(1 - E/2E_q)}, \qquad 0 < E < E_q \tag{7.4}$$

$$\kappa(E) = \sqrt{\frac{2m_e}{\hbar^2}(E_g - E)(1 - \frac{E_g - E}{2(E_g - E_q)}}, \quad E_q < E < E_g \tag{7.5}$$

$$E_q = E_g \frac{m_e}{m_e + m_h} \tag{7.6}$$

E_q is changed in both sides of the device according to the effective mass. κ can be calculated both in energy space and position as the position in the bandstructure and energy space have a one-to-one relationship. The tunneling rates for different effective masses are also added up if the bands are degenerate (For example the light and heavy hole in InAs). The energy and transverse momentum do not change during band to band tunneling owing to lack of any scatterers. For a given energy slice within the tunneling gap (energy space between the conduction band of n-type and valence band of p-type material), the transverse momentum can increase from zero until some finite value where the effective energy bang gap goes to zero. The effective energy gap for tunneling for a given bias and transverse momentum is given by

$$\begin{aligned} E_{T,eff} = E_T - V_{bias} - \hbar^2 k_\perp^2/(2m_n) - \hbar^2 k_\perp^2)/(2m_p) \\ E_T = V_n^{'} + V_p^{'} \end{aligned} \tag{7.7}$$

where $m_n(m_p)$ is the effective mass on the n(p) side of the device and $V_n^{'}$ and $V_p^{'}$ are the distances of the Fermi level from band edges on the n and p side respectively (fig. 7.4a). The doping is usually considerably lower on the n-side compared to the p-side because of lower electron effective mass of III-V materials are much lower than the hole effective mass. The diode characteristic farther away from equilibrium is calculated using the ideal diode equation assuming the quasi-Fermi levels move according to the bias conditions. The complex momentum as a function of transverse momentum is shown in fig. 7.4b for a InAs/$\text{In}_{0.2}\text{Ga}_{0.8}\text{As}$ junction with the doping concentration as specified in the caption. The magnitude of the complex momentum increases as a function of transverse momentum analogous to the increased effective band gap (7.7) and thus the integrand in 7.3 increases in magnitude leading to lower contribution from higher transverse momentums.

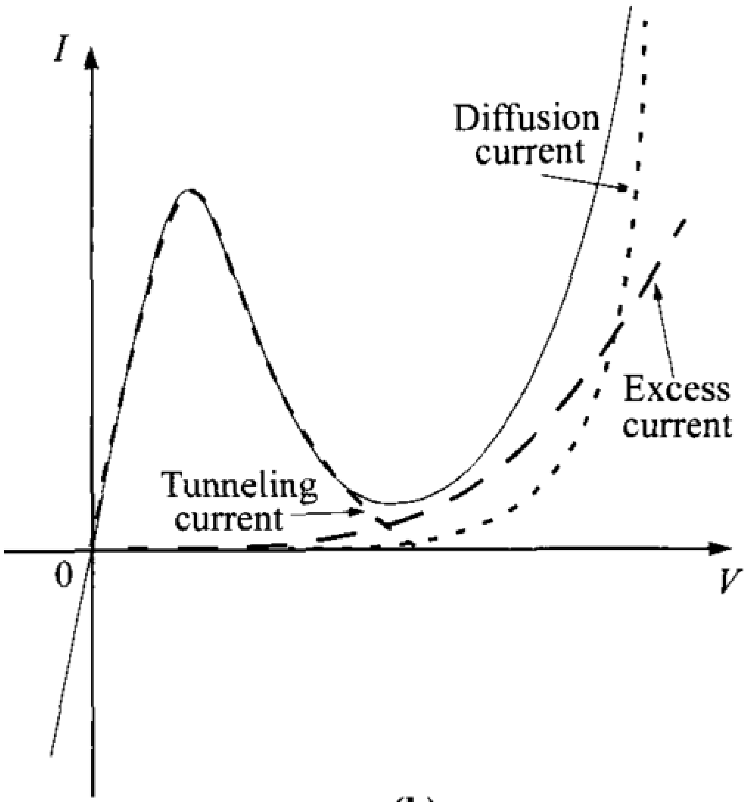
Diode Regime

The last regime of operation of the esaki diode is given by the diode's forward biased regime where the standard diode equation applies.

$$\begin{aligned} I = I_s \left(\exp \frac{qV_D}{k_B T} - 1 \right) \\ I_s = q \left(\frac{D_N}{L_N} \frac{n_i^2}{N_A} + \frac{D_P}{L_P} \frac{n_i^2}{N_D} \right) \end{aligned} \tag{7.8}$$

where $D_{N(P)}$, $L_{N(P)}$ and $N_{A(D)}$ are the diffusion constants, diffusion length of the electrons(holes) and the acceptor(donor) concentrations respectively. The diode current thus decreases with increase in doping concentration for a given bias because of both reduction in minority concentration and diffusion constant. This gives the majority of current at fig. 7.2d.

Valley current



(a)

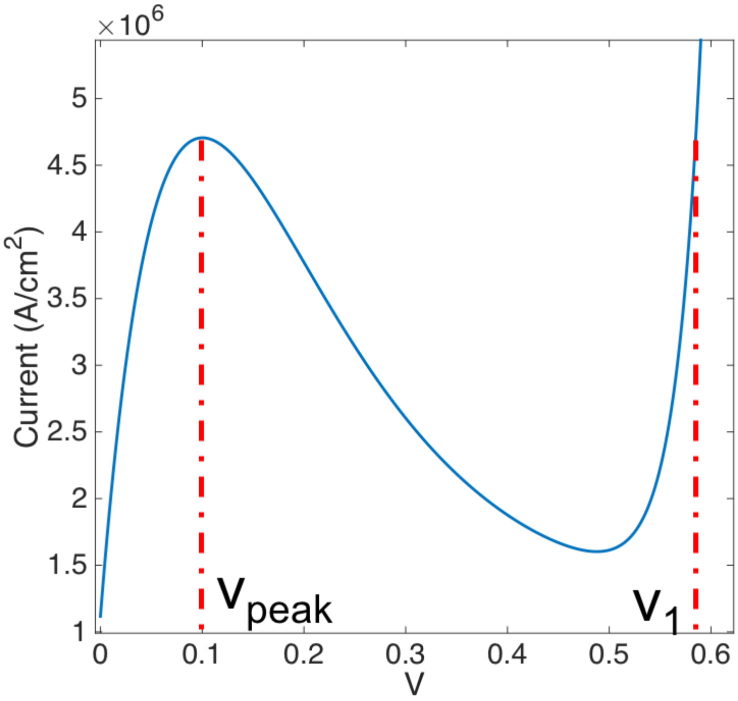
Figure 7.5: The total static characteristics are broken down into three current components The tunneling component first increases and then decreases. The excess current due to additional transport path increases as a function of voltage while the diffusion current starts increasing from the knee voltage. Adapted from Physics of Semiconductor Devices, S.M Sze and Kwok K. Ng.

Finally the current between the peak tunneling value and the diode region should ideally go to a very small value except for the presence of trap assisted tunneling in most materials. So, an electron can go through several phase changes before getting from n-side to p-side of the device. The current in this region is therefore difficult to calculate analytically and usually a empirical formula is used to express the valley current. Most known devices show peak to valley ratio of 2 to 4. So, if we chose the valley current to be equal to $J_v = J_{peak}/2$, the current in the valley is given by the following:

$$J_v = J_v \exp C_4(V - V_v) \tag{7.9}$$

where V_v is the position of the local minimum and C_4 is an empirical fitting parameter. For our calculations and the estimation of MR both these parameters are not significant as our results depend on V_{peak} and V_1 in fig. 7.6. The total current is thus given by the sum of current from the three regions of operations which is illustrated by fig. 7.5a.

A sample I-V characteristic is shown in fig. 7.6, where the difference between V_{peak} and V_1 is 0.45 V for a peak current of $4.6e6 Acm^{-2}$. The proposed MR would therefore be 300%.



(a)

Figure 7.6: Calculated IV characteristics for a InAs/$In_{0.2}Ga_{0.8}As$ heterojunction. Peak voltage and current is calculated using Kanes formalism [10], while for higher voltages an ideal diode equations are used. For the region between the peak voltage and diode like characteristic, an interpolation is used giving a peak-to-valley ratio of around 2 [6].

7.3 Design Considerations

Although the expected MR from this device is significant, it has certain physical constraints on its aspect ratio for enough current flux to be diverted into the tunneling junction. The two parameters that can changed in the device are W and L as shown in fig. 7.1a. W dictates the amount of current that is available to be directed into the tunneling junction, while L dictates the amount of current flux needed to observe the MR. When a magnetic field B is applied on a conducting channel carrying a current density J with a mobility of μ, the deflected current perpendicular to the magnetic field (in the direction of Lorentz force) is give by $J\mu B$ [11]. If the out of plane thickness of the device is t, the total current being carried by the ohmic junction is given by tWJ, while the current deflected into the tunneling junction when a magnetic field is applied out of the plane is given by $tWJ\mu B$.

Now, suppose the current density through the tunneling junction is J_{tun} and the tunneling probability is T, the flux at the junction is given by

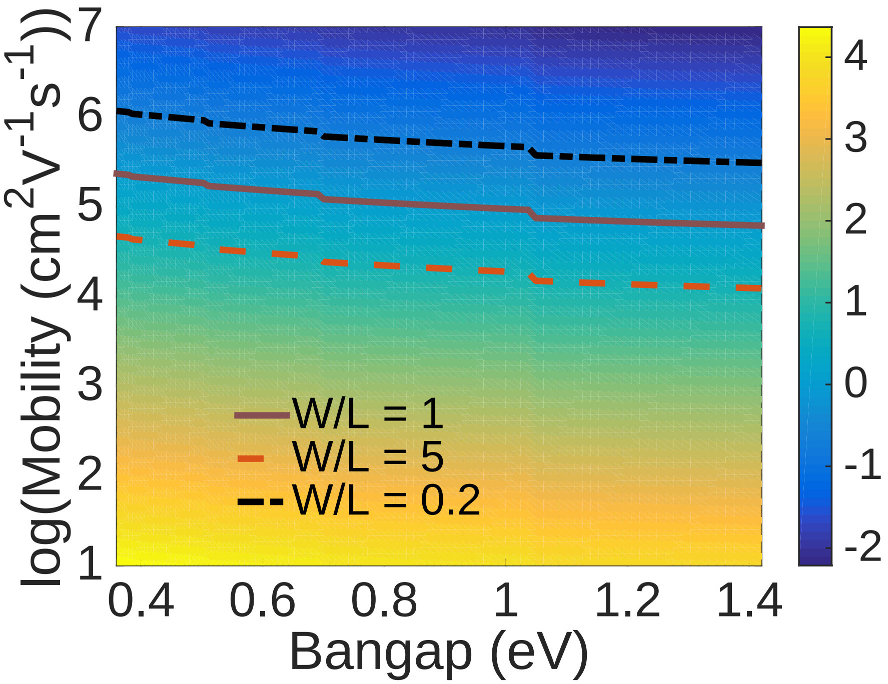
$$\frac{J_{tun}}{T}tL \tag{7.10}$$

To move the diode operation from V_{peak} to V_1 the deflected current from the ohmic channel has to provide an increase in current flux into the junction. To provide a 5% change in the flux, the following equation has to be satisfied

$$tWJB = 0.05\frac{J_{tun}}{T}tL \tag{7.11}$$

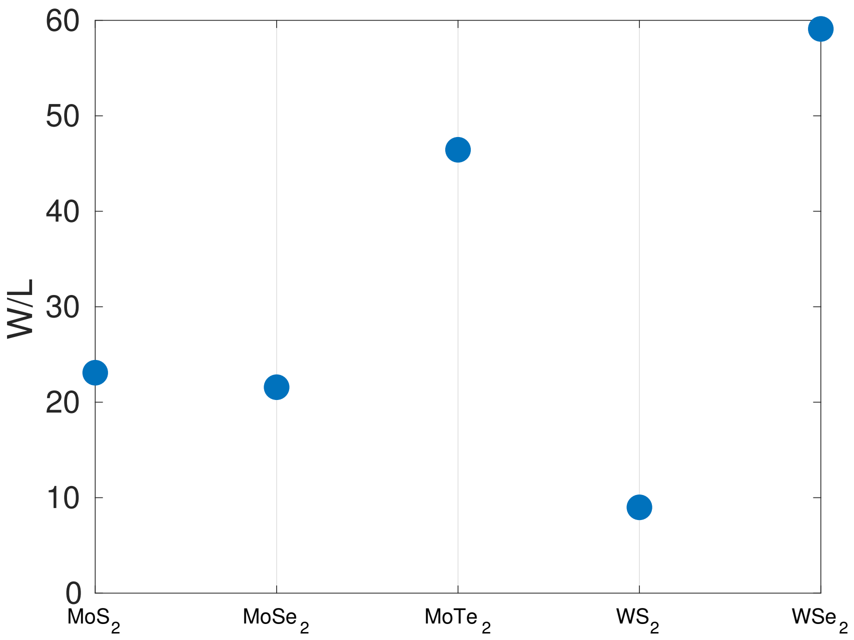
$$\frac{W}{L} = \frac{0.05J_{tun}}{TB} \tag{7.12}$$

Fig. 7.7 shows the design constraints on an InAs/$\mathrm{In_xGa_{1-x}As}$ junction. X was varied from 0 to 1 for the p side of the device thus changing the band gap from 0.36 to 1.42 eV [12]. InAs is used in the n side of the device because of its high mobility. The plot shows the aspect ration $\frac{W}{L}$ as a function of mobility of the n-type material and the band gap of the p-type material while assuming a constant current flux of $10^6 A/cm^2$ is flowing in the ohmic channel. As expected from eqn. 4 the aspect ratio reduces for a material of high mobility while it changes very slightly for change in band gap. This is because the $\frac{W}{L}$ depends on $\frac{J_{tun}}{T}$ and both J_{tun} and T decreases for a material of higher band gap thus rendering the flux into the tunnel junction to be fairly constant across different materials. Fig. 7.8 shows design consideration for TMD devices which can be layered to provide materials of varying thickness [13]. Fig. 7.8a shows the possibilities with InAs in the n side while 5b shows the possibilities with a TMD on either side of the tunneling junction. $\mathrm{WS_2}$ shows the best possibility with an aspect ratio of 8 for a current density of $10e6Acm^{-2}$ in the ohmic layer, while for TMD-TMD heterojunction we see a best possible ratio of 30. The aspect ratio for TMDs increase because of increased band gap and effective mass. The calculations shown are for a high mobility of $10^4 cm^2/Vs$. Even though TMDs were theoretically expected to have

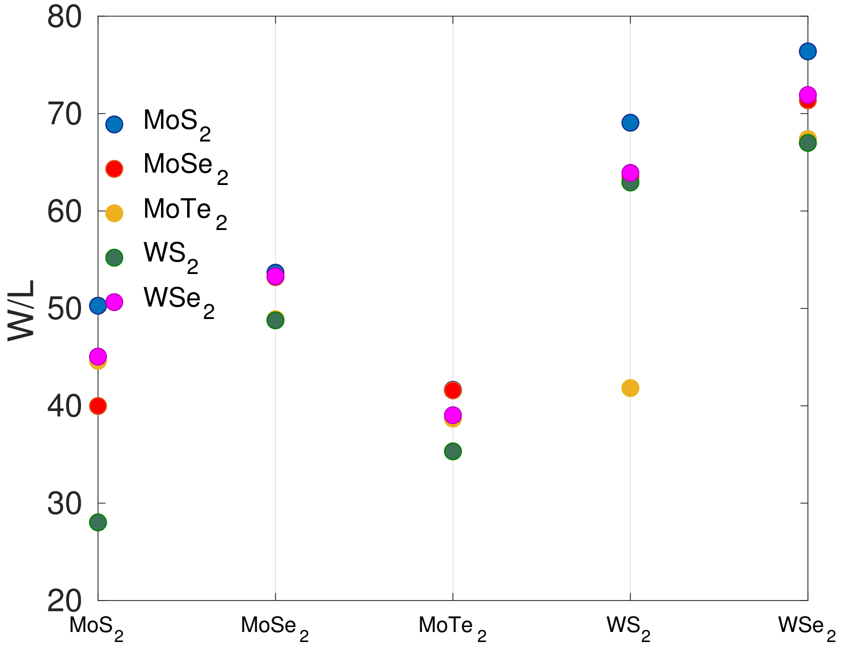


(a)

Figure 7.7: Required mobility for a current flux of $10^6 A/cm^2$ in the ohmic channel to provide 5 % of the peak current flux flowing into the Esaki diode junction for a given band gap. The band gap is set by the composition of $\mathrm{In_xGa_{1-x}As}$. Mobility of the ohmic layer is crucial to efficiency as the current bending is proportional to B due to Lorentz force. The required mobility can be decreased by increasing the aspect ratio of the ohmic layer.



(a)



(b)

Figure 7.8: Required W/L ratio for a) InAs/TMD heterostructure and b) TMD/TMD heterostructure assuming mobility of $4 * 10^4 cm^2/Vs$. The lowest aspect ratio could be achieved for a InAs/WS$_2$ junction and a WS$_2$/MoS$_2$ junction for a current flux of $10^6 A/cm^2$ and a magnetic field of 1 Tesla. Materials with higher mobility are desired for a feasible device. TMD devices suffer because of high effective masses and large band gaps thus increasing the required deflection of current flux.

high mobility, the experimental values seen have been considerably lower thus increasing the constraints for a TMD-TMD junction.

7.4 Conclusion

The gain in significance of magnetism in the storage application has led to the search for faster and cheaper detection of the storage state. In this work, we proposed a technique to detect MR of up to 400% using a tunneling junction and exploiting Lorentz forces with its primary advantage being that it doesn't rely on magnetic materials for the detection. Engineering materials of high mobility and large aspect ratio could enable this device to potentially replace MTJ based detection in the future, but there is a gap between the current technology and required aspect ratio for these devices.

Chapter 8

Conclusion and Future Work

In this thesis we looked at TMDs, while discussing their potential in low power transistor applications. While looking at four different TMDs, we went into further detail in investigating MoS_2 transistors using more rigorous simulations to look at its electrical performance in the ultimate scaling limit. Finally, we considered the effect of doping on the contact resistance in TMDs to understand how the current state of the art compares to maximum achievable limits. In addition to looking at applications in the low power electronics, we proposed two other novel applications based on heterostructures of confined materials - a one-dimensional spin channel and an Esaki diode that could show large magneto-resistance. In this chapter, we will summarize the results that we arrived at in this thesis while also highlighting some questions that arose from the discussions.

Quantum Transport Simulator

We developed a quantum transport simulator within the NEGF formalism as described in Chapter 2. The scaling of transistor channels to the ultimate scaling limit of 5 nm gate lengths necessitate a full quantum mechanics based transport solver which can take tunneling and inter-valley interactions into account. The simulator was parallelized over the energy domain, with it being embarrassingly parallel by using sub-routines within MPI (Message Passing Interface). Scalability of up to 5000 processors was demonstrated in the developed simulator. In the process, we implemented two different algorithms - the recursive Green's function (RGF) algorithm and the nested dissection (ND) approach to tackle the computationally intensive task of calculating the Green's function of the device by inverting a largely sparse matrix. The simulator was put into use in studying various aspects of low-dimensional electronics in chapter 3-6 the results of which we'll summarize in the following sections.

8.1 Low power electronics applications

Comparative Studies of Various TMD Materials

In chapter 3, we investigated four different TMDs -MoS_2, $MoSe_2$, WS_2 and WSe_2 using an effective mass Hamiltonian. For the electronic properties of the TMDs, DFT calculations were done using VASP with GGA-PBE pseudopotentials to obtain the inter-valley distance in energy space between the K and Σ_{min} points as well as the curvature at each of the conduction band minima. The transport was then studied by using metal contacts at the source and drain for one to five layer devices to see if any increase in current could be obtained by increasing the number of layers. From the results, we realized for the first time that screening effects can be significant in these layered devices because of the large DOS (large effective mass) - therefore, monolayer devices showed the best transistor performance for all the different TMDs. Any increase in the number of layers would lead to a degradation of electrostatics as the layers further from the gate was shielded by the charge in the layers closest to it [76]. One of the most interesting results was the absence of direct source-drain tunneling for any of these devices because of large effective mass which reduces the probability of tunneling - most of the current was found to be at the top of the barrier thus eliminating significant short channel effects. It was also observed that there was a difference in transistor performance of different TMDs s the effective mass of each of them were different, for example MoS_2 showed the best ON/OFF ratio for a given on current compared with WSe_2 as the larger effective mass of WSe_2 led to degraded electrostatics because of large quantum capacitance.

Screening Effects in Ultimately Scaled MoS_2 Transistors

Even though we looked at transistor applications of different TMDs in great detail in chapter. 4, there was a need for a more rigorous study of a TMD which included all the relevant physics from the band structure for different layered materials. We used a TB Hamiltonian including the s,p and d orbitals to calculate the transistor performance of 1-3 layered MoS_2 transistors. The first band structure effect we observed was a large surface energy in the outer layers of a 3-layer MoS_2 crystal [77]. This leads to the middle layer carrying larger carrier concentration in the OFF state when the gate is not screened by the outer layers. Similar screening effects as seen in the previous study was also observed here with monolayer transistor showing the best performance with it being the only transistor achieving an ON/OFF ratio of 10^5 . Interestingly, double-gated bilayer transistors had similar performance as single-gated monolayer transistor showing that each layer of the former screened the other layer from the farther gate. We also saw negative differential resistance in the output characteristics which is a purely bandstructure effect arising from the bandwidth-limited first band. The above effect can be mitigated with the inclusion of scattering so that electrons are scattered to bands higher in energy as shown in [108]. Further studies on scaling such as thinning of gate oxides and doping of contacts revealed how important contact engineering will be for these layered materials.

Effects of Doping on Contact Resistance

In chapter 3, we used metal contacts to see the behavior of few-layer TMD transistors at the ultimate scaling limit, while moving to doped contacts in chapter 4 to avoid reflections from mismatched contacts to capture the maximum drivability of MoS_2 transistors. The electrostatic short channel effects were found to depend significantly on the doping concentration of the source-drain contacts as the SS degraded for larger charge in the contacts. On the other hand, the ON current improved with a higher doping concentration. Further, theoretically predicted large mobility of these confined materials had not been observed in early experimental demonstration of TMD transistors because of the good contacts to these materials had not been engineered yet. In chapter 5, we calculated the best possible band-limited contact resistance achievable in TMDs and realized that the best results seen in recent studies were very close to the theoretical minimum. Overall, contact resistance could be a bottleneck for TMDs or any layered materials as the number of modes available to inject charge into the channel are limited by the weak Van der Waals interaction between the layers, while raised S/D contacts are used to compensate for the same in SOI structures. With this, we conclude the studies on the applicability of few-layer TMDs in conventional transistors. We realized that it is possible to achieve very good ON/OFF of 10^5 ratio due to absence of leakage due to tunneling and excellent electrostatics, while balancing the trade-off with the ON current levels will be the primary challenge in large scale implementation of these devices.

8.2 Novel spintronics applications

One-dimensional Spin Channel

We looked at two novel applications of heterostructures of layered materials where the confined nature of the same could help in achieving unique physical phenomena. In chapter 6, we investigated the possibility of realizing a 1-D spin channel at the junction of two different TMDs when a non-equilibrium condition is applied along the junction. We realized that it was possible to achieve a spin polarization of up to 1% in a junction of MoS_2 and WSe_2, while we discussed how the effective mass could be varied to change the magnitude of the effect. The one-dimensionality of the spin polarization was dictated by the lack of transverse spin force unlike the well-known spin-orbit splitting induced spin hall effect and lack of an out of plane momentum. Even though the effect was found to be small, the realization of a 1-D spin channel could be a very interesting result especially because of the unique ability to control the emergence of it via atomically sharp junctions in 2-D heterostructure. This adds to the spin transport functionalities in TMDs that have been explored in other studies such as the induction of "designer" spin-orbit interaction in van der Waals heterostructures of layered materials resulting in topological insulating states and other quantum spin hall effect devices [120, 92].

Large Magneto-resistance in Esaki Diodes

In chapter 7, we explored how large magnetoresistance could be realized in Esaki diodes by using a magnetic field to modulate current flux into the junction. Depending on the direction of the magnetic field, the Esaki diode could operate in two different bias points of different resistance thus allowing detection of the magnetic field. We looked at different material combinations to understand the design considerations such as aspect ratio that was needed to observe the effect so that the devices could be physically realized. Engineering layered materials with large mobility and aspect ratio could help in fabricating such devices in the future as the quest for large MR devices continues to realize more efficient storage devices.

8.3 Future Directions

In this thesis, we found that few-layer TMD materials have a significant potential in replacing Si for low power electronics, while there's a need for substantial studies to be carried out to address challenges concerning layered materials. A major bottleneck in achieving the appreciable current levels and mobility in few-layer materials has been the development of good contacts to the channel, the lack of which has artificially reduced the mobility of TMD transistors. At the same time, proper encapsulating environments, such BN has been found to enhance the mobility [64]. There is a need to understand how the encapsulation and how fields from the dielectric and the substrate affects the mobility of vdW materials via increased scattering, which could help in designing higher mobility devices in the future. As better contacts to few-layer TMDs are being developed [32, 23, 24, 117] it has also become essential to understand various transport mechanisms from metal contacts to the layered TMDs [71]. The optimization of contact materials with different electronic structure properties in order to reduce reflections at the contact could become important in the ultimately scaled limit as reflections at the boundaries assumes a significant portion of the device resistance. We realized that the current state of the art is very close to the intrinsic limit of contact resistance, and it'll be essential to carefully engineer the contacts to be competitive with conventional materials.

In bulk materials, as the Σ_{min} valley gets closer to the K valley, it could be an interesting study to look at the anisotropic transport in these materials given the Σ_{min} has an anisotropic mass at the conduction band minimum. This could result in interesting physics as the direction of transport can be optimized to increase the injection velocity while keeping appreciable quantum capacitance. We also realized a need for dissipative transport as the lack of inelastic scattering events led to an artificial NDR behavior in the MoS transistors. Including scattering physics possibly increases the ON current in these materials when doing dissipative calculations compared to ballistic calculations thus improving the drive current. Ultimately, a comprehensive device model accurately including contact physics, reflections at the boundary, bandstructure effects and scattering could be developed for better quantitative estimates of transistor performance.

Other Low-dimensional Materials

Along with TMDs, other low dimensional materials are also being explored for ultimate scalability in order to improve the electrostatics. Black phosphorus for example is a layered material with a highly anisotropic effective mass, with almost an order of magnitude difference between its longitudinal ($0.17\mathrm{m_e}$) and transport effective mass ($1.20\mathrm{m_e}$) which enables it to have a large average carrier velocity as well as a large DOS [61]. At the same time, silicon nanowires as well as confined device structures made of III-V materials such as InGaAs are also being studied for application in the next technology node. Even though the electronic structure for these materials in their bulk phases are reasonably well understood, studies need to be done to understand the changes in their electronic properties when they're confined. In highly confined structures, the minimum energy of the sub-bands increase in energy resulting in a larger band-gap, which could reduce leakage current in low band gap materials like InAs, but could render the transistors to be limited by the low quantum capacitance as the gap between sub-bands increases. There is a need however to understand the quantitative extent of this effect as wavefunction penetration of the semiconductor material into the gate oxide is included in the modeling efforts thus mitigating the limitation of quantum capacitance.

There could also be a need to move away from RGF for NEGF calculations for highly confined structures such as Si nanowires as the aspect ratio of these devices could make it more amenable to the nested dissection approach where more complex sparsity can be exploited in addition to the block tridiagonal nature of the Hamiltonian. For example, the size of the Hamiltonian for a slice of 4 nm InGaAs nanowire could be as big as 2000 for a sp^3d^5s* TB Hamiltonian, which would result in a prohibitively large computational burden if the conventional RGF algorithm is used. It could be interesting to see the computational advantages that new algorithms can provide for these ultra-confined materials.

Non-idealities in Transport

Even though ballistic transport gives a good qualitative estimate of the performance of a scaled device, further physics may be included in the future to account for scattering. Inter-valley scattering due to acoustic phonons has been found to mitigate the fake negative differential resistance as described in chapter 4, while the inclusion of self-heating could show degraded drive currents as the effects of scattering increases. We also saw singularities in the Hamiltonian, when calculating the Green's function for layered $\mathrm{MoS_2}$ transistors which needed to be filtered for calculating the current. These singularities were dependent on the broadening levels used in calculating the self-energy matrices used at the contacts which change the reflections at the boundaries, which points towards a search for more stable TB parameters. In addition to the above, traps and interfacial defects at the TMD/insulator junction should also be modeled to estimate any degradation to current levels because of it.

Applications in Magnetism

In addition to the predicted one-dimensional spin channel in chapter 5, other spintronics application can be pursued in few-layer devices by either artificially inducing electric fields (either in-plane or out of plane) or exploiting the inherent large spin-orbit splitting in the valence band of the TMDs. The valence band of the TMDs is at the Γ point for few-layer TMDs as it becomes an indirect band gap material with a widened spin-splitting because of inter-layer coupling. This could lead to interesting spin applications if we move away from electrons as the majority spin carrier. The breaking of inversion symmetry along with spin-orbit coupling in TMDs could lead to coupling of valley and spin physics resulting in long spin lifetimes [123]. We hope that this discussion on the electronic structure of few-layer devices and their potential applications in both conventional transistor design and novel devices could help spur further work in this very interesting material system which has become one of the defining materials of the 21$^{\text{st}}$ century.

Bibliography

[1] LS Adam et al. “A physical model for implanted nitrogen diffusion and its effect on oxide growth”. In: *Electron Devices Meeting, 2000. IEDM’00. Technical Digest. International.* IEEE. 2000, pp. 507–510.

[2] Samarth Agarwal, Gerhard Klimeck, and Mathieu Luisier. “Leakage-reduction design concepts for low-power vertical tunneling field-effect transistors”. In: *IEEE Electron Device Letters* 31.6 (2010), pp. 621–623.

[3] Sapan Agarwal and Eli Yablonovitch. “Using dimensionality to achieve a sharp tunneling FET (TFET) turn-on”. In: *Device Research Conference (DRC), 2011 69th Annual.* IEEE. 2011, pp. 199–200.

[4] MP Anantram, Mark S Lundstrom, and Dmitri E Nikonov. “Modeling of nanoscale devices”. In: *Proceedings of the IEEE* 96.9 (2008), pp. 1511–1550.

[5] J. Appenzeller et al. “Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors”. In: *Phys. Rev. Lett.* 93 (19 2004), p. 196805. DOI: `10.1103/PhysRevLett.93.196805`. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.93.196805`.

[6] Luigi Atzori, Antonio Iera, and Giacomo Morabito. “The internet of things: A survey”. In: *Computer networks* 54.15 (2010), pp. 2787–2805.

[7] Uygar E Avci and Ian A Young. “Heterojunction TFET scaling and resonant-TFET for steep subthreshold slope at sub-9nm gate-length”. In: *Electron Devices Meeting (IEDM), 2013 IEEE International.* IEEE. 2013, pp. 4–3.

[8] Uygar E Avci et al. “Understanding the feasibility of scaled III–V TFET for logic by bridging atomistic simulations and experimental results”. In: *VLSI technology (VLSIT), 2012 symposium on.* IEEE. 2012, pp. 183–184.

[9] Phaedon Avouris et al. “Carbon nanotube electronics”. In: *Proceedings of the IEEE* 91.11 (2003), pp. 1772–1784.

[10] Giorgio Baccarani, Matthew R Wordeman, and Robert H Dennard. “Generalized scaling theory and its application to a 1/4 micrometer MOSFET design”. In: *IEEE Transactions on Electron Devices* 31.4 (1984), pp. 452–462.

[11] Behtash Behin-Aein et al. “Proposal for an all-spin logic device with built-in memory”. In: *Nature nanotechnology* 5.4 (2010), pp. 266–270.

[12] Simone Bertolazzi, Jacopo Brivio, and Andras Kis. “Stretching and breaking of ultrathin MoS2”. In: *ACS nano* 5.12 (2011), pp. 9703–9709.

[13] Kirill I Bolotin et al. “Ultrahigh electron mobility in suspended graphene”. In: *Solid State Communications* 146.9 (2008), pp. 351–355.

[14] Daniele Braga et al. “Quantitative determination of the band gap of WS2 with ambipolar ionic liquid-gated transistors”. In: *Nano letters* 12.10 (2012), pp. 5218–5223.

[15] DJ Chadi. “Spin-orbit splitting in crystalline and compositionally disordered semiconductors”. In: *Physical Review B* 16.2 (1977), p. 790.

[16] Bhim Chamlagain et al. “Mobility improvement and temperature dependence in MoSe2 field-effect transistors on parylene-C substrate”. In: *Acs Nano* 8.5 (2014), pp. 5079–5088.

[17] Jiwon Chang, Leonard F. Register, and Sanjay K. Banerjee. “Atomistic full-band simulations of monolayer MoS2 transistors”. In: *Applied Physics Letters* 103.22, 223509 (2013), pp. –. DOI: `http://dx.doi.org/10.1063/1.4837455`. URL: `http://scitation.aip.org/content/aip/journal/apl/103/22/10.1063/1.4837455`.

[18] Jiwon Chang, Leonard F Register, and Sanjay K Banerjee. “Atomistic full-band simulations of monolayer MoS2 transistors”. In: *Applied Physics Letters* 103.22 (2013), p. 223509.

[19] Claude Chappert, Albert Fert, and Frédéric Nguyen Van Dau. “The emergence of spin electronics in data storage”. In: *Nature materials* 6.11 (2007), pp. 813–823.

[20] Tawinan Cheiwchanchamnangij and Walter RL Lambrecht. “Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS 2”. In: *Physical Review B* 85.20 (2012), p. 205302.

[21] Ming-Hui Chiu et al. “Determination of band alignment in transition metal dichalcogenides heterojunctions”. In: *arXiv preprint arXiv:1406.5137* (2014).

[22] Yang-Kyu Choi et al. “FinFET process refinements for improved mobility and gate work function engineering”. In: *Electron Devices Meeting, 2002. IEDM'02. International.* IEEE. 2002, pp. 259–262.

[23] Hsun-Jen Chuang et al. “High mobility WSe2 p-and n-type field-effect transistors contacted by highly doped graphene for low-resistance contacts”. In: *Nano letters* 14.6 (2014), pp. 3594–3601.

[24] Hsun-Jen Chuang et al. “Low-resistance 2D/2D ohmic contacts: A universal approach to high-performance WSe2, MoS2, and MoSe2 transistors”. In: *Nano letters* 16.3 (2016), pp. 1896–1902.

[25] R Coehoorn et al. “Electronic structure of MoSe 2, MoS 2, and WSe 2. I. Band-structure calculations and photoelectron spectroscopy”. In: *Physical Review B* 35.12 (1987), p. 6195.

[26] Saptarshi Das and Joerg Appenzeller. "WSe2 field effect transistors with enhanced ambipolar characteristics". In: *Applied physics letters* 103.10 (2013), p. 103501.

[27] S Dasgupta et al. "Sub-kT/q switching in strong inversion in PbZr 0.52 Ti 0.48 O 3 gated negative capacitance FETs". In: *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits* 1 (2015), pp. 43–48.

[28] Supriyo Datta. "Nanoscale device modeling: the Greens function method". In: *Superlattices and microstructures* 28.4 (2000), pp. 253–278.

[29] Supriyo Datta. *Quantum transport: atom to transistor*. Cambridge University Press, 2005.

[30] Bruce Doris et al. "Extreme scaling with ultra-thin Si channel MOSFETs". In: *Electron Devices Meeting, 2002. IEDM'02. International*. IEEE. 2002, pp. 267–270.

[31] Jason K Ellis, Melissa J Lucero, and Gustavo E Scuseria. "The indirect to direct band gap transition in multilayered MoS2 as predicted by screened hybrid density functional theory". In: *Applied Physics Letters* 99.26 (2011), p. 261908.

[32] Chris D English et al. "Improved contacts to MoS2 transistors by ultra-high vacuum metal deposition". In: *Nano letters* 16.6 (2016), pp. 3824–3830.

[33] Leo Esaki and Yuriko Miyahara. "A new device using the tunneling process in narrow pn junctions". In: *Solid-State Electronics* 1.1 (1960), 13IN515IN717–14IN616IN821.

[34] D Esseni et al. "Low field mobility of ultra-thin SOI N-and P-MOSFETs: Measurements and implications on the performance of ultra-short MOSFETs". In: *Electron Devices Meeting, 2000. IEDM'00. Technical Digest. International*. IEEE. 2000, pp. 671–674.

[35] Paul Martin Fahey, PB Griffin, and JD Plummer. "Point defects and dopant diffusion in silicon". In: *Reviews of modern physics* 61.2 (1989), p. 289.

[36] Hui Fang et al. "Degenerate n-doping of few-layer transition metal dichalcogenides by potassium". In: *Nano letters* 13.5 (2013), pp. 1991–1995.

[37] Hui Fang et al. "Strong interlayer coupling in van der Waals heterostructures built from single-layer chalcogenides". In: *Proceedings of the National Academy of Sciences* 111.17 (2014), pp. 6198–6202.

[38] Kartik Ganapathi, Youngki Yoon, and Sayeef Salahuddin. "Analysis of InAs vertical and lateral band-to-band tunneling transistors: Leveraging vertical tunneling for improved performance". In: *Applied Physics Letters* 97.3 (2010), p. 033504.

[39] Yongji Gong et al. "Vertical and in-plane heterostructures from WS2/MoS2 monolayers". In: *Nature materials* 13.12 (2014), pp. 1135–1142.

[40] E Gourmelon et al. "MS 2 (M= W, Mo) photosensitive thin films for solar cells". In: *Solar energy materials and solar cells* 46.2 (1997), pp. 115–121.

[41] Jing Guo et al. "Atomistic simulation of carbon nanotube field-effect transistors using non-equilibrium Greens function formalism". In: *Journal of Computational Electronics* 3.3 (2004), pp. 373–377.

[42] Suyog Gupta et al. "7-nm FinFET CMOS design enabled by stress engineering using Si, Ge, and Sn". In: *IEEE Transactions on Electron Devices* 61.5 (2014), pp. 1222–1230.

[43] R Haydock, Volker Heine, and MJ Kelly. "Electronic structure based on the local atomic environment for tight-binding bands". In: *Journal of Physics C: Solid State Physics* 5.20 (1972), p. 2845.

[44] Qiyuan He et al. "Fabrication of Flexible MoS2 Thin-Film Transistor Arrays for Practical Gas-Sensing Applications". In: *Small* 8.19 (2012), pp. 2994–2999.

[45] Stephan Henzler. *Power management of digital circuits in deep sub-micron CMOS technologies*. Vol. 25. Springer Science & Business Media, 2006.

[46] U Hetmaniuk, Y Zhao, and MP Anantram. "A nested dissection approach to modeling transport in nanodevices: Algorithms and applications". In: *International Journal for Numerical Methods in Engineering* 95.7 (2013), pp. 587–607.

[47] Adrian M Ionescu and Heike Riel. "Tunnel field-effect transistors as energy-efficient electronic switches". In: *Nature* 479.7373 (2011), pp. 329–337.

[48] Asif Islam Khan et al. "Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures". In: *Applied Physics Letters* 99.11 (2011), p. 113501.

[49] Carlo Jacoboni and Paolo Lugli. *The Monte Carlo method for semiconductor device simulation*. Springer Science & Business Media, 2012.

[50] Sanghyun Jo et al. "Mono-and bilayer WS2 light-emitting transistors". In: *Nano letters* 14.4 (2014), pp. 2019–2025.

[51] EO Kane. "Zener tunneling in semiconductors". In: *Journal of Physics and Chemistry of Solids* 12.2 (1960), pp. 181–188.

[52] Leonid V Keldysh et al. "Diagram technique for nonequilibrium processes". In: *Sov. Phys. JETP* 20.4 (1965), pp. 1018–1026.

[53] Hafiz MW Khalil et al. "Highly stable and tunable chemical doping of multilayer WS2 field effect transistor: Reduction in contact resistance". In: *ACS applied materials & interfaces* 7.42 (2015), pp. 23589–23596.

[54] M Abul Khayer and Roger K Lake. "Effects of band-tails on the subthreshold characteristics of nanowire band-to-band tunneling transistors". In: *Journal of Applied Physics* 110.7 (2011), p. 074508.

[55] Jack S Kilby. "Invention of the integrated circuit". In: *IEEE Transactions on electron devices* 23.7 (1976), pp. 648–654.

[56] Raseong Kim, Supriyo Datta, and Mark S Lundstrom. “Influence of dimensionality on thermoelectric device performance”. In: *Journal of Applied Physics* 105.3 (2009), p. 034506.

[57] Gerhard Klimeck. “Quantum and semi-classical transport in NEMO 1-D”. In: *Journal of Computational Electronics* 2.2 (2003), pp. 177–182.

[58] Andor Kormányos et al. “Spin-orbit coupling, quantum dots, and qubits in monolayer transition metal dichalcogenides”. In: *Physical Review X* 4.1 (2014), p. 011034.

[59] Georg Kresse and Jürgen Furthmüller. “Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set”. In: *Physical review B* 54.16 (1996), p. 11169.

[60] Roger Lake, Dejan Jovanovic, and CRISTIAN Rivas. “Nonequilibrium Greens functions in semiconductor device modeling”. In: *Progress in Nonequilibrium Green Functions* (2003), pp. 143–158.

[61] Kai-Tak Lam, Zhipeng Dong, and Jing Guo. “Performance limits projection of black phosphorous field-effect transistors”. In: *IEEE Electron Device Letters* 35.9 (2014), pp. 963–965.

[62] Stefano Larentis, Babak Fallahazad, and Emanuel Tutuc. “Field-effect transistors and intrinsic mobility in ultra-thin MoSe2 layers”. In: *Applied Physics Letters* 101.22 (2012), p. 223104.

[63] Mark E Law. “Process modeling for future technologies”. In: *IBM Journal of Research and Development* 46.2.3 (2002), pp. 339–346.

[64] Gwan-Hyoung Lee et al. “Highly stable, dual-gated MoS2 transistors encapsulated by hexagonal boron nitride with gate-controllable contact, resistance, and threshold voltage”. In: *ACS nano* 9.7 (2015), pp. 7019–7026.

[65] Hee Sung Lee et al. “MoS2 nanosheet phototransistors with thickness-modulated optical energy gap”. In: *Nano letters* 12.7 (2012), pp. 3695–3700.

[66] MH Lee et al. “Prospects for ferroelectric HfZrOx FETs with experimentally CET= 0.98 nm, SSfor= 42mV/dec, SSrev= 28mV/dec, switch-off¡ 0.2 V, and hysteresis-free strategies”. In: *Electron Devices Meeting (IEDM), 2015 IEEE International*. IEEE. 2015, pp. 22–5.

[67] Jing Li et al. “Design paradigm for robust spin-torque transfer magnetic RAM (STT MRAM) from circuit/architecture perspective”. In: *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* 18.12 (2010), pp. 1710–1723.

[68] Xiaolin Li et al. “Chemically derived, ultrasmooth graphene nanoribbon semiconductors”. In: *science* 319.5867 (2008), pp. 1229–1232.

[69] Gui-Bin Liu et al. “Electronic structures and theoretical modelling of two-dimensional group-VIB transition metal dichalcogenides”. In: *Chemical Society Reviews* 44.9 (2015), pp. 2643–2663.

[70] Wei Liu et al. “Role of metal contacts in designing high-performance monolayer n-type WSe2 field effect transistors”. In: *Nano letters* 13.5 (2013), pp. 1983–1990.

[71] Mathieu Luisier et al. “First-principles simulations of 2-D semiconductor devices: Mobility, IV characteristics, and contact resistance”. In: *Electron Devices Meeting (IEDM), 2016 IEEE International*. IEEE. 2016, pp. 5–4.

[72] Mathieu Luisier et al. “Ultimate device scaling: Intrinsic performance comparisons of carbon-based, InGaAs, and Si field-effect transistors for 5 nm gate length”. In: *Electron Devices Meeting (IEDM), 2011 IEEE International*. IEEE. 2011, pp. 11–2.

[73] Jesse Maassen et al. “Full band calculations of the intrinsic lower limit of contact resistivity”. In: *Applied Physics Letters* 102.11 (2013), p. 111605.

[74] Kin Fai Mak et al. “Atomically thin MoS 2: a new direct-gap semiconductor”. In: *Physical Review Letters* 105.13 (2010), p. 136805.

[75] L. F. Mattheiss. “Band Structures of Transition-Metal-Dichalcogenide Layer Compounds”. In: *Phys. Rev. B* 8 (8 1973), pp. 3719–3740. DOI: `10.1103/PhysRevB.8.3719`. URL: `http://link.aps.org/doi/10.1103/PhysRevB.8.3719`.

[76] Varun Mishra et al. “Dependence of intrinsic performance of transition metal dichalcogenide transistors on materials and number of layers at the 5 nm channel-length limit”. In: *IEEE IEDM Tech. Dig* (2013), pp. 136–139.

[77] Varun Mishra et al. “Screening in ultrashort (5 nm) channel MoS 2 transistors: A full-band quantum transport study”. In: *IEEE Transactions on Electron Devices* 62.8 (2015), pp. 2457–2463.

[78] Kaizad Mistry et al. “A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging”. In: *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*. IEEE. 2007, pp. 247–250.

[79] Alejandro Molina-Sanchez and Ludger Wirtz. “Phonons in single-layer and few-layer MoS 2 and WS 2”. In: *Physical Review B* 84.15 (2011), p. 155413.

[80] Gordon E Moore et al. “Lithography and the future of Moores law”. In: *Proc. Spie*. Vol. 2437. 8. 1995, pp. 2–17.

[81] Sivakumar Mudanai et al. “Capacitance compact model for ultrathin low-electron-effective-mass materials”. In: *IEEE Transactions on Electron Devices* 58.12 (2011), pp. 4204–4211.

[82] Laurence Nagel and Ronald Rohrer. “Computer analysis of nonlinear circuits, excluding radiation (CANCER)”. In: *IEEE Journal of Solid-State Circuits* 6.4 (1971), pp. 166–182.

[83] Laurence William Nagel and Donald O Pederson. *SPICE (simulation program with integrated circuit emphasis)*. Tech. rep. 1973.

[84] Nanoacademic Technologies. *Nanoskif*. `http://www.nanoacademic.com/`. 2013.

[85] Branislav K Nikolić et al. “Nonequilibrium spin Hall accumulation in ballistic semiconductor nanostructures”. In: *Physical review letters* 95.4 (2005), p. 046601.

[86] Dmitri E Nikonov, George Bourianoff, Tahir Ghani, et al. “Proposal of a spin torque majority gate logic”. In: *Electron Device Letters, IEEE* 32.8 (2011), pp. 1128–1130.

[87] Dmitri E Nikonov, George I Bourianoff, and Paolo A Gargini. “Power dissipation in spintronic devices out of thermodynamic equilibrium”. In: *Journal of superconductivity and novel magnetism* 19.6 (2006), pp. 497–513.

[88] KS Novoselov. “Nobel lecture: Graphene: Materials in the flatland”. In: *Reviews of Modern Physics* 83.3 (2011), p. 837.

[89] Dmitry Ovchinnikov et al. “Electrical transport properties of single-layer WS2”. In: *ACS nano* 8.8 (2014), pp. 8174–8181.

[90] Tomás Palacios. “Graphene electronics: thinking outside the silicon box”. In: *Nature nanotechnology* 6.8 (2011), pp. 464–465.

[91] John P Perdew, Kieron Burke, and Matthias Ernzerhof. “Generalized gradient approximation made simple”. In: *Physical review letters* 77.18 (1996), p. 3865.

[92] Xiaofeng Qian et al. “Quantum spin Hall effect in two-dimensional transition metal dichalcogenides”. In: *Science* (2014). ISSN: 0036-8075. DOI: `10.1126/science.1256815`. eprint: `http://science.sciencemag.org/content/early/2014/11/19/science.1256815.full.pdf`. URL: `http://science.sciencemag.org/content/early/2014/11/19/science.1256815`.

[93] Branimir Radisavljevic et al. “Single-layer MoS2 transistors”. In: *Nature nanotechnology* 6.3 (2011), pp. 147–150.

[94] Filip A Rasmussen and Kristian S Thygesen. “Computational 2D materials database: electronic structure of transition-metal dichalcogenides and oxides”. In: *The Journal of Physical Chemistry C* 119.23 (2015), pp. 13169–13183.

[95] Zhibin Ren. “Nanoscale MOSFETs: Physics, simulation and design”. PhD thesis. Purdue University, 2001.

[96] Tania Roy et al. “Dual-gated MoS2/WSe2 van der Waals tunnel diodes and transistors”. In: *Acs Nano* 9.2 (2015), pp. 2071–2079.

[97] Sayeef Salahuddin and Supriyo Datta. “Interacting systems for self-correcting low power switching”. In: *Applied Physics Letters* 90.9 (2007), p. 3503.

[98] Sayeef Salahuddin and Supriyo Datta. “Use of negative capacitance to provide voltage amplification for low power nanoscale devices”. In: *Nano letters* 8.2 (2008), pp. 405–410.

[99] MP Lopez Sancho et al. “Highly convergent schemes for the calculation of bulk and surface Green functions”. In: *Journal of Physics F: Metal Physics* 15.4 (1985), p. 851.

[100] Ankit Sharma and Kaushik Roy. "Design Space Exploration of Hysteresis-Free HfZrO x-Based Negative Capacitance FETs". In: *IEEE Electron Device Letters* 38.8 (2017), pp. 1165–1167.

[101] William Shockley. "Electrons and holes in semiconductors". In: *&* 471 (1952), p. 237.

[102] N Singh et al. "Ultra-narrow silicon nanowire gate-all-around CMOS devices: Impact of diameter, channel-orientation and low temperature on device performance". In: *Electron Devices Meeting, 2006. IEDM'06. International*. IEEE. 2006, pp. 1–4.

[103] Thomas Skotnicki et al. "The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance". In: *IEEE Circuits and Devices Magazine* 21.1 (2005), pp. 16–26.

[104] John C Slater and George F Koster. "Simplified LCAO method for the periodic potential problem". In: *Physical Review* 94.6 (1954), p. 1498.

[105] Andrea Splendiani et al. "Emerging photoluminescence in monolayer MoS2". In: *Nano letters* 10.4 (2010), pp. 1271–1275.

[106] Mark A Stettler. "Device and process modeling: 20 Years at Intel's other fab". In: *Simulation of Semiconductor Processes and Devices (SISPAD), 2014 International Conference on*. IEEE. 2014, pp. 13–16.

[107] Kunihiro Suzuki et al. "Source/drain contact resistance of silicided thin-film SOI MOSFET's". In: *IEEE Transactions on Electron Devices* 41.6 (1994), pp. 1007–1012.

[108] A. Szabo, R. Rhyner, and M. Luisier. "Ab-initio simulations of MoS2 transistors: From mobility calculation to device performance evaluation". In: *Electron Devices Meeting (IEDM), 2014 IEEE International*. 2014, pp. 30.4.1–30.4.4. DOI: `10.1109/IEDM.2014.7047142`.

[109] Simon M Sze and Kwok K Ng. *Physics of semiconductor devices*. John wiley & sons, 2006.

[110] Yuan Taur and Tak H Ning. *Fundamentals of modern VLSI devices*. Cambridge university press, 2009.

[111] Yuan Taur, Tak H Ning, et al. *Fundamentals of modern VLSI devices*. Vol. 2. Cambridge University Press Cambridge, 1998.

[112] Shiyang Tian et al. "A detailed physical model for ion implant induced damage in silicon". In: *IEEE Transactions on Electron Devices* 45.6 (1998), pp. 1226–1238.

[113] R Venugopal et al. "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches". In: *Journal of Applied Physics* 92.7 (2002), pp. 3730–3739.

[114] Suresh Vishwanath et al. "Controllable growth of layered selenide and telluride heterostructures and superlattices using molecular beam epitaxy". In: *Journal of Materials Research* 31.7 (2016), pp. 900–910.

[115] áP Vogl, Harold P Hjalmarson, and John D Dow. "A semi-empirical tight-binding theory of the electronic structure of semiconductors". In: *Journal of Physics and Chemistry of Solids* 44.5 (1983), pp. 365–378.

[116] Caihua Wan et al. "Geometrical enhancement of low-field magnetoresistance in silicon". In: *Nature* 477.7364 (2011), pp. 304–307.

[117] Jingli Wang et al. "High Mobility MoS2 Transistor with Low Schottky Barrier Contact by Using Atomic Thick h-BN as a Tunneling Layer". In: *Advanced Materials* 28.37 (2016), pp. 8302–8308. ISSN: 1521-4095. DOI: `10.1002/adma.201602757`. URL: `http://dx.doi.org/10.1002/adma.201602757`.

[118] L Wang and MW Wu. "Electron spin diffusion in monolayer MoS 2". In: *Physical Review B* 89.20 (2014), p. 205401.

[119] Qing Hua Wang et al. "Electronics and optoelectronics of two-dimensional transition metal dichalcogenides". In: *Nature nanotechnology* 7.11 (2012), pp. 699–712.

[120] Zhe Wang et al. "Origin and Magnitude of 'Designer' Spin-Orbit Interaction in Graphene on Semiconducting Transition Metal Dichalcogenides". In: *Phys. Rev. X* 6 (4 2016), p. 041020. DOI: `10.1103/PhysRevX.6.041020`. URL: `https://link.aps.org/doi/10.1103/PhysRevX.6.041020`.

[121] Cory E Weber et al. "Modeling of NMOS performance gains from edge dislocation stress". In: *Electron Devices Meeting (IEDM), 2011 IEEE International*. IEEE. 2011, pp. 34–4.

[122] SA Wolf et al. "Spintronics: a spin-based electronics vision for the future". In: *Science* 294.5546 (2001), pp. 1488–1495.

[123] Di Xiao et al. "Coupled spin and valley physics in monolayers of MoS 2 and other group-VI dichalcogenides". In: *Physical Review Letters* 108.19 (2012), p. 196802.

[124] Lingming Yang et al. "Chloride molecular doping technique on 2D materials: WS2 and MoS2". In: *Nano letters* 14.11 (2014), pp. 6275–6280.

[125] Youngki Yoon, Kartik Ganapathi, and Sayeef Salahuddin. "How good can monolayer MoS2 transistors be?" In: *Nano letters* 11.9 (2011), pp. 3768–3773.

[126] Bin Yu et al. "FinFET scaling to 10 nm gate length". In: *Electron Devices Meeting, 2002. IEDM'02. International*. IEEE. 2002, pp. 251–254.

[127] Ferdows Zahid et al. "A generic tight-binding model for monolayer, bilayer and bulk MoS2". In: *arXiv preprint arXiv:1304.0074* (2013).

[128] Ping Zhou et al. "Energy reduction for STT-RAM using early write termination". In: *Computer-Aided Design-Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference on*. IEEE. 2009, pp. 264–268.

[129] Jian-Gang (Jimmy) Zhu and Chando Park. "Magnetic tunnel junctions". In: *Materials Today* 9.11 (2006), pp. 36 –45. ISSN: 1369-7021. DOI: `https://doi.org/10.1016/S1369-7021(06)71693-5`. URL: `http://www.sciencedirect.com/science/article/pii/S1369702106716935`.