Improving the Performance of Nanoscale Field-Effect Transistors Through Electrostatic Engineering

by

Robert Kevin Arran Bennett

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering – Nanotechnology

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Acknowledgements

I would like to begin by thanking my supervisor, Professor Yoon, for welcoming me into his research group and for offering me his guidance and patience over the last two years. Thank you for all the help you have given me in both my research and in my academic career.

I would also like to thank all of my fellow group members: AbdulAziz, Andrew, Demin, Hyunjae, Manasa, Mayuri, Richard, and Yiju. I should give special thanks to Demin in particular for teaching me the fundamentals of our NEGF and DFT simulations (and for answering my endless questions) and to AbdulAziz for developing the simulation software that made my research possible. I am also grateful to the Waterloo Institute for Nanotechnology (WIN), the Government of Ontario, and the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support.

I've been fortunate enough to have many teachers and mentors who have encouraged me to pursue a career in science and engineering. Thanks go to Sean Ealey, Mitchell Plant, Ross Holt, Dapeng Liu, Collin Tittle, Gordon Hall, William Wang, Yuxing Yao, Professor Wang, and Professor Backhouse for helping me get to where I am today.

I am grateful to my family for always supporting my research career, which may have started I was about four years old and I stuck a twig in a beehive because "I wanted to know what would happen." Thank you for helping me preserve my curiosity while improving my methodology.

And of course, my thanks go out to Sarah for her love and support over the years and for giving me a reason why.

Abstract

The continued scaling of field-effect transistors (FETs) requires that nearly every aspect of these devices be optimized to ensure that they can continue to meet practical performance requirements. However, scaling the channel lengths of FETs naturally enhances electrostatic and quantum mechanical short-channel effects, thereby increasing leakage currents in the OFF-state, reducing driving currents in the ON-state, and making it difficult for FETs attain optimal switching behaviours. To mitigate these detrimental effects, it is imperative to (i) thoroughly understand the electrostatic operation of nanoscale FETs and (ii) establish novel design strategies to mitigate short-channel effects.

In this thesis, I address these two challenges by studying the electrostatic operation of nanoscale FETs using simulation techniques. In particular, I use the non-equilibrium Green's function method, an atomistic quantum transport simulation technique, to study the electrostatic operation of MOSFETs and to assess the utility of novel electrostatic design strategies for nanoscale FETs.

The body of this thesis is divided into three main works. In the first, I study how individual elements of a metal-oxide-semiconductor FET's (MOSFET's) semiconductor's anisotropic permittivity affect device performance, and I establish electrostatic-based guidelines for selecting optimal semiconductors for future MOSFETs. Next, I study how replacing an FET's conventional isotropic insulators (i.e. gate insulator and spacers) with anisotropic insulators can improve the performance of both conventional MOSFETs and tunnel FETs, and I propose novel insulator architectures to further optimize the performance of these devices. Finally, in my third study, I examine how fringe-induced barrier lowering, an electrostatic short-channel effect created by implementing high-κ gate insulators, can be exploited to suppress quantum mechanical short-channel effects (source-to-drain tunneling) to improve the overall performance of nanoscale MOSFETs. The operating principles and design rules established in these three works extend the current picture of the electrostatic operation and design rules for nanoscale FETs to help device designers continue to scale FETs while meeting essential performance benchmarks.

Contents

List of Figures

List of Tables

S: Look! [*Points to the volcano.*] Can your fortune telling explain that?

Cuts to shot of the volcano, smoke rising from the top in great plumes.

M: [*Amused.*] Can your science explain why it rains?

S: [*Turns to him; yells angrily.*] Yes! Yes, it can!

1 Introduction

1.1 Scaling of Field-Effect Transistors

The advancement of modern computational technologies has been driven largely by the continued miniaturization, or *scaling*, of the metal-oxide-semiconductor field-effect transistor (MOSFET). The motivation behind miniturization is straightforward: the smaller MOSFETs are, the more we can put on a single computer chip. Over the last 60 years, the lengths of MOSFETs have decreased exponentially, with modern computer processors achieving billions of transistors on a single $1\ \text{cm}^2$ chip. However, MOSFETs that comprise modern computer chips have channel lengths of around 10 nm, and scaling devices beyond these lengths poses immense challenges, both in terms of device design and fabrication. Consequently, the scaling of silicon-based MOSFETs over the last decade has been achieved mostly by decreasing the effective footprint of their interconnects rather than by continuing to decrease their physical channel lengths.

Despite these challenges, the continued scaling of FETs' channels remains a topic of great interest in both academia and industry. FETs need to be extremely well optimized from nearly every perspective to scale them into the sub-10 nm regime, requiring researchers to derive models that capture the unique behaviors of nanoscale FETs (which often different significantly compared to the behaviors of their μm scale counterparts) and to engineer novel design strategies to optimize sub-10 nm FETs. In particular, the electrostatic integrity of devices tends to become compromised as devices are scaled to short channel lengths, making the electrostatic engineering of nanoscale FETs an essential step in realizing the sub-10 nm regime.

Therefore, the goals of the work presented in this thesis are (i) to provide a deeper understanding of the electrostatic operation of nanoscale FETs and (ii) to present novel design strategies for enhancing nanoscale FETs' electrostatic integrity. This thesis is organized as follows: firstly, I discuss why electrostatic integrity is essential for any high-performing FET and explain why scaling devices to extremely short channel lengths compromises their electrostatic integrity. Next, I review previous literature to present the current state-of-the-art in electrostatic engineering to contextualize the work of this thesis. Finally, I describe the methodology I used to explore the electrostatic operation of nanoscale FETs before presenting three new advances in electrostatic engineering in the body of this thesis.

1.2 Electrostatic and Quantum Mechanical Short-Channel Effects

The characteristic behaviours of MOSFETs can be understood using a capacitance network where the source, drain, and gate electrodes are coupled to a point in MOSFET's channel through three capacitors: $C_S^{(\|)}$, $C_D^{(\|)}$,

and $C_G^{(\perp)}$. Here, I use a superscripted parallel ($\parallel$) or perpendicular ($\perp$) sign to denote that the the capacitor acts in the in-plane or out-of-plane direction, respectively. $C_S^{(\parallel)}$ and $C_D^{(\parallel)}$ couple the channel to the source and drain, respectively, and arise due to lateral charge separation within an FET's semiconductor. The gate capacitance, $C_G^{(\perp)}$, couples the channel to the gate electrode through the semiconductor and insulator. The capacitances offered by the semiconductor and insulator in the out-of-plane direction are denoted by $C_{SC}^{(\perp)}$ and $C_{\text{ox}}^{(\perp)}$, respectively, and the overall $C_G^{(\perp)}$ is the series capacitance of $C_{\text{ox}}^{(\perp)}$ and $C_{SC}^{(\perp)}$. The locations of these capacitors in an FET are shown schematically in Fig. 1.1(a), where V_S, V_D, V_G, and ψ_{ch} are the potentials of the source, drain, gate, and a point in the channel, respectively. While we could assign ψ_{ch} to be the potential at any point in the channel, it is most useful to define ψ_{ch} at the point in which the potential in the channel is greatest. Consequently, we assign ψ_{ch} to be located in the center of the channel so that it is identical to the height of the potential barrier, as illustrated schematically in Fig. 1.1(b).

We can already begin to understand the operating principles of a MOSFET using this simplified model. If we increase V_G by some amount ΔV_G then ψ_{ch} will likewise change by some amount $\Delta\,\psi_{ch}$, thereby inducing a change in the drain current, I_D. Ideally, a MOSFET's I_D should be as responsive to changes in V_G as possible. In other words, we generally wish to maximize $\partial I_D/\partial V_G$, as this allows us to maximize the strength of our signal, which is practical for several reasons: it allows us to maximize the frequency at which we can switch MOSFETs off and on [2], it minimizes MOSFETs' power consumptions [3], and it boosts the sensitivity of MOSFET-based sensors [4].

In the most ideal case, the potential of the channel will respond perfectly to a change in the gate electrode, i.e. $\Delta\psi_{ch} = \Delta V_G$. However, from the top-of-the-barrier capacitor model shown in Fig. 1.1, we may note that

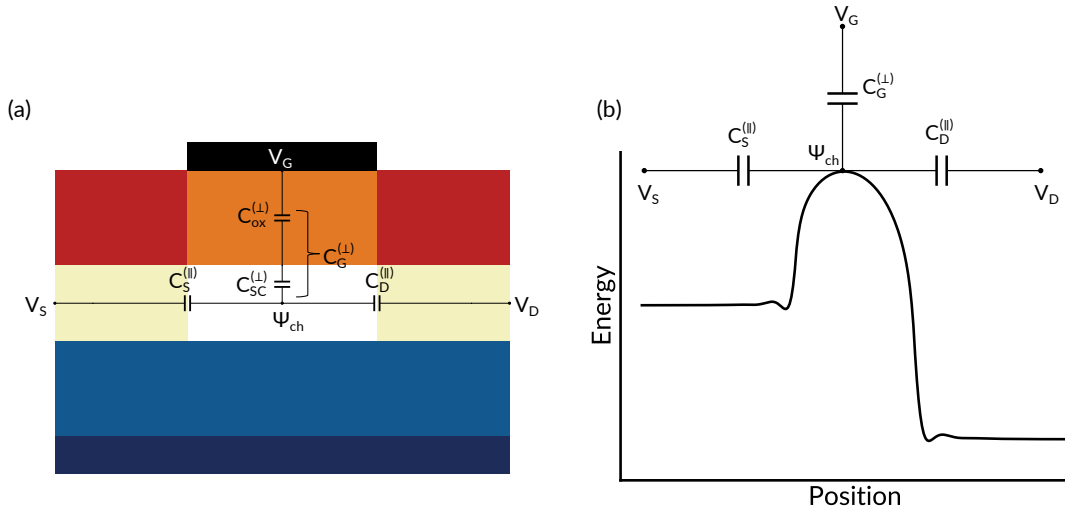

Figure 1.1: **(a)** A diagram showing how key potentials in an FET are coupled through various capacitors. Adapted from [1], © 2020 IEEE. **(b)** A top-of-the-barrier model where the conduction band edge is qualitatively plotted as a function of the lateral position in the device. Here, the capacitors from (a) are used to illustrate that the top of the barrier is taken as ψ_{ch}, which is in turn coupled to the source, drain, and gate electrodes. Adapted from [2].

V_S and V_D also influence ψ_{ch}. By applying a capacitive voltage divider, we can estimate the influence of each electrode on ψ_{ch}:

$$\psi_{ch}(C_S^{(\parallel)} + C_D^{(\parallel)} + C_G^{(\perp)}) = C_S^{(\parallel)}V_S + C_D^{(\parallel)}V_D + V_G C_G^{(\perp)} \tag{1.1}$$

$$\psi_{ch} = \frac{C_S^{(\parallel)}V_S + C_D^{(\parallel)}V_D + V_G C_G^{(\perp)}}{C_S^{(\parallel)} + C_D^{(\parallel)} + C_G^{(\perp)}} \tag{1.2}$$

$$\frac{\partial \psi_{ch}}{\partial V_G} = \frac{C_G^{(\perp)}}{C_S^{(\parallel)} + C_D^{(\parallel)} + C_G^{(\perp)}} \tag{1.3}$$

$$\frac{\partial \psi_{ch}}{\partial V_S} = \frac{C_S^{(\parallel)}}{C_S^{(\parallel)} + C_D^{(\parallel)} + C_G^{(\perp)}} \tag{1.4}$$

$$\frac{\partial \psi_{ch}}{\partial V_D} = \frac{C_D^{(\parallel)}}{C_S^{(\parallel)} + C_D^{(\parallel)} + C_G^{(\perp)}} \tag{1.5}$$

From Equation (1.3), $\partial\psi_{ch}/\partial V_G$ achieves its maximum value of 1 (corresponding to the ideal case of $\Delta V_G = \Delta\psi_{ch}$) when $C_G^{(\perp)} >> (C_S^{(\parallel)} + C_D^{(\parallel)})$. In this case, $\partial\psi_{ch}/\partial V_S \approx 0$ and $\partial\psi_{ch}/\partial V_D \approx 0$ from Equation (1.4) and Equation (1.5), indicating that a change in the source/drain voltage will have little effect on the channel's potential. In this case, the gate has perfect control over the channel's potential, whereas the channel's potential is entirely unresponsive to changes in the potentials of the source and drain. Such a device is said to have good gate control. However, if $C_S^{(\parallel)}$ and/or $C_D^{(\parallel)}$ are on the same order (or greater than) $C_G^{(\perp)}$, then $\partial\psi_{ch}/\partial V_G$ will be considerably less than 1 and $\partial\psi_{ch}/\partial V_S$ and $\partial\psi_{ch}/\partial V_D$ will both be non-zero. Such a device is said to have weak gate control and will generally offer poor performance for practical applications. Consequently, device engineers typically seek to maximize $C_G^{(\perp)}$ while minimizing $C_S^{(\parallel)}$ and $C_D^{(\parallel)}$ to maximize an FET's performance. At the same time, however, we also wish to minimize the physical dimensions of MOSFETs to fit as many devices as possible on a single computer chip. These two design considerations conflict: since $C_S^{(\parallel)}$ and $C_D^{(\parallel)}$ are geometric capacitances, they are inversely proportional to the channel length L_{ch} and become larger as the length of MOSFETs are scaled down. In other words, the drain and source electrodes naturally exert greater influence on the channel's potential as the lengths of devices decreases, worsening device performance [2].

The influence of the source and drain's electric fields on the channel's potential gives rise to a number of *short-channel effects* that burden nanoscale FETs. These effects are pervasive among scaled devices and can cause FETs to have reduced ON currents, slower switching speeds, and in extreme cases can lead to leakage currents that prevent FETs from ever entering the OFF state. Therefore, it is extremely important for device engineers to take electrostatic integrity into account when designing nanoscale FETs to ensure that FETs remain able to meet key performance metrics even at the short channel lengths required for practical device applications.

A device's electrostatic integrity is often quantified by the scaling length λ. λ represents the distance that electric field lines penetrate laterally from a device's drain into its channel and is given by Equation (1.6) [5],

$$\lambda = \sqrt{\frac{\epsilon^{(SC)}}{\epsilon_{ox}} t_{ox} t_{ch}}, \tag{1.6}$$

where ϵ_{ox} is the permittivity of the MOSFET's gate insulator, $\epsilon^{(SC)}$ is the permittivity of the semiconductor, t_{ox} is the thickness of the gate insulator, and t_{ch} is the thickness of the semiconductor. A device with $C_G^{(\perp)}/C_D^{(\parallel)} \geq 10$ typically exhibits strong electrostatic control, whereas devices tend to suffer from short channel effects when $C_G^{(\perp)}/C_D^{(\parallel)} < 10$. It should be noted, however, that Equation (1.6) is only an estimate for λ, and that λ is difficult to quantify precisely, especially for short-channel devices. However, Equation (1.6) still offers a reasonable estimate for λ and is still regularly applied to describe the electrostatic integrity of modern FETs [6].

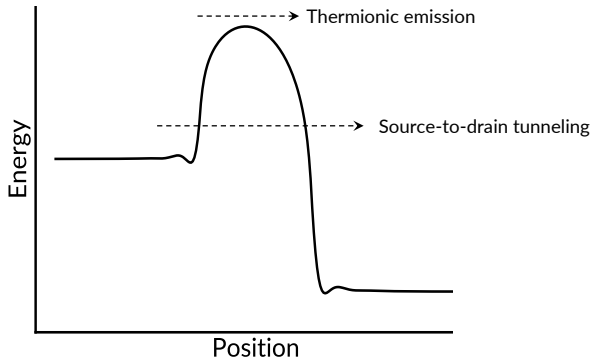

Figure 1.2: A schematic illustration showing the difference between thermionic emission (where high-energy electrons contribute to I_D by traveling over the potential barrier) and source-to-drain tunneling (where lower-energy electrons tunnel through the potential barrier).

Alongside electrostatic short-channel effects, quantum mechanical short-channel effects can severely degrade the performance of ultra-scaled FETs. In particular, MOSFETs with sub-10 nm channel lengths tend to suffer from source-to-drain tunneling, a quantum mechanical short-channel effect where electrons with energies below the height of the potential barrier tunnel through the barrier to contribute to the total drain current, as illustrated in Fig. 1.2. A MOSFET's source-to-drain tunneling current grows exponentially as the MOSFET's channel length decreases, and in severe cases these tunneling currents can prevent MOSFETs from entering the OFF-state, even at extremely negative gate biases [7].

1.3 Recent Advances in the Electrostatic Engineering of Field-Effect Transistors

1.3.1 Electrostatics of Semiconductors

Decreasing the physical thickness of a MOSFET's semiconductor is one of the most common approaches for increasing an FET's electrostatic integrity. Thick semiconductors exhibit relatively poor electrostatic control because the physical thickness of the semiconductor itself acts as a capacitor that weakens the coupling between the bottom of the semiconductor (i.e., the portion of the semiconductor opposite to the gate electrode). This behavior is reflected in Equation (1.3), as $C_G^{(\perp)}$ is the series capacitance of $C_{\text{ox}}^{(\perp)}$ and $C_{SC}^{(\perp)}$, where $C_{SC}^{(\perp)} \propto \frac{1}{t_{SC}}$. This behavior is also reflected in Equation (1.6), where λ decreases as t_{ch} decreases, thereby decreasing $C_G^{(\perp)}/C_D^{(\parallel)}$ and improving the device's electrostatic integrity. However, mobilities of electrons and holes of common semiconductors (e.g., silicon, gallium arsenide) decrease rapidly as the semiconductor's thickness approaches nanometer length scales, posing a practical limit for further decreasing t_{ch}. This limitation has one of the main reasons why commercial silicon-based MOSFETs have not been scaled to channel lengths less than 10 nm.

New possibilities in scaled devices emerged in 2010 when two-dimensional (2-D) MoS_2 was found to be a direct band gap semiconductor that exhibits reasonable electron and hole mobilities even at monolayer thicknesses [8]. This generated a surge of research interest within the field of two-dimensional semiconductors and has since lead to the discovery of over 100 2-D semiconductors have been studied for use in FETs in either a theoretical (e.g., using density functional theory coupled with quantum transport simulations) or experimental capacity [9]. While two-dimensional semiconductors still face significant challenges before they can be implemented in practical FETs, such as improving their ambient stability, improving their contact resistances, and doping them without introducing unacceptable levels of strain, they remain the most promising class of materials for enabling scaled FETs with strong electrostatic integrities.

Aside from engineering the thickness of semiconductors, several recent works have focused instead on the semiconductor's permittivity, which is another key parameter for determining the capacitances contained within an FET's semiconductor (labelled in Fig. 1.1). Generally speaking, the electrostatic coupling offered by the semiconductor tends to deteriorate device performance by increasing the drain's influence over the channel. This is also reflected in Equation (1.6), where λ increases as the semiconductor's permittivity increases. Recently, reports have emerged that suggest that due to a phenomenon termed *permittivity reduction,* the permittivity of some semiconductors may not be constant throughout the material. For example, polarized bonds in thin slabs silicon cause the material's local permittivity to decrease, leading to a non-uniform permittivity that is small at the surface and that increases towards the bulk value of silicon ($12\epsilon_0$) in the center of the slab. In 2019, Chen et al. [10] investigated how permittivity reduction could impact the performance of ultra-scaled MOSFETs by using non-equilibrium Green's function simulations to simulate the performance of MOSFETs with differing dielectric profiles. As a test case, they first investigated silicon-based MOSFETs where the dielectric constant of silicon was varied from 4 to 12. As shown in

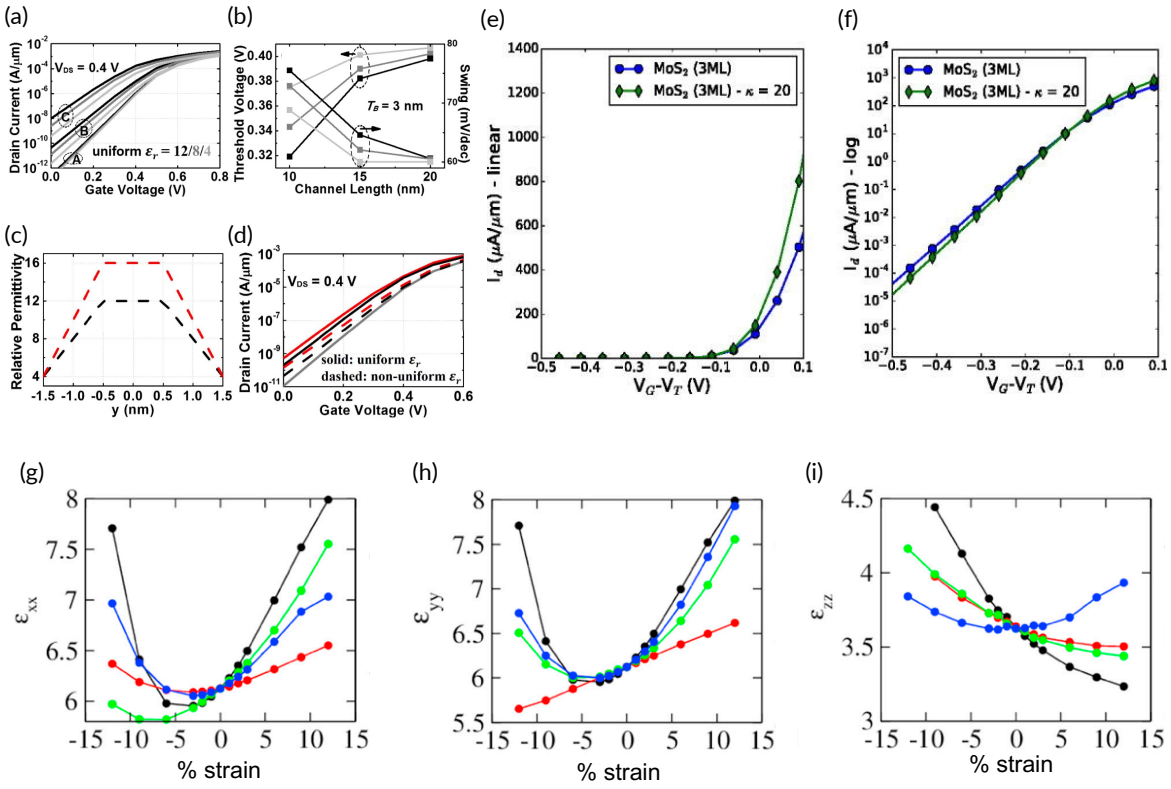

Figure 1.3: **(a)** Simulated transfer characteristics for MOSFETs where the semiconductor's dielectric constant was varied between 4 and 4, and whose channel lengths/semiconductor thicknesses are 20/3 nm (group A), 10/3 nm (group B), and 10/4 nm (group C). **(b)** Threshold voltage and subthreshold swing as a function of channel length for MOSFETs whose semiconductor's dielectric constant is varied from 4 to 12 and whose semiconductor's thickness is fixed at 3 nm. **(c)** Various semiconductor dielectric constant profiles used for subsequent simulations. **(d)** Simulated transfer characteristics for MOSFETs with the profiles shown in (c) (dashed red and black lines) and for MOSFETs with constant semiconductor dielectric constants of 16 (solid red line), 12 (solid black line) and 4 (solid gray line). **(e)** and logarithmic **(f)** transfer characteristics for 3-layer MoS_2-based MOSFETs with a base dielectric constant of 10 and an artificial dielectric constant of 20. In plane **(g,h)** and out-of-plane **(i)** permittivities of MoS_2 as a function of strain, where black, blue, red, and green lines correspond to isotropic, shear (xy), uniaxial (x), and uniaxial (y) strains, respectively.
(a – d) are taken from ref. [10] and are © 2019 IEEE.[a]
(e) and (f) are taken from ref. [11] and are © 2017 AIP publishing.[b]
(g – i) are taken from ref. [12] and are © 2016 Elsevier.[c]

[a] See copyright letter provided at the end of this thesis.
[b] Reprinted from Journal of Applied Physics, 121, A. Lu et al., On the electrostatic control achieved in transistors based on multilayered MoS2: A first-principles study, 044505, 2017, with the permission of AIP Publishing.
[c] Reprinted from Solid State Communications, 227, A. Maniadaki et al., Strain engineering of electronic properties of transition metal dichalcogenide monolayers, 33 – 39, 2016, Copyright 2016, with permission from Elsevier.

Fig. 1.3(a) and Fig. 1.3(b), the MOSFETs' OFF-currents and subthreshold swings were both minimized when the semicondutor's permittivity was set to small values, whereas the threshold voltage increased as the dielectric constant increased as a result of drain-induced barrier lowering. Next, they simulated the

performance of MOSFETs where the permittivity of silicon followed the functions shown in Fig. 1.3(c), where the y direction corresponds to the out-of-plane direction (i.e., perpendicular to the direction in which current is carried). As shown in Fig. 1.3(d), MOSFETs whose semiconductors followed these dielectric profiles had lower subthreshold swings and OFF-currents compared to MOSFETs whose semiconductors' permittivities were set to the peak values in the profiles (i.e. 16 for the solid red line, 12 for the solid black line), but could not outperform a semiconductor whose permittivity was set to the minimum value in the dielectric profile (i.e. 4 for the grey line).

Similarly, two-dimensional layered semiconductors, such as MoS_2 and black phosphorus, exhibit thickness-dependent permittivities where both their in-plane and out-of-plane permittivities decrease from their bulk value as the semiconductor's thickness decreases. This in turn affects source/channel and drain/channel couplings and modifies electrostatic shielding effects between subsequent layers. In 2017, Lu et al. demonstrated that electrostatic shielding between layers of two-dimensional semiconductors could greatly impact the performance of two-dimensional semiconductors by simulating the transfer characteristics of MoS_2-based MOSFETs with varying numbers of layers. Interestingly, they observed that the ON current increased and subthreshold swing decreased when the semiconductor's permittivity increased from ~10 to 20 [Figs. 1.3(e,f)], which is in disagreement with scaling trends for the semiconductor's permittivity predicted by Equation (1.6). However, the reason for this discrepancy was not commented on in the original publication, but will be investigated in greater detail in Chapter 3 of this thesis.

Although a material's permittivity is typically thought of as a fixed constant, in 2016 Maniadaki et al. [12] demonstrated that a semiconductor's in-plane and out-of-plane permittivities could be tuned by straining the material's atomic lattice. For example, the in-plane permittivity of MoS_2 could be tuned between ~5.5 and 8 by applying compressive/tensile strains [Figs. 1.3(g),(h)], whereas the out-of-plane permittivity of MoS_2 could be tuned between $\approx$ 3.2 and 4.5 [Fig. 1.3(i)]. As strain engineering is regularly employed to tune other essential electronic properties of semiconductors, such as mobility and band gap [13], the ability to simultaneously tune dielectric constants through strain engineering could allow promising semiconductors to be further optimized post-fabrication.

1.3.2 Electrostatics of Insulators

The most straightforward approach to enhance a MOSFET's electrostatic integrity is by increasing the capacitance offered by its gate insulator. As $C_{\text{ox}}^{(\perp)} \propto \epsilon_{ox}/t_{ox}$, this may be achieved by decreasing the insulator's physical thickness and/or by using insulators with higher permittivities, called *high-κ insulators*. To facilitate the comparisons of various gate oxides, device designers often discuss gate insulators in terms of their equivalent oxide thicknesses (EOTs), calculated as:

$$\text{EOT} = \frac{\epsilon^{(\text{SiO}_2)}}{\epsilon_{ox}} t_{ox} \tag{1.7}$$

where $\epsilon^{(\mathrm{SiO_2})}$ is the permittivity of silicon dioxide (3.9). The EOT quantifies the capacitance of an insulator by describing how thick a layer of silicon dioxide would have to be to provide an equivalent gate capacitance. For instance, a hypothetical insulator with a permittivity of 39 and a physical thickness of 10 nm has an EOT of 1 nm, which means it offers the same gate capacitance as 1 nm of silicon dioxide. To quantify how EOT affects a MOSFET's electrostatic integrity, we can substitute Equation (1.7) into Equation (1.6) to obtain

$$\lambda = \sqrt{\frac{\epsilon^{(SC)}}{\epsilon^{(\mathrm{SiO_2})}} t_{ch} \mathrm{EOT}}. \tag{1.8}$$

Let us consider a hypothetical MOSFET made with a semiconductor with $\epsilon^{(SC)}$ = 10 (a typical value for semiconductors), t_{ch} = 1 nm, and with a target L_{ch} = 10 nm. To achieve strong electrostatic control, we require that $\lambda \approx 1$ nm, which in turn requires the gate insulator to offer an EOT of 0.39 nm. Although silicon dioxide can indeed be grown at physical thicknesses of < 1 nm, such a small EOT cannot practically be obtained using silicon dioxide (or other insulators with low permittivities) because electrons and holes can readily tunnel from the gate electrode to the semiconductor when the insulator is made too thin, which would prevent MOSFETs from ever entering the OFF-state. Consequently, researchers have instead turned towards implementing high-κ insulators in MOSFETs to achieve small EOTs while still maintaining large *physical* thicknesses to suppress gate leakage.

However, implementing high-κ insulators into FETs creates several drawbacks. Firstly, the interfaces between semiconductors and high-κ insulators tend to contain large numbers of interfacial defects, such as trapped charges and recombination sites, which negatively impact device performance. This is particularly problematic for 2-D semiconductors, as their crystal structures do not align well with those of conventional insulators. Consequently, the search for high-κ insulators compatible with 2-D semiconductors has been identified as a key step that must be fulfilled before 2-D semiconductor-based FETs can be realized for most practical applications [14].

Another notable problem with high-κ insulators is that while their high permittivities couple the channel more strongly to the gate, they also tend to enhance lateral fringing fields that run horizontally through FETs. These fringing fields negatively contribute to circuit-level performance metrics of MOSFETs (e.g. by increasing FETs' intrinsic delays) and also deteriorate their ideal source-to-drain transfer characteristics by relinquishing some of the gate's control over the center of the channel. Consequently, while high-κ insulators outperform low-κ insulators at a given physical thickness, they tend to offer worse performance than low-κ insulators at the same EOTs by increasing both SS and OFF-currents, as shown in Fig. 1.4(a). This phenomenon, referred to as *fringe induced barrier lowering* (FIBL), is well known and has been studied since the early 2000s.

To circumvent FIBL while still ensuring that oxides could retain physical thicknesses large enough to prevent gate leakage, Salmani-Jelodar et al. proposed using the MOSFET structure in Fig. 1.4(b), where a thin (< 1 nm) layer of SiO_2 served as a buffering layer and a larger layer composed of a targeted high-κ gate

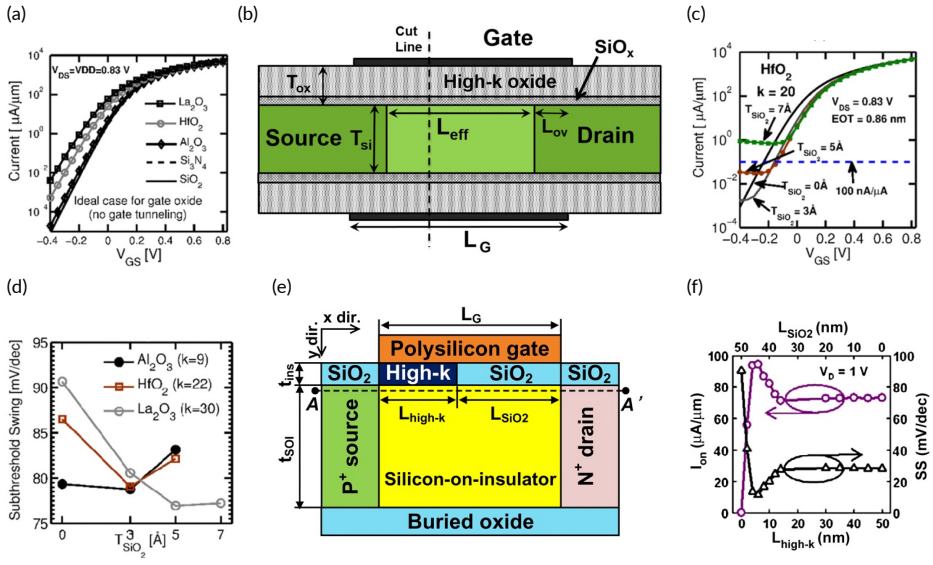

Figure 1.4: **(a)** Transfer characteristics of a MOSFET with various gate insulators, each of whose EOTs are fixed at 0.83 nm. Insulators are listed in the legend in order of decreasing permittivity. **(b)** MOSFET structure to reduce FIBL while maintaining a large physical thickness to prevent source-to-drain tunneling. **(c)** Transfer characteristics of an La_2O_3-based MOSFET with buffer layers of varying thickness, where the EOT of each gate stack is held constant at 0.86 nm. **(d)** SS as a function of buffer layer thickness for 3 high-κ insulators. **(e)** Hetero-gate-dielectric MOSFET architecture used to enhance TFETs. **(f)** I_{on} and SS as functions of geometric parameters as labeled in (e).
(a) – (d) were taken from ref. [15] and are © 2016 IEEE.
(e) – (f) were taken from ref. [16] and are © 2010 IEEE.

dielectric served as the principal gate insulator. For example, Fig. 1.4 (c) shows the transfer characteristics of the architecture in Fig. 1.4(b) with an La_2O_3 gate insulator (ϵ_{ox} = 30) for buffer layers of varying thickness, where EOT was held constant at 0.86 nm for every configuration. The MOSFET with no buffering layer [i.e. the thickness of SiO_2 (t_{SiO_2}) = 0] exhibited an OFF-state current of 10^{-2} μA/μm at V_G = -0.4 V, whereas the OFF current decreased when t_{SiO_2} was set to 3 or 5 Å. However, at 7 Å, the insulator's physical thickness became too small to prevent gate leakage, leading to an extremely high SS in the OFF-state. Similarly, the authors found that the SS's of MOSFETs using other gate insulators could be improved using this buffering layer strategy [Fig. 1.4(d), also at a fixed EOT of 0.86 nm], thus presenting a unique way to maintain large physical oxide thicknesses while still mitigating FIBL.

Finally, recent works have sought to optimize the electrostatics of tunnel field-effect transistors (TFETs, a type of FET whose current is driven by source-to-drain tunneling rather than thermionic emission) using hetero-gate-dielectrics, which are gate insulators that are composed of two or more distinct lateral regions with differing permittivities. Hetero-gate-dielectrics were first proposed by Choi and Lee in 2010 [16], who demonstrated that the hetero-gate-dielectric architecture in Fig. 1.4(e) could improve both the SS's and ON-currents and TFETs. In this architecture, a low-κ insulator, such as SiO_2, spans most of the device, and a small high-κ insulator, such as HfO_2, is inserted between the source/channel interface and the center of the channel. The abrupt transition in the lateral dielectric constant induces a dip in both the valence band and

conduction band edges near the source side of the channel. Consequently, the tunneling distance is decreased when the device is in the ON-state, whereas the tunneling distance remains relatively unchanged when the device is in the OFF-state, leading to an improved current in the ON-state without compromising the SS, as shown in Fig. 1.4(f). Using this optimized architecture, the authors demonstrated that the TFET's ON-state current improved by 30% while SS decreased by 60%.

2 Methodology

2.1 Motivation for Simulating Nanoscale Field-Effect Transistors

In this thesis, I exclusively use computational methods to study the electrostatics of nanoscale FETs and to explore new strategies for engineering FETs' electrostatics. While an FETs' electrostatics can also be studied experimentally, simulation studies carry several distinct advantages that make them especially useful for this task. In particular, much of the work presented in the body of this thesis studies how varying the dielectric constants of various regions in FETs affects devices' transfer characteristics. This type of study is difficult to carry out experimentally because varying a region's dielectric constant by using different materials can result in experimental differences (e.g., processing conditions, interfacial defects), which may introduce confounding factors. Furthermore, intermediate results, such as an FET's electrostatic profile, conduction/valence band edge profiles, current spectra, and charge carrier densities, can be difficult to measure experimentally, though they are readily extracted from device simulations. Consequently, many innovations in electrostatic engineering reported in previous literature were initially studied through simulation methods before their subsequent experimental demonstration.

2.2 Non-Equilibrium Green's Function Simulations

The non-equilibrium Green's function (NEGF) method is a simulation method frequently used to model quantum transport in nanoscale materials and devices. The NEGF method is extremely popular because it is quantum mechanical and atomisticly rigorous and hence can accurately model devices with arbitrarily small dimensions. Other popular modeling techniques for semiconductor devices, such as the semi-classical drift diffusion model, fail to capture the quantum mechanical phenomena (e.g., electron wave interference and quantum mechanical tunneling) that influence nanoscale devices.

The NEGF method models quantum transport using Green's functions and the electrostatic potential using Poisson's equation. Green's functions calculate the spatial distribution of charge carriers as a function of the device's electrostatic potential, whereas Poisson's equation calculates the device's electrostatic potential as a function of the spatial distribution of charge carriers. The NEGF method therefore aims to find a self-consistent electrostatic potential and charge carrier distribution which simultaneously satisfy both sets of equations. Finally, once the density of charge carriers is known, the current flowing through the device may be easily calculated, as shown schematically in Fig. 2.1.

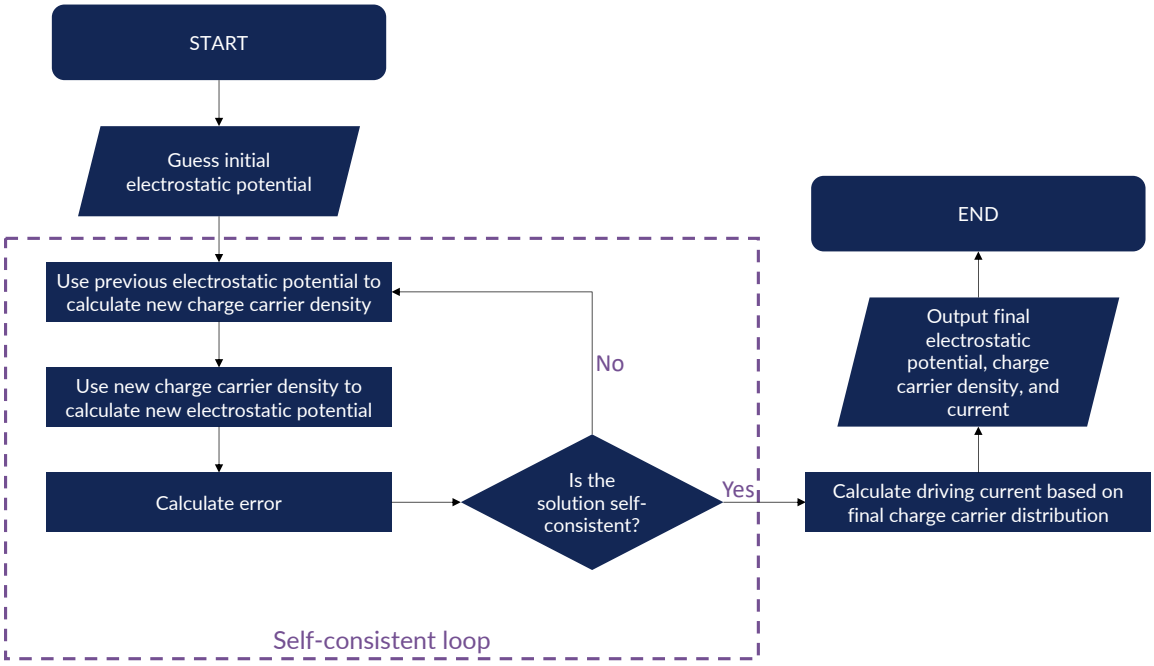

Figure 2.1: A flow chart illustrating the NEGF simulation method.

The work presented in this thesis focuses specifically on engineering the electrostatics of devices, which are modeled exclusively through Poisson's equation. Therefore, this chapter will briefly present the transport equations used in the NEGF method before discussing the modeling of Poisson's equation in detail.

2.2.1 Transport Equations

The basis for the NEGF formalism lies in the Green's function for Schrodinger's equation, given as:

$$G(E) = [(E + \delta i)I - H - q\psi - \Sigma_S - \Sigma_D]^{-1}, \tag{2.1}$$

where E is energy, $i = \sqrt{-1}$, δ is an infinitesimally small positive number, H is a tight-binding (or tight-binding-like) Hamiltonian matrix that describes the electronic properties of the semiconductor, q is the elementary charge, ψ is the electrostatic potential, and Σ_S and Σ_D are the self-energies of the source and drain, respectively. We use Equation (2.1) to find the quantum mechanical electron and hole densities, $G^n(E)$ and $G^p(E)$, as follows:

$$G^n(E) = G(E)\Sigma_\alpha^{in}G^\dagger(E), \tag{2.2}$$

$$G^p(E) = G(E)\Sigma_\alpha^{out}G^\dagger(E), \tag{2.3}$$

where a superscripted dagger denotes the complex conjugate, and where Σ_α^{in} and Σ_α^{out} are referred to as the in-scattering and out-scattering functions, respectively. These functions are calculated using the self-energies of the source and drain as follows:

$$\Sigma_\alpha^{in} = i[\Sigma_S(E) - \Sigma_S^\dagger(E)]f(E), \tag{2.4}$$

$$\Sigma_\alpha^{out} = i[\Sigma_D(E) - \Sigma_D^\dagger(E)][1 - f(E)]. \tag{2.5}$$

Here, $f(E)$ is the Fermi function, which describes the probably of an electron existing at a particular energy level:

$$f(E) = [1 + \exp(\frac{E - E_f}{k_B T})]^{-1}, \tag{2.6}$$

where E_f is the fermi energy, k_B is the Boltzmann constant, and T is the temperature. Once the solutions to the above equations are known, we can find the density of electrons and holes (ρ_n and ρ_p) simply by integrating across all available energies:

$$\rho_n(r) = 2\int_{-\infty}^{\infty} \frac{dE}{2\pi} G^n(E), \tag{2.7}$$

$$\rho_p(r) = 2\int_{-\infty}^{\infty} \frac{dE}{2\pi} G^p(E). \tag{2.8}$$

And finally, we calculate the density of current flowing through the device as:

$$J_\alpha = \frac{2e}{\hbar}\int \frac{dE}{2\pi} \mathrm{Tr}\Big[\Sigma_s(E)[G^n(E) + G^p(E)] - \Gamma_\alpha G^n(E)\Big], \tag{2.9}$$

where Tr is the trace operator.

Equation (2.9) is the end-result of an NEGF simulation and only needs to be computed once we are certain that the charge carrier densities are properly satisfied. To obtain this self-consistent solution, we feed $\rho_n(r)$ and $\rho_p(r)$ into Poisson's equation to calculate the device's electrostatic potential, allowing us to resolve Equation (2.1). When these variables do not change significantly between successive iterations then we consider the equations to be satisfied self-consistently, and only then can we correctly calculate the device's current.

2.2.2 Poisson's Equation

Poisson's equation of electrostatics, or simply *Poisson's Equation*, is a cornerstone of electrostatics and is an essential component for our NEGF device simulations. The derivation of Poisson's equation is straightforward. We begin with the divergence form of Gauss's law in two dimensions:

$$\nabla \cdot D(x, z) = \rho(x, z) \tag{2.10}$$

where ∇ is the gradient operator, D is the displacement field, ρ is the electric charge density, and x and z are spatial coordinates. The displacement field is given by:

$$D(x, y, z) = \bar{\bar{\epsilon}}(x, z)E(x, z) \tag{2.11}$$

where $\bar{\bar{\epsilon}}$ is the dielectric tensor and $E(x, y, z)$ is the electric field. Note that I have written the dielectric tensor as a functional of spatial position to ensure that Equation (2.11) remains valid in heterogenuous media.

We substitute Equation (2.11) into Equation (2.10) to obtain:

$$\nabla \cdot \bar{\bar{\epsilon}}(x, z)E(x, z) = \rho(x, z) \tag{2.12}$$

Finally, we substitute $E(x, z) = -\nabla\psi(x, z)$ (where ψ is the potential) into Equation (2.12) to obtain Poisson's equation:

$$\nabla \cdot [\bar{\bar{\epsilon}}(x, z)\nabla\psi(x, z)] = -\rho(x, z) \tag{2.13}$$

Numerical schemes for solving Equation (2.13) have been presented in previous literature. However, to my knowledge, derivations for these schemes are only presented for the *isotropic* Poisson's equation, where $\bar{\bar{\epsilon}}(x, z)$ is replaced by a scalar permittivity. To derive a numerical scheme to solve the anisotropic Poisson's

equation [i.e. as it is written in Equation (2.13)], I will follow a derivation presented in previous literature for the isotropic Poisson's equation but without making the simplification that $\bar{\bar{\epsilon}}(x,z)$ is a scalar. The original derivation that I will be following was presented by Nagel in 2012 [17].

We use the five-point star method to obtain a numerical scheme with which we may solve Poisson's equation. In this approach, ψ is discretized in a rectangular grid in the x and z dimensions, as illustrated in Fig. 2.2(a). In the most general case, we consider the central point, $\psi^{(i,j)}$, at an interface between four materials whose permittivities are labeled in Fig. 2.2(a). We wish to find a numerical scheme to solve for $\psi_{i,j}$ in terms of the potentials of the neighbouring points along the grid, which will allow us to build a system of equations to solve for $\psi(x,z)$ across the device using standard linear algebra techniques.

To work Equation (2.13) into a form which we may solve numerically, we note that integrating ρ across the differential area $dxdz$ will yield the differential charge of that area, Q. We integrate both sides and apply divergence theorem to the left hand side of Equation (2.13) to obtain:

$$\iint \nabla \cdot [\bar{\bar{\epsilon}}(x,z)\nabla\psi(x,z)]d\Omega = -\iint \rho d\Omega = -Q(x,z) \tag{2.14}$$

$$\oint [\bar{\bar{\epsilon}}\nabla\psi] \cdot dn = -Q(x,z) \tag{2.15}$$

$$\oint \left(\begin{bmatrix} \epsilon_{\parallel}(x,z) & 0 \\ 0 & \epsilon_{\perp}(x,z) \end{bmatrix} \cdot \begin{bmatrix} \partial/\partial x \\ \partial/\partial z \end{bmatrix} \psi(x,z) \right) \cdot dn = -Q(x,z) \tag{2.16}$$

$$\oint \left(\epsilon_{\parallel}(x,z)\frac{\partial\psi(x,z)}{\partial x} + \epsilon_{\perp}(x,z)\frac{\partial\psi(x,z)}{\partial z} \right) \cdot dn = -Q(x,z) \tag{2.17}$$

We expand the left side of Equation (2.17) by treating it as a series of smaller integrals around the exterior of the square. Following the notation of [17], we refer to the edges of the square as S_1, S_2, S_3, and S_4 [as labelled in Fig. 2.2)(b)], and use $\int_{S_n}$ to denote the integral along side S_n. We wish to integrate clockwise around the square, so we write:

$$\oint \left(\epsilon_{\parallel}(x,z)\frac{\partial\psi(x,z)}{\partial x}\hat{x} + \epsilon_{\perp}(x,z)\frac{\partial\psi(x,z)}{\partial z}\hat{z} \right) \cdot dn = \int_{S_1} + \int_{S_2} + \int_{S_3} + \int_{S_4} = -\rho\Delta x\Delta z \tag{2.18}$$

where we noted that $Q \approx \rho\Delta x\Delta z$, where Δx and Δz are the spacings between consecutive grid points along the x and z dimensions, as labeled in Fig. 2.2. Next, we write out and solve the integrals, with a sample

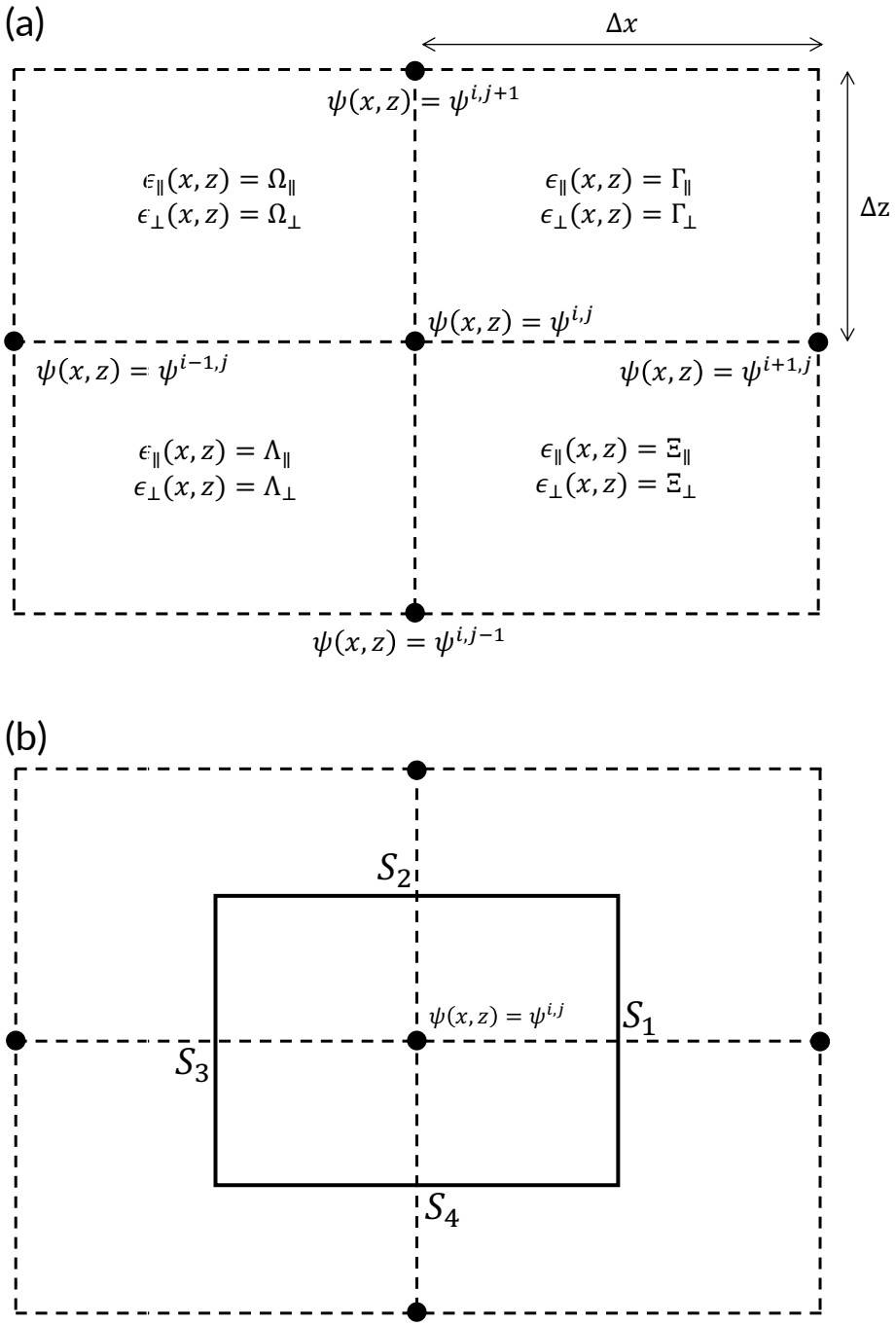

Figure 2.2: **(a)** The five-point star grid used when solving Poisson's equation numerically. Each region in the grid has its own unique permittivities, as labeled. **(b)** The same, showing the square with differential area $dxdz$ used to set up the integrals shown in Equations (2.20) – (2.23). Adapted from [17].

integration for S_1 shown below. Similar integrals for S_2 - S_4 are obtained by updating both the limits of integration and carrying out the dot products with the correct vectors.

$$\int_{S_1} = \int_{\frac{dz}{2}}^{-\frac{dz}{2}} \left(\epsilon_{\parallel}(x,z) \frac{\partial \psi(x,z)}{\partial x} \hat{x} + \epsilon_{\perp}(x,z) \frac{\partial \psi(x,z)}{\partial z} \hat{z} \right) \cdot (-\hat{x}) dz = \int_{-\frac{dz}{2}}^{\frac{dz}{2}} \epsilon_{\parallel}(x,z) \frac{\partial \psi(x,z)}{\partial x} dz \tag{2.19}$$

And we apply a central difference scheme to obtain:

$$\int_{S_1} \approx \Delta z \left(\frac{\Gamma_{\parallel} + \Xi_{\parallel}}{2} \right) \left(\frac{V^{(i+1,j)} - V^{(i,j)}}{\Delta x} \right) = \frac{\Delta z}{\Delta x} \left(\frac{\Gamma_{\parallel} + \Xi_{\parallel}}{2} \right) (V^{(i+1,j)} - V^{(i,j)}) \tag{2.20}$$

This procedure, when carried out on the remaining sides of the square, yields:

$$\int_{S_2} \approx \frac{\Delta z}{\Delta x} \left(\frac{\Gamma_{\perp} + \Omega_{\perp}}{2} \right) (V^{(i,j+1)} - V^{(i,j)}) \tag{2.21}$$

$$\int_{S_3} \approx \frac{\Delta x}{\Delta z} \left(\frac{\Omega_{\parallel} + \Lambda_{\parallel}}{2} \right) (V^{(i-1,j)} - V^{(i,j)}) \tag{2.22}$$

$$\int_{S_4} \approx \frac{\Delta z}{\Delta x} \left(\frac{\Lambda_{\perp} + \Xi_{\perp}}{2} \right) (V^{(i,j-1)} - V^{(i,j)}) \tag{2.23}$$

We then substitute these results back into Equation (2.18) and rearrange to obtain:

$$A_1 V^{(i-1,j)} + A_2 V^{(i,j-1)} + A_3 V^{(i+1,j)} + A_4 V^{(i,j+1)} + A_5 V^{(i,j)} = -\rho^{(i,j)} \tag{2.24}$$

where:

$$A_1 = \frac{\Omega_\parallel + \Lambda_\perp}{2(\Delta z)^2} \tag{2.25}$$

$$A_2 = \frac{\Lambda_\perp + \Xi_\perp}{2(\Delta z)^2} \tag{2.26}$$

$$A_3 = \frac{\Gamma_\parallel + \Xi_\parallel}{2(\Delta x)^2} \tag{2.27}$$

$$A_4 = \frac{\Gamma_\parallel + \Omega_\parallel}{2(\Delta x)^2} \tag{2.28}$$

$$A_5 = \frac{\Omega_\parallel + \Gamma_\parallel + \Lambda_\parallel + \Xi_\parallel}{2(\Delta x)^2} + \frac{\Omega_\perp + \Gamma_\perp + \Lambda_\perp + \Xi_\perp}{2(\Delta z)^2} \tag{2.29}$$

Next, we need to consider our boundary conditions. The simplest boundary condition we consider is used to model the applied gate voltage. In this case, we simply assign:

$$\psi^{(i,j)} = V_G \tag{2.30}$$

where V_G is the applied gate voltage.

Next, we consider the boundary conditions around the simulation domain which do not correspond to a fixed voltage. A simple and effective way to handle these boundaries is to apply Neumann boundary conditions:

$$\frac{\partial \psi}{\partial n} = \frac{\psi_b - \psi_i}{\Delta n} \tag{2.31}$$

where n is the direction normal to the surface (i.e. x for the left edge, $-x$ for the right edge, z for the top edge, and $-z$ for the bottom edge), ψ_b is the potential at a point along the boundary we are considering, and ψ_i is the potential at the nearest grid point adjacent to ψ_n such that ψ_i does not itself fall on the edge of the device. Here, we assume that $\frac{\partial \psi}{\partial n} = 0$ along the boundary, allowing us to write:

$$\psi^{(i,j)} - \psi^{(i+1,j)} = 0 \qquad \text{for the leftmost boundary} \tag{2.32}$$

$$\psi^{(i,j)} - \psi^{(i-1,j)} = 0 \qquad \text{for the rightmost boundary} \tag{2.33}$$

$$\psi^{(i,j)} - \psi^{(i,j-1)} = 0 \qquad \text{for the topmost boundary} \tag{2.34}$$

$$\psi^{(i,j)} - \psi^{(i,j+1)} = 0 \qquad \text{for the bottommost boundary} \tag{2.35}$$

This allows us to adapt our previous coefficients from each case to ensure consistency across our solution. For example, for the leftmost boundary condition, we would assign $A_1 = A_2 = A_4 = 0$, $A_3 = -1$, and $A_5 = 1$ so that Equation (2.24) is identical to Equation (2.31) (where $\rho^{(i,j)} = 0$ in this case).

Our final boundary conditions correspond to the corners of the device. In this case, we apply the boundary condition expressed in Equation (2.31) once in each direction and then add the results to obtain:

$$\frac{\partial \psi}{\partial n_1} = 0 = \frac{\psi_c - \psi_{e_1}}{\Delta n_1} \tag{2.36}$$

$$\frac{\partial \psi}{\partial n_2} = 0 = \frac{\psi_c - \psi_{e_2}}{\Delta n_2} \tag{2.37}$$

$$0 = 2\psi_c - \psi_{e_1} - \psi_{e_2} \tag{2.38}$$

where ψ_c is the potential at the corner we are considering, n_1 and n_2 are the two directions normal to the corner, and ψ_{e_1} and ψ_{e_2} are the potentials of the grid points next to the corner we are considering in the directions n_1 and n_2, respectively. We follow this procedure for each corner and obtain:

$$2\psi^{(i,j)} - \psi^{(i+1,j)} - \psi^{(i,j+1)} = 0 \qquad \text{for the bottom left corner} \tag{2.39}$$

$$2\psi^{(i,j)} - \psi^{(i-1,j)} - \psi^{(i,j+1)} = 0 \qquad \text{for the bottom right corner} \tag{2.40}$$

$$2\psi^{(i,j)} - \psi^{(i+1,j)} - \psi^{(i,j-1)} = 0 \qquad \text{for the top left corner} \tag{2.41}$$

$$2\psi^{(i,j)} - \psi^{(i-1,j)} - \psi^{(i,j-1)} = 0 \qquad \text{for the top right corner} \tag{2.42}$$

This result once again allows us to adapt our previous coefficients from each case to ensure consistency across our solution. For example, for the bottom left corner, we would assign $A_1 = A_2 = 0$, $A_3 = A_4 = -1$, and $A_5 = 2$, so that Equation (2.24) matches Equation (2.38).

After following the above procedures for boundary conditions, we will have established equations to describe the potential at every point in our simulation domain, giving us N equations that correspond to N unknown variables. If the length of the device is L_x and the thickness of the device is L_z, then we will have $N_x = L_x/\Delta x$ discretization points in the x direction and $N_z = L_z/\Delta z$ discretization points in the z direction, giving us a total of $N = N_x N_z$ equations. To finally solve for the potential across the device, we build a system of equations in the form of:

$$\bar{\bar{A}}\bar{\psi} = \bar{\rho} \tag{2.43}$$

where $\bar{\bar{A}}$ is a sparse $N_x N_z \times N_x N_z$ matrix built using the coefficients A_1 - A_5 along the entire device and $\bar{\psi}$ and $\bar{\rho}$ are vectors of length $N_x N_z$ which contain the charge densities and potentials at every point in the device. We fill in $\bar{\bar{A}}$ using Equations (2.25) – (2.29) for points inside of the device and by using the boundary conditions outlined above for the points along the exterior of the simulation boundary, yielding a closed system of linear equations. For small devices, Equation (2.43) may be solved directly through inverting $\bar{\bar{A}}$. For larger devices, inverting an $N_x N_z \times N_x N_z$ sized matrix directly may be computationally inefficient and/or lead to numerical instability, so we may instead employ an iterative Newton-Raphson method, as described in Chapter 3.2 of [18], to solve Equation (2.43) more efficiently.

3 Dielectric Anisotropy in Semiconductors

The work documented in this chapter was published in June 2020 and is © 2020 IEEE (R. K. A. Bennett, D. Yin, and Y. Yoon, "Assessing the Role of a Semiconductor's Anisotropic Permittivity in Hafnium Disulfide Monolayer Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 67, pp. 2607–2613, June 2020, doi: 10.1109/TED.2020.2985023) [1].

3.1 Introduction

2-D FETs are widely studied using computational methods, such as the non-equilibrium Green's function (NEGF) formalism, where the device's charge density is solved self-consistently with its electrostatic potential. The potential is modelled using Poisson's equation, which is a function of the permittivity throughout the device. The permittivities of many 2-D semiconductors are represented by diagonal tensors whose components are the permittivity along the in-plane directions, $\epsilon_{\parallel}^{(SC)}$, and the permittivity in the out-of-plane (stacking) direction, $\epsilon_{\perp}^{(SC)}$ [19]. Many recent studies that used computational methods to study 2-D FETs with semiconductors whose permittivities are anisotropic have taken this anisotropy into account [11, 20, 21, 22, 23, 24]. Many similar works instead represent the permittivity as a scalar [25, 26, 27, 28, 29], thereby assuming the permittivity of the semiconductor is isotropic, thereby making an approximation that I shall refer to as the *isotropic approximation*.[1] The inconsistent use of the isotropic approximation makes comparing otherwise similar studies difficult, as there have been no investigations on the consequences of using the isotropic approximation when modelling materials with anisotropic permittivities. Furthermore, many studies report different values of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for TMD monolayers. To illustrate this, I have plotted $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for TMD monolayers from different sources in Fig. 3.1(a).

Because there have been no works that systematically studied the role of a semiconductor's anisotropic permittivity on the performance of 2-D-material-based MOSFETs, it is impossible to anticipate the differences in device performance that could arise from implementing the isotropic approximation or from using the full anisotropic case with an "incorrect" $\epsilon_{\parallel}^{(SC)}$ and/or $\epsilon_{\perp}^{(SC)}$ when simulating the performance of these devices. In this chapter, I address this gap in knowledge by simulating quantum transport in 2-D FETs. In Chapter 3.3.1, I compare results obtained using the full anisotropic case to results obtained using two different isotropic approximations. I also investigate how the isotropic approximation affects results for devices with different gate oxides and channel lengths. In Chapter 3.3.2, I study the role of $\epsilon_{\parallel}^{(SC)}$ and

[1] Note that I have assumed a work used the anisotropic case if it explicitly said it was doing so or if the authors reported $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ separately (or cited a source for their permittivity that did so, if the work did not report the permittivity directly), while I assumed that a study used the isotropic approximation if the authors reported the dielectric constant as a scalar.

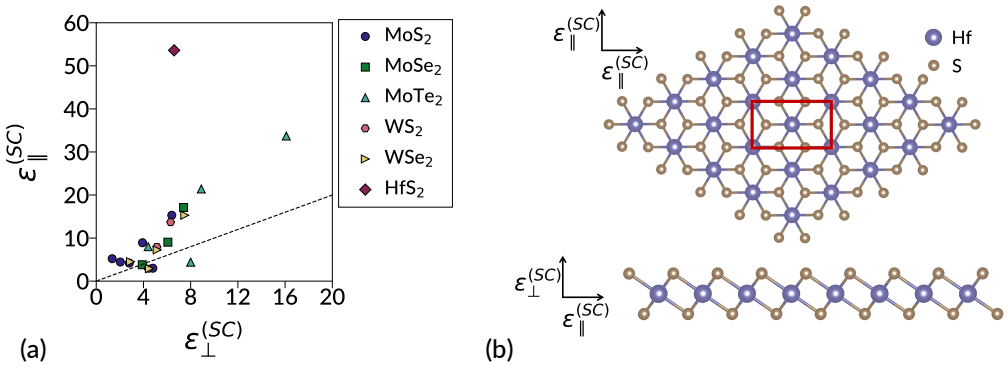

Figure 3.1: **(a)** Various reported values in literature for $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for TMDs. The black dashed line corresponds to $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$. We wish to highlight the anisotropic permittivies of these TMDs and show that there is a discrepancy in the reported $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for these materials (values taken from [11, 19, 20, 21, 26, 30]). **(b)** the lattice structure of 1T HfS_2 from a top-down view (top) and from a side view (bottom). The red box in the top view shows the rectangular supercell used for DFT calculations. The directions of anisotropic permittivities $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ are labelled to the left of each diagram [1]. © 2020 IEEE.

$\epsilon_{\perp}^{(SC)}$ directly, and explain how they influence device performance by considering a capacitor model of FETs. My analysis shows that increasing $\epsilon_{\parallel}^{(SC)}$ worsens device performance by providing capacitance between the source/channel and channel/drain, and increasing $\epsilon_{\perp}^{(SC)}$ improves device performance by increasing the gate capacitance, functioning similarly to the gate oxide.

3.2 Methodology

Devices were simulated using the NEGF method whereby the electrostatic potential was solved self-consistently with the charge density, as described in Chapter 2. In order to study the effect of varying a semiconductor's anisotropic permittivity, I varied $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ when solving the anisotropic form of Poisson's equation [see Equation (2.13)] in the semiconductor.

I simulated 2-D FETs that were comprised of a monolayer of 1T-phase hafnium disulfide (HfS_2), shown in Fig. 3.1(b), with electron transport in the $\Gamma \rightarrow X$ direction. The tight-binding-like Hamiltonian used in [1] was used in the NEGF solver in order to describe the electronic structure of the 1T HfS_2 monolayer. In all cases, I assumed ballistic transport due to the short channel lengths discussed in this work. I used HfS_2 as a model semiconductor because its permittivity is highly anisotropic, with $\epsilon_{\parallel}^{(SC)} = 53.6$ and $\epsilon_{\perp}^{(SC)} = 6.6$ [19] (note that here and for the remainder of this chapter, all permittivities discussed shall refer to the permittivity of the material relative to the vacuum permittivity constant, ϵ_0). The extreme difference in $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ of HfS_2 was beneficial because I wished to highlight the role of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ separately, and using a TMD with a highly anisotropic permittivity helped to distinguish roles from each element of the permittivity individually.

The structure of MOSFETs used in all simulations is shown in Fig. 3.2(a), where a monolayer of HfS_2, which served as the source, channel, and drain, was placed upon a 10 nm thick layer of silicon dioxide (SiO_2) with isotropic permittivity $\epsilon^{(SiO_2)}$ = 3.9. The source and drain were both 10 nm long and were n-doped to a concentration of 10^{14} cm^{-2}. The channel was undoped. In this chapter, the gate oxide is discussed in terms of its equivalent oxide thickness (EOT), which indicates how thick a layer of SiO_2 would have to be to provide an equivalent capacitance. EOT was calculated as EOT = $t_{ox}SiO_2/\epsilon^{(ox)}$, where t_{ox} is the thickness of the oxide and $\epsilon^{(ox)}$ is the (isotropic) permittivity of the oxide. In this chapter, I studied how varying EOT affected device performance by holding t_{ox} constant at 2.92 nm and varying $\epsilon^{(ox)}$. I chose this method of varying EOT based on scaling limits for the gate oxide thickness and trends in research towards high-κ oxides, though it should be noted that this approach does not capture some of the changes in fringing fields that could arise if EOT were varied through changing t_{ox} instead. Unless specified otherwise, all simulations discussed in this chapter were performed at V_D - V_S = 0.5 V, where V_D and V_S are the applied voltages at the drain and source.

In scaling studies, I considered devices with channel lengths (L_{ch}'s) of 6 nm $\leq L_{ch} \leq$ 20 nm. This range was selected because L_{ch} = 6 nm represents aggressively scaled devices with noticeable short-channel effects (SCEs), whereas SCEs become much less noticeable around L_{ch} = 20 nm. For similar reasons, I also considered devices with EOTs in the range of 0.2 nm $\leq$ EOT $\leq$ 1.0 nm. The L_{ch} and EOT of the nominal device were 12 nm and 0.5 nm, respectively. These nominal values were chosen because they are both close to the center of these ranges, and hence offered a useful intermediate between these two extremes.

3.3 Results

3.3.1 Assessment of the Isotropic Approximation

We begin by examining the transfer characteristics of the nominal device using the full anisotropic case with $\epsilon_{\parallel}^{(SC)}$ = 53.6 and $\epsilon_{\perp}^{(SC)}$ = 6.6. I also investigated the same device under an isotropic approximation where the permittivity was set to the in-plane dielectric constant of HfS_2 in what I shall refer to as the *isotropic in-plane case* (i.e. setting $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$ = 53.6), and under an isotropic approximation where the permittivity was set to the out-of-plane dielectric constant of HfS_2 in what I shall refer to as the *isotropic out-of-plane case* (i.e. setting $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$ = 6.6). I considered these two isotropic cases separately to determine how each element of the anisotropic permittivity influenced the transfer characteristics of each device, and to assess which element of the permittivity (if either) could be used in an isotropic approximation that accurately reproduced the results of the full anisotropic case.

Drain current (I_D) as a function of gate voltage (V_G) is plotted in Fig. 3.2(b). At V_G = -0.1 V, I_D for the anisotropic case was 3.5 times greater than that of the isotropic in-plane case and 120 times greater than that of the isotropic out-of-plane case, though these I_D's became similar as V_G increased. The subthreshold swing [SS = $\partial V_G/\partial(\log_{10} I_D)$] was highest for the anisotropic case at 75 mV/dec, lower for the isotropic in-plane case at 71 mV/dec, and lowest for the isotropic out-of-plane case at 66 mV/dec. Since a smaller SS indicates the

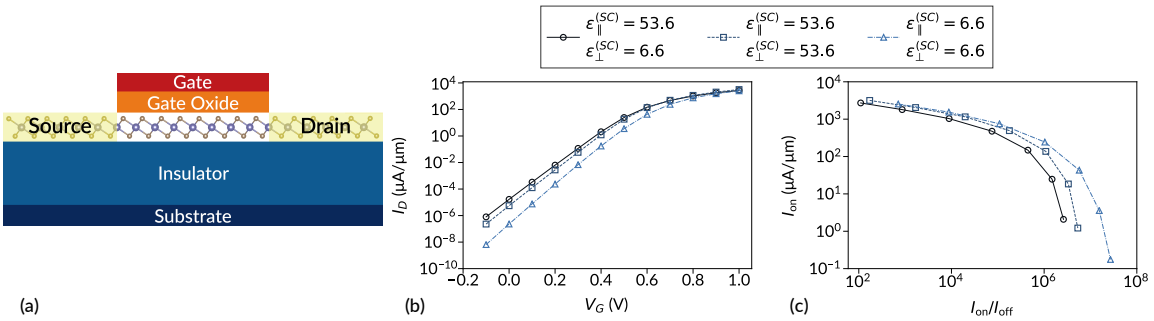

Figure 3.2: **(a)** A diagram of the FET structure used for simulations (not to scale). **(b)** Transfer characteristics of the nominal device (L_{ch} = 12 nm, EOT = 0.5 nm) under three configurations of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$: the full anisotropic case with $\epsilon_{\parallel}^{(SC)}$ = 53.6 and $\epsilon_{\perp}^{(SC)}$ = 6.6 (SS = 75 mV/dec, DIBL = 110 mV/V), the isotropic case with $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$ = 53.6 (SS = 71 mV/dec, DIBL = 80 mV/V), and the isotropic case with $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$ = 6.6 (SS = 66 mV/dec, DIBL = 50 mV/V). **(c)** I_{on}/I_{off} curves for each configuration of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ [1]. © 2020 IEEE.

device transitions to the on-state more quickly, we observe that using either isotropic approximation results in an artificial increase to the I_{on}/I_{off} ratio [Fig. 3.2(c)]. For example, at I_{on} = 500 μA/μm, I_{on}/I_{off} of the isotropic in-plane case was 2.4 times greater than that of the anisotropic case, and I_{on}/I_{off} of the isotropic out-of-plane case was 7.9 times greater than that of the anisotropic case. I_{on}/I_{off} was measured across a voltage window of 0.5 V. For example, if I_{on} was measured at V_G = 1.0 V, then I_{off} was measured at V_G = 0.5 V. The drain-induced barrier lowering [DIBL = $\Delta V_{TH}/\Delta V_D$, where V_{TH} is the threshold voltage] followed a similar trend as SS, with DIBLs of 110, 80, and 50 mV/V for the anisotropic, isotropic in-plane, and isotropic out-of-plane cases, respectively. Clearly, implementing the isotropic approximation overestimated device performance and underestimated SCEs.

Next, I simulated a series of devices with varying EOTs and L_{ch}'s under the anisotropic case and both isotropic approximations. I investigated EOT scaling by holding L_{ch} = 12 nm and varying EOT from 0.2 to 1 nm. At EOT = 0.2 nm, SS = 66 mV/dec for the anisotropic case, 63 mV/dec for the isotropic in-plane case, and 62 mV/dec for the isotropic out-of-plane case [Fig. 3.3(a)]. As EOT increased, the SS of the anisotropic and the isotropic in-plane cases increased at similar rates, while the SS of the isotropic out-of-plane case increased more gradually, causing the isotropic out-of-plane approximation to underestimate the SS at higher EOTs. Next, I considered L_{ch} scaling by holding EOT = 0.5 nm and varying L_{ch} from 6 to 20 nm. At L_{ch} = 6 nm, SS = 130 mV/dec for the anisotropic case, 117 mV/dec for the isotropic in-plane case, and 95 mV/dec for the isotropic out-of-plane case [Fig. 3.3(b)]. This difference became smaller as channel length increased; at L_{ch} = 20 nm, the SS's of the three arrangements were within 3 mV/dec of one another. DIBL plotted as functions of EOT [Fig. 3.3(c)] and L_{ch} [Fig. 3.3(d)] followed similar trends as those for SS.

To understand these different scaling behaviours, we can consider the characteristic length commonly used to estimate the prominence of SCEs in transistors [5]:

$$\lambda = \sqrt{\frac{\epsilon^{(SC)}}{\epsilon^{(\text{ox})}} t_{ch} t_{ox}} = \sqrt{\frac{\epsilon^{(SC)}}{\text{SiO}_2} t_{ch} \text{EOT}} \tag{3.1}$$

where t_{ch} is the thickness of the channel, t_{ox} is the thickness of the oxide, and $\epsilon^{(SC)}$ is the *isotropic* permittivity of the semiconductor. Devices with $L_{ch}/\lambda \geq 5 - 10$ have been shown to have good immunity to SCEs, and SCEs become more prominent as L_{ch}/λ decreases [31].

To my knowledge, while there exist more complicated expressions for λ [32], there have been no studies that report scaling equations that account for anisotropy in the permittivity of the semiconductor. However, Equation (3.1) was derived by considering how electric fields penetrate through the device in the in-plane direction, which is determined by $\epsilon_{\parallel}^{(\text{SC})}$. Furthermore, in all cases in Fig. 3.2(b) and Fig. 3.3, SS and DIBL increased when $\epsilon_{\parallel}^{(\text{SC})}$ increased and $\epsilon_{\perp}^{(\text{SC})}$ decreased. Since a large L_{ch}/λ is characteristic of devices with minimal SCEs, it is apparent that we should use $\epsilon_{\parallel}^{(\text{SC})}$ in place of $\epsilon^{(SC)}$ in (1.6) for semiconductors with anisotropic permittivities to obtain:

$$\lambda = \sqrt{\frac{\epsilon_{\parallel}^{(\text{SC})}}{\text{SiO}_2} t_{ch} \text{EOT}} \tag{3.2}$$

Using Equation (3.2), we consider our scaling results. At EOT = 0.2 nm, $L_{ch}/\lambda > 9$ for all three cases. Each device therefore had good electrostatic control, resulting in similar values of SS and DIBL. As EOT increased, λ increased with a slope proportional to $\sqrt{\epsilon_{\parallel}^{(\text{SC})}}$. Since SS and DIBL increase as λ increases, SS and DIBL increased quickly in Figs. 3.3(a,c) for the anisotropic case and the isotropic in-plane case (where $\epsilon_{\parallel}^{(\text{SC})}$ = 53.6), and more gradually for the isotropic out-of-plane case (where $\epsilon_{\parallel}^{(\text{SC})}$ = 6.6). We can use similar reasoning to understand the effects of scaling L_{ch} in Figs. 3.3(b,d). In this case, however, λ remained constant for each configuration, and the ratio L_{ch}/λ increased as L_{ch} increased. λ was larger for the anisotropic and isotropic in-plane cases than for the isotropic out-of-plane case, and hence the values of SS and DIBL we observed were

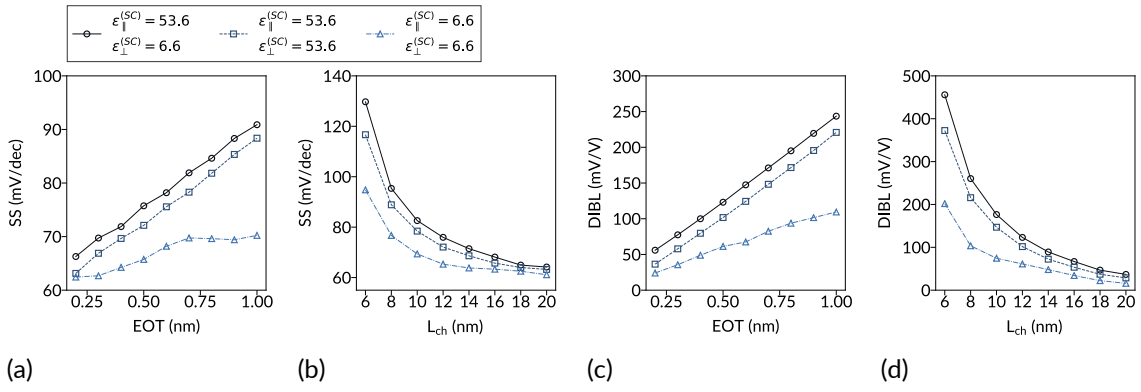

Figure 3.3: SS as a function of **(a)** EOT and **(b)** L_{ch}. DIBL as a function of **(c)** EOT and **(d)** L_{ch} [1]. © 2020 IEEE.

larger for the anisotropic and isotropic in-plane cases as well. As L_{ch} increased, L_{ch}/λ increased for all three cases, causing SS and DIBL to approach ~60 mV/dec and ~24 mV/V, respectively, for each configuration.

While this discussion explains some trends observed for EOT and L_{ch} scaling, Equation (3.2) cannot explain the difference between the anisotropic and isotropic in-plane case because λ is not a function of $\epsilon_{\perp}^{(SC)}$. To understand these differences, we must investigate the role of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ directly.

3.3.2 Investigating the Role of In-Plane and Out-of-Plane Permittivities

To better understand the role of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$, I measured the SS's of a series of devices while varying both elements of the anisotropic permittivity individually. The results from Fig. 3.3(b) and Fig. 3.3(d) showed that differences in $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ more significantly impacted devices with small channel lengths. These more aggressively scaled devices are currently a topic of intense research, and thus it is especially important to understand how $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ affect devices with very small channel lengths. Therefore, I performed these simulations using devices with L_{ch} = 6 nm. As shown in Fig. 3.4(a), SS increased as $\epsilon_{\parallel}^{(SC)}$ increased (holding $\epsilon_{\perp}^{(SC)}$ constant), and SS decreased as $\epsilon_{\perp}^{(SC)}$ increased (holding $\epsilon_{\parallel}^{(SC)}$ constant) when EOT = 0.5 nm.

Consider starting in the bottom left corner of Fig. 3.4(a) and moving along the diagonal $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$ (shown by the dashed arrow). This corresponds to increasing the isotropic permittivity of HfS_2, which has competing effects: SS increases as $\epsilon_{\parallel}^{(SC)}$ increases and decreases as $\epsilon_{\perp}^{(SC)}$ increases. Increasing $\epsilon_{\perp}^{(SC)}$ has a greater effect, causing the overall SS to decrease along the diagonal. This is interesting because it contradicts the conventional wisdom that increasing isotropic permittivity should enhance SCEs [as reflected in (1.6) and demonstrated in a recent simulation study [10]]. To further investigate this phenomenon, I repeated these simulations for devices with an EOT of 1 nm. As shown in Fig. 3.4(b), the trends are similar; increasing $\epsilon_{\parallel}^{(SC)}$ increases SS, and increasing $\epsilon_{\perp}^{(SC)}$ decreases SS. However, the effect of increasing $\epsilon_{\perp}^{(SC)}$ was smaller when the EOT was larger. As a result, SS increases as isotropic permittivity increases along the diagonal $\epsilon_{\parallel}^{(SC)} = \epsilon_{\perp}^{(SC)}$.

It is well established in literature that SS can be minimized by maximizing a device's gate control. A device is said to have good gate control when the potential barrier in the channel is affected strongly by V_G and minimally by V_D. This is achieved by maximizing gate capacitance, $C_G^{(\perp)}$, while minimizing the capacitance between the source and channel, $C_S^{(\parallel)}$, and capacitance between the channel and drain, $C_D^{(\parallel)}$ [2]. Therefore, to understand these results, consider a capacitance network model for a MOSFET, as shown in Fig. 3.4(c). To emphasize which direction a capacitor acts in, I use a superscripted parallel symbol ($\parallel$) when charge separation occurs in the in-plane direction, and a superscripted perpendicular symbol ($\perp$) when charge separation occurs in the out-of-plane direction.

$C_G^{(\perp)}$ is given by the series capacitance $C_G^{(\perp)-1} = C_{ox}^{(\perp)-1} + C_{HfS_2}^{(\perp)-1}$, where $C_{ox}^{(\perp)}$ is the capacitance of the oxide and $C_{HfS_2}^{(\perp)}$ is the capacitance of the HfS_2 monolayer, i.e. $C_{HfS_2}^{(\perp)} = -\partial Q_{HfS_2}/\partial \psi_{ch}$, where Q_{HfS_2} is the charge of the HfS_2 monolayer and ψ_{ch} is the potential of the HfS_2 monolayer. As discussed in [2], $C_{HfS_2}^{(\perp)}$

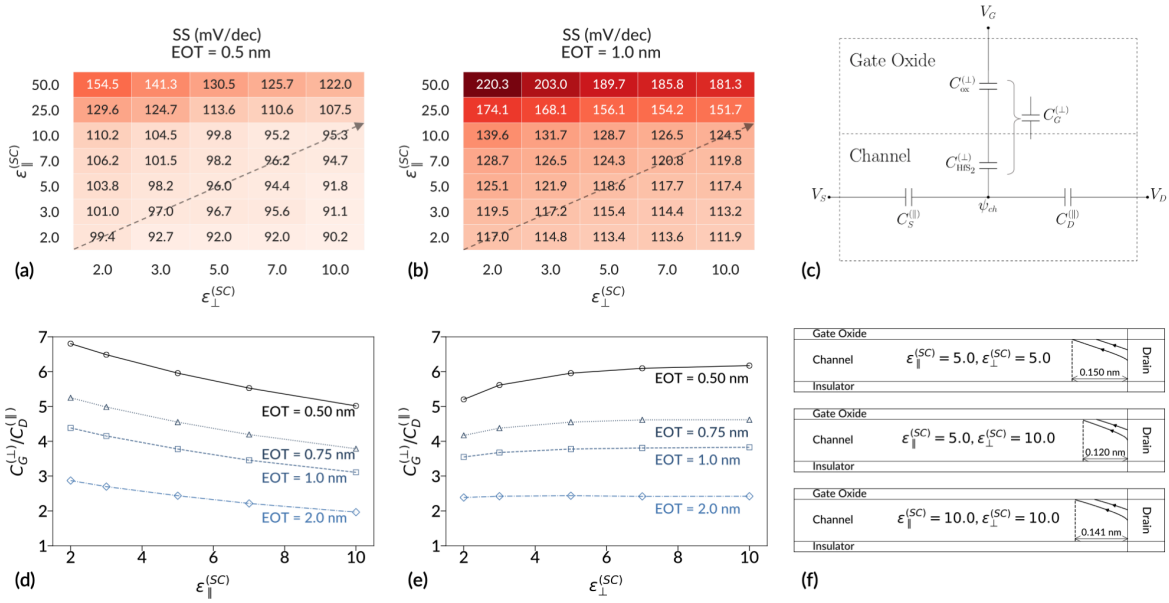

Figure 3.4: Heatmaps showing SS as a function of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for a device with **(a)** with EOT = 0.5 nm and **(b)** EOT = 1.0 nm. SS increases as $\epsilon_{\parallel}^{(SC)}$ increases or $\epsilon_{\perp}^{(SC)}$ decreases in both (a) and (b). However, SS decreases along the diagonal (shown by the black dashed arrow) in (a), and increases along the diagonal in (b). **(c)** A schematic of the capacitor model of a MOSFET. ψ_{ch} is the potential used in the top-of-the-barrier capacitor model to derive an expression for $C_G^{(\perp)}/C_D^{(\parallel)}$. **(d)** $C_G^{(\perp)}/C_D^{(\parallel)}$ plotted as a function of $\epsilon_{\parallel}^{(SC)}$. **(e)** $C_G^{(\perp)}/C_D^{(\parallel)}$ plotted as a function of $\epsilon_{\perp}^{(SC)}$. **(f)** Electric field lines that begin at the drain and terminate on the gate oxide for a device with $\epsilon_{\parallel}^{(SC)}$ = 5.0 and $\epsilon_{\perp}^{(SC)}$ = 5.0 (top), $\epsilon_{\parallel}^{(SC)}$ = 5.0 and $\epsilon_{\perp}^{(SC)}$ = 10.0 (middle), and $\epsilon_{\parallel}^{(SC)}$ = 10.0 and $\epsilon_{\perp}^{(SC)}$ = 10.0 (bottom). Increasing $\epsilon_{\parallel}^{(SC)}$ or decreasing $\epsilon_{\perp}^{(SC)}$ causes these electric field lines to penetrate farther into the channel. Note that these diagrams show a zoomed-in region around the channel/drain interface rather than the entire channel [1]. © 2020 IEEE.

(referred to as the "capacitance of the semiconductor" for a generic semiconductor), is always positive because the amount of negative charge carriers increases as the channel's potential increases.

Analytical expressions for $C_{\text{ox}}^{(\perp)}$, $C_{\text{HfS}_2}^{(\perp)}$, $C_S^{(\parallel)}$, and $C_D^{(\parallel)}$ are quite complicated, so we instead consider how these quantities vary with relevant device parameters:

$$
\begin{aligned}
&\epsilon^{(\text{ox})} \text{ increases or } t_{ox} \text{ decreases} \quad &\rightarrow \quad &C_{\text{ox}}^{(\perp)} \text{ increases} \\
&\epsilon_{\perp}^{(SC)} \text{ increases} \quad &\rightarrow \quad &C_{\text{HfS}_2}^{(\perp)} \text{ increases} \\
&\epsilon_{\parallel}^{(SC)} \text{ increases} \quad &\rightarrow \quad &C_D^{(\parallel)}, C_S^{(\parallel)} \text{ increase}
\end{aligned}
$$

Thus, gate control improves as $\epsilon^{(\text{ox})}$ and $\epsilon_{\perp}^{(SC)}$ increase, and worsens as $\epsilon_{\parallel}^{(SC)}$ and EOT increase. This explains the trends in Figs. 3.4(a,b). As $\epsilon_{\parallel}^{(SC)}$ increased, $C_D^{(\parallel)}$ and $C_S^{(\parallel)}$ increased, causing SS to increase. As $\epsilon_{\perp}^{(SC)}$ increased, $C_G^{(\perp)}$ increased, causing SS to decrease. However, when capacitors are in series, the series capacitance is influenced mostly by the smallest of the individual capacitors. Therefore, when $C_{\text{ox}}^{(\perp)}$ was made small by using a large EOT, increasing $\epsilon_{\perp}^{(SC)}$ only marginally increased $C_G^{(\perp)}$, whereas increasing $\epsilon_{\perp}^{(SC)}$ had a greater

impact on $C_G^{(\perp)}$ when the EOT was small. This explains why SS decreased along the diagonal in Fig. 3.4(a) (EOT = 0.5 nm) and increased along the diagonal in Fig. 3.4(b) (EOT = 1 nm). Increasing $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ at the same rate (i.e. moving along the diagonal starting in the bottom-left hand corner) increased $C_G^{(\perp)}$ more quickly than $C_S^{(\parallel)}$ and $C_D^{(\parallel)}$ only when $C_{ox}^{(\perp)}$ was large.

To verify this discussion, I used a top-of-the-barrier capacitor model to measure the ratio $C_G^{(\perp)}/C_D^{(\parallel)}$ as a function of $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ in simulated devices with L_{ch} = 6 nm and EOTs of 0.5, 0.75, 1.0, and 2.0 nm at V_G = 0 V. In this model, the top of the potential barrier in the channel is given by ψ_{ch}. Using the capacitor diagram in Fig. 3.4(c) and following Equation (10.24) in [2] while assuming that charge is negligible in the subthreshold region and neglecting the buried oxide, ψ_{ch} is written as:

$$\psi_{ch} = \frac{C_G^{(\perp)} V_G + C_D^{(\parallel)} V_D + C_S^{(\parallel)} V_S}{C_G^{(\perp)} + C_D^{(\parallel)} + C_S^{(\parallel)}} \tag{3.3}$$

If $V_S = V_D$, the top of the potential barrier will be in the center of the channel, and $C_D^{(\parallel)} = C_S^{(\parallel)}$. This allows us to rewrite Equation (3.3) as:

$$\frac{C_G^{(\perp)}}{C_D^{(\parallel)}} = 2\frac{V_D - \psi_{ch}}{\psi_{ch} - V_G} \tag{3.4}$$

I extracted V_G, V_D, and ψ_{ch} from the NEGF simulation results and then Equation (3.4) to calculate $C_G^{(\perp)}/C_D^{(\parallel)}$. These simulations were performed at $V_S = V_D$, which ensured that the top of the potential barrier would be in the center of the channel, and consequently $C_D^{(\parallel)} = C_S^{(\parallel)}$ (otherwise, $C_G^{(\perp)}/C_D^{(\parallel)}$ would fail to capture effects from both the source and drain). This also ensured that no tunnelling current would be present, which is necessary when using a top-of-the-barrier model [33].

As shown in Fig. 3.4(d), increasing $\epsilon_{\parallel}^{(SC)}$ caused $C_G^{(\perp)}/C_D^{(\parallel)}$ to decrease, which occurred because increasing $\epsilon_{\parallel}^{(SC)}$ increased $C_D^{(\parallel)}$. $C_D^{(\parallel)}$ is not a function of EOT, so $C_D^{(\parallel)}$ increased at the same rate regardless of the EOT. As shown in Fig. 3.4(e), increasing $\epsilon_{\perp}^{(SC)}$ caused $C_G^{(\perp)}/C_D^{(\parallel)}$ to increase, which occurred because increasing $\epsilon_{\perp}^{(SC)}$ increased $C_{HfS_2}^{(\perp)}$, which in turn increased $C_G^{(\perp)}$. This effect was much more significant at low EOTs because $C_{HfS_2}^{(\perp)}$ had a greater influence on $C_G^{(\perp)}$ when $C_{ox}^{(\perp)}$ was large, and hence variations in $\epsilon_{\perp}^{(SC)}$ more noticeably affected $C_G^{(\perp)}$. These trends are consistent with those observed in Fig. 3.4(a,b) and validate the above discussion that changes in gate control can be understood on the basis of changes in $C_D^{(\parallel)}$ and $C_S^{(\parallel)}$ from $\epsilon_{\parallel}^{(SC)}$ and changes in $C_G^{(\perp)}$ from $\epsilon_{\perp}^{(SC)}$.

The physical reason why increasing $C_G^{(\perp)}/C_D^{(\parallel)}$ causes SS to decrease is because the electric field emanating from the drain cannot penetrate as far into the channel when a device has good gate control [2, 34]. To verify that increasing $\epsilon_{\parallel}^{(SC)}$ increased SS and increasing $\epsilon_{\perp}^{(SC)}$ decreased SS due to changes in gate control, I have plotted the electric field lines that emanate from the drain into the gate for $\epsilon_{\parallel}^{(SC)}/\epsilon_{\perp}^{(SC)}$ = 5/5, 5/10, and 10/10 in Fig. 3.4(f). These simulations were performed with V_G = 0.2 V and EOT = 0.5 nm. The electric field lines

extended 0.150 nm into the channel for $\epsilon_{\parallel}^{(SC)}/\epsilon_{\perp}^{(SC)}$ = 5/5, 0.120 nm into the channel for $\epsilon_{\parallel}^{(SC)}/\epsilon_{\perp}^{(SC)}$ = 5/10, and 0.141 nm into the channel for $\epsilon_{\parallel}^{(SC)}/\epsilon_{\perp}^{(SC)}$ = 10/10. These trends are consistent with the earlier results of Fig. 3.4. Furthermore, just as was the case with Fig. 3.4(a) (where EOT was also 0.5 nm), increasing $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ at the same rate (i.e. going from $\epsilon_{\parallel}^{(SC)}/\epsilon_{\perp}^{(SC)}$ = 5/5 to 10/10) improved gate control because the increase in $C_G^{(\perp)}$ overpowered the increase in $C_S^{(\parallel)}$ and $C_D^{(\parallel)}$.

EOT is a useful metric because it allows one to assess a gate oxide by indicating how thick a layer of SiO_2 would have to be to give the equivalent $C_{ox}^{(\perp)}$. However, the results from Fig. 3.4 indicate that EOT alone cannot be used to assess $C_G^{(\perp)}$ (the actual physical quantity of interest) in devices with aggressively scaled oxides. Therefore, as researchers continue to scale devices, it will become increasingly necessary to report the capacitance from the semiconductor alongside the EOT so that devices can be meaningfully compared.

3.4 Conclusions

The isotropic approximation (i.e. assuming that the permittivity of the semiconductor material is isotropic) becomes inaccurate as the channel lengths of FETs are scaled down. At large EOTs, the isotropic approximation is fairly accurate when the isotropic permittivity is set to the in-plane permittivity, but becomes inaccurate when the isotropic permittivity is set to the out-of-plane permittivity. Therefore, given the trajectory of semiconductor devices, implementing models which account for the full anisotropic permittivity of devices will be essential.

Furthermore, my results show that SCEs are minimized when the semiconductor's in-plane permittivity is small and the out-of-plane permittivity is large. At small EOTs, an equivalent increase in $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ will reduce SCEs because the change in gate capacitance is enough to offset the change in the capacitance between the source and channel as well as the channel and drain. This contradicts predictions made by commonly-cited scaling equations that anticipate lowering the isotropic permittivity should reduce SCEs. Therefore, as researchers continue to develop FETs with aggressively scaled gate oxides, it will become necessary for future works to also report the capacitance of the semiconductor material to properly assess gate control. Finally, due to discrepancies in reported simulated values of the anisotropic permittivity for up-and-coming semiconducting materials, it may be necessary to obtain reliable experimental data for $\epsilon_{\parallel}^{(SC)}$ and $\epsilon_{\perp}^{(SC)}$ for these materials to ensure that future simulation works accurately reflect reality.

4 Dielectric Anisotropy in Insulators

The work documented in this chapter was published in February 2021 and is © 2021 IEEE (R. K. A. Bennett and Y. Yoon, "Using Anisotropic Insulators to Engineer the Electrostatics of Conventional and Tunnel Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 68, no. 2, pp. 865-872, Feb. 2021, doi: 10.1109/TED.2020.3044559) [35].

Many previous works on engineering field-effects transistors' (FETs') electrostatics have sought to optimize capacitances and electric fields that act in both the lateral (in-plane/current-carrying) and perpendicular (out-of-plane/stacking) directions of these devices [41]. Meanwhile, experimental works in the materials sciences have explored insulators whose permittivities are directionally dependent. Throughout this chapter, I refer to these materials as "anisotropic insulators" and have tabulated the in-plane and out-of-plane permittivities ($\epsilon_{\parallel}$ and $\epsilon_{\perp}$) of some of these materials, including homogenous chemical compounds and composites with tailorable dielectric anisotropy, in Table 4.1 (note that the composite materials listed have μm-scale fillers that would need to be scaled down before they could be used in FETs). In principle, these insulators could be used in FETs to separately control directional capacitances and electric fields to engineer devices more precisely. However, to date, the only dielectric material with an anisotropic permittivity that has been studied in FETs is hexagonal boron nitride (hBN). To the best of my knowledge, other anisotropic insulators have not yet been fabricated using CMOS-compatible technologies, which makes it difficult to study their utility in FETs experimentally. Consequently, while the effects of varying the isotropic permittivity of an FET's gate insulator and spacers have been thoroughly studied in literature, the effect of separately varying elements of these insulators' anisotropic permittivities has never been studied before, and the degree to which devices can be improved by implementing anisotropic insulators is unknown.

In this chapter, I address these gaps in knowledge by using NEGF simulations to systematically study metal-oxide-semiconductor FETs (MOSFETs) and tunnel FETs (TFETs) that use anisotropic materials as

Table 4.1: Reported permittivities of sample anisotropic insulators [35]. © 2021 IEEE.

Material	$\epsilon_{\parallel}$	$\epsilon_{\perp}$	
hBN (bulk)	6.9	3.8	[19]
Rutile TiO_2	80	170	[36]
Tetragonal ZrO_2	41.6	14.9	[37]
$CaZrO_3$[a]	34.25	62.4	[38]
TiO_2 composite	9.6	90	[39]
Polyvinylidene fluoride composite[b]	16	27	[40]

[a] $\epsilon_{\parallel}$ was taken as the average value of the dielectric constant in the two in-plane directions (34.8 and 33.7)

[b] At 8% volume fraction of Bi_2S_3 nanorod filler

these devices' gate insulators and spacers. Here, I find that for MOSFETs, the in-plane permittivities of the gate insulator and spacers ($\epsilon_{\parallel}^{(ox)}$ and $\epsilon_{\parallel}^{(sp)}$) should be minimized to reduce fringing effects, while the out-of-plane permittivities of the gate insulator and spacers ($\epsilon_{\perp}^{(ox)}$ and $\epsilon_{\perp}^{(sp)}$) should be maximized to improve gate control. I take advantage of this to show how anisotropic insulators that span the source, channel, and drain could obviate the need for low-κ spacers. I also demonstrate that the subthreshold performance and ON current (I_{on}) of TFETs can be dramatically improved by using anisotropic gate insulators, and introduce a new hetero-gate dielectric for TFETs based on changing the gate insulator's in-plane permittivity.

4.1 Methodology

Devices were simulated using the NEGF method, where transport equations were solved self-consistently with the electrostatic potential, as discussed in Chapter 2. Here, I study the effect of dielectric anisotropy in an FET's insulators by varying elements of the spatially-dependent permittivity tensor while solving Poisson's equation [Equation (2.13)] in the FET's gate insulator and/or spacers.

MOSFETs followed the single-gated design shown in Fig. 4.1(a), where a monolayer of black phosphorous (BP) served as the semiconductor. TFETs followed the double-gated design shown in Fig. 4.1(b), as stronger electrostatic control is essential for high-performing TFETs. Monolayer BP is suboptimal for TFETs due to its large band gap (E_g = 1.52 eV) [42], so I instead used bilayer BP (E_g = 1.12 eV [43]) as the semiconductor in TFETs. In all simulations, transport occurred along the armchair direction of BP. In this chapter, I chose to use a 2-D material for the channel of FETs because 2-D materials can offer smaller computational burdens compared to conventional bulk semiconductors due to their relatively small tight-binding Hamiltonian matrices. In particular, I used BP as a model 2-D semiconductor because it not only offers a high computational efficiency but also exhibits excellent performance when used in FETs.

The electronic structure of BP was described using a tight-binding approximation where interactions between neighbouring atoms were described in terms of previously reported hopping parameters [44]. The semiconductor's in-plane and out-of-plane permittivities were set to 4.56 and 1.36 for monolayer BP and 7.41 and 1.52 for bilayer BP based on a previous study [24]. Note that here and for the remainder of this chapter, all permittivities have been scaled by ϵ_0.

The channels for MOSFETs and TFETs were 15 and 20 nm long, respectively, and the sources and drains of both were 20 nm long. The sources and drains of MOSFETs and the drains of TFETs were n-doped, while the sources of TFETs were p-doped. The concentrations of dopants in the source and drain were 1.5 $\times$ 10^{13} cm^{-2} for MOSFETs and 2 $\times$ 10^{13} cm^{-2} for TFETs, and the channels were intrinsic. The buried oxide for MOSFETs was 20 nm thick, and all gate insulators were 6.55 nm thick. V_D - V_S = 0.5 and 0.7 V for MOSFETs and TFETs, respectively, where V_D and V_S are the potentials of the drain and source.

For simplicity, I assumed all transport was ballistic. As the mean free path in thin BP is $\lambda \approx 14$ nm [45, 46], there might be slight scattering at channel lengths considered in this chapter. Nevertheless, a simulation

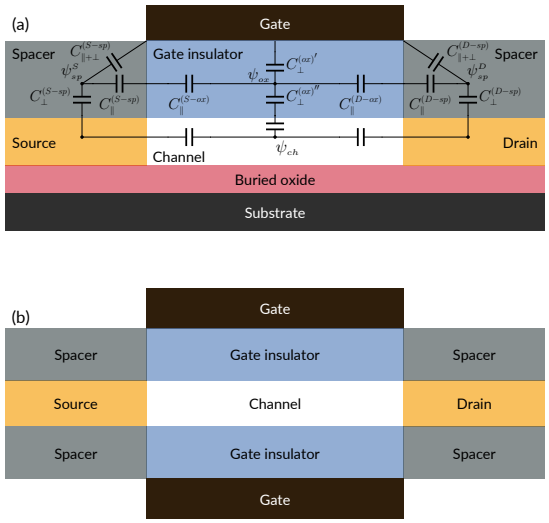

Figure 4.1: Device structures for **(a)** MOSFETs and **(b)** TFETs. The overlain capacitors and potentials in (a) will be referenced in Chapter 4.2.1 to understand the role of an insulator's anisotropic permittivity [35]. © 2021 IEEE.

study has shown that neglecting scattering does not significantly affect the performance of BP FETs at L_{ch} = 20 nm[47], and other NEGF studies have also used the ballistic approximation to study BP FETs with the same channel length [48]. Therefore, this approximation should not significantly affect the results reported in this chapter.

Except for where otherwise noted, MOSFETs and TFETs used spacers made of silicon dioxide (SiO_2). Gate insulators are discussed in terms of their equivalent oxide thickness (EOT), calculated as EOT = $t_{ox}(\epsilon^{(SiO_2)}/\epsilon_{\perp}^{(ox)})$, where t_{ox} is the thickness of the insulator and $\epsilon^{(SiO_2)}$ is the permittivity of SiO_2. An FET's OFF-state voltage V_{OFF} is the gate voltage V_G at which the drain current I_D is equal to the OFF current $I_{off} = 10^{-6}$ µA/µm. Subthreshold swing (SS) is measured at V_{OFF}, and I_{on} is measured at an ON-state voltage $V_{ON} = V_{OFF} + V_{DD}$, where the power supply voltage V_{DD} = 0.5 V for MOSFETs and 0.7 V for TFETs. Intrinsic delay is calculated as $\tau = (Q_{ON} - Q_{OFF})/I_{on}$, where Q_{ON} and Q_{OFF} are the charges in the channel region at $V_G = V_{ON}$ and $V_G = V_{OFF}$ [49].

4.2 Results and Discussion

4.2.1 MOSFETs

I begin by simulating the I_D-V_G characteristics of MOSFETs while varying elements of the gate insulator's permittivity. Fig. 4.2(a) shows the I_D-V_G characteristics of devices with $\epsilon_{\parallel}^{(ox)}/\epsilon_{\perp}^{(ox)}$ = 25/10 (SS = 96.6 mV/dec), 25/25 (SS = 75.2 mV/dec), and 2/25 (SS = 70.8 mV/dec). Increasing $\epsilon_{\perp}^{(ox)}$ from 10 to 25 strongly reduces I_{off}

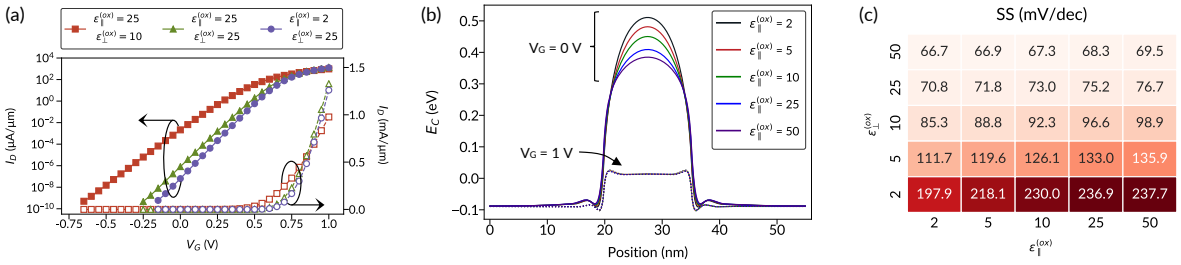

Figure 4.2: **(a)** I_D-V_G characteristics for MOSFETs with varied $\epsilon_{\parallel}^{(ox)}/\epsilon_{\perp}^{(ox)}$. Solid lines/filled markers and dashed lines/open markers correspond to the left (log scale) and right (linear scale) y axes, respectively. **(b)** Conduction band (E_C) profiles of devices at V_G = 0 and 1 V while varying $\epsilon_{\parallel}^{(ox)}$ ($\epsilon_{\perp}^{(ox)}$ = 10 and $V_D = V_S$ = 0 V). **(c)** Heatmap of SS as a function of $\epsilon_{\perp}^{(ox)}$ and $\epsilon_{\parallel}^{(ox)}$ [35]. © 2021 IEEE.

and increases I_{on}, thus decreasing SS. Decreasing $\epsilon_{\parallel}^{(ox)}$ from 25 to 2 decreases I_{off} noticeably while only slightly decreasing I_{on}, causing SS to decrease moderately when $\epsilon_{\parallel}^{(ox)}$ decreases. Drain-induced barrier lowering (DIBL) follows the same trend as SS for these three configurations of $\epsilon_{\parallel}^{(ox)}/\epsilon_{\perp}^{(ox)}$, which have DIBL values of 299, 129, and 98 mV/V, respectively. Gate capacitance C_G is proportional to $\epsilon_{\perp}^{(ox)}$, which is why increasing $\epsilon_{\perp}^{(ox)}$ from 10 to 25 improves device performance. As this has been studied extensively in past works [2], I instead focus more on $\epsilon_{\parallel}^{(ox)}$, whose role is less understood.

Fig. 4.2(b) shows that at $V_D = V_S$ = 0 V, increasing $\epsilon_{\parallel}^{(ox)}$ decreases the height of the potential energy barrier at V_G = 0 V (OFF-state), while the barrier's height is nearly independent of $\epsilon_{\parallel}^{(ox)}$ at V_G = 1 V (ON-state). This is why I_{off} increases and I_{on} is mostly unchanged as $\epsilon_{\parallel}^{(ox)}$ increases. The reasons for these trends in barrier heights can be understood on the basis of the inner fringing capacitances, which act within the insulator in the in-plane direction and are hence associated with $\epsilon_{\parallel}^{(ox)}$. Increasing fringing capacitances reduces gate control [50], which is why we observe a reduction in barrier height at V_G = 0 V. These fringing capacitances are screened out as the device enters the ON-state [41], which is why the barriers are almost independent of $\epsilon_{\parallel}^{(ox)}$ at V_G = 1 V.

The role of $\epsilon_{\perp}^{(ox)}$ and $\epsilon_{\parallel}^{(ox)}$ can be better understood by considering the capacitor network in Fig. 4.1(a). ψ_{ch} is the potential in the channel, while ψ_{ox}, ${\psi_{sp}}^S$, and ${\psi_{sp}}^D$ are the potentials of generic points in the gate insulator, source-side spacer, and drain-side spacer. Two points in the capacitance network are highly coupled (i.e. mutually influence each other) when the capacitance between them is large. Increasing $\epsilon_{\perp}^{(ox)}$ increases ${C_{\perp}^{(ox)}}'$ and ${C_{\perp}^{(ox)}}''$, which in turn couple ψ_{ch} more strongly to ψ_{ox} and ψ_{ox} more strongly to V_G, resulting in enhanced gate control. However, increasing $\epsilon_{\parallel}^{(ox)}$ increases $C_{\parallel}^{(S-ox)}$ and $C_{\parallel}^{(D-ox)}$. These capacitors couple ψ_{ox} to ${\psi_{sp}}^S$ and ${\psi_{sp}}^D$, which are coupled to the source and drain. Therefore, increasing $\epsilon_{\parallel}^{(ox)}$ increases the influence of the source and the drain on the channel, thereby decreasing gate control.

$C_{\parallel}^{(S-ox)}$ and $C_{\parallel}^{(D-ox)}$ can be changed in isotropic insulators by varying the gate oxide's isotropic permittivity $\epsilon_{iso}^{(ox)}$, which also changes ${C_{\perp}^{(ox)}}'$ and ${C_{\perp}^{(ox)}}''$. This overpowers the changes in the inner fringing capacitances, causing device performance to improve as $\epsilon_{iso}^{(ox)}$ increases. Consequently, previous works have not been able to quantify how varying $\epsilon_{\parallel}^{(ox)}$ impacts device performance. To address this, we may consider SS as a function of $\epsilon_{\perp}^{(ox)}$ and $\epsilon_{\parallel}^{(ox)}$ in Fig. 4.2(c). Regardless of the $\epsilon_{\perp}^{(ox)}$ used, decreasing $\epsilon_{\parallel}^{(ox)}$ decreases SS, thereby improving device

performance. This effect is stronger when $\epsilon_{\perp}^{(ox)}$ is small, though it is still quite noticeable even at higher $\epsilon_{\perp}^{(ox)}$'s. For example, at $\epsilon_{\perp}^{(ox)}$ = 25 (EOT = 1 nm), SS decreases by 4.4 mV/dec when $\epsilon_{\parallel}^{(ox)}$ decreases from 25 to 2.

To further understand the physical mechanisms by which changing $\epsilon_{\parallel}^{(ox)}$ affects MOSFETs, we consider the fundamental electrostatics of FETs with anisotropic gate oxides. Previous works have studied the effects of electrostatic screening of charges in anisotropic media and showed that potential contour lines penetrate farther into the dielectric film when the in-plane permittivity is smaller than the out-of-plane permittivity, and become flat when the in-plane permittivity is larger than the out-of-plane permitttivity [51, 52]. Figs. 4.3(a,b) show the contour plots of potential energy in the gate oxide for devices with $\epsilon_{\parallel}^{(ox)}$ = 2 and $\epsilon_{\parallel}^{(ox)}$ = 50 (with $\epsilon_{\perp}^{(ox)}$ = 10, V_G = 0 V, and $V_D = V_S$ for both) and show that they follow the same behaviour: at $\epsilon_{\parallel}^{(ox)}$ = 2, the contour lines stretch vertically into the gate oxide, whereas the contour lines are much flatter at $\epsilon_{\parallel}^{(ox)}$ = 50. Since the shapes of these contour lines are already understood, I focus on analysing how they affect device performance. The plots to the right of Figs. 4.3(a,b) show close-ups of the potentials in the regions enclosed in the red dashed lines around the channel/drain interface. When $\epsilon_{\parallel}^{(ox)}$ =2, the contour lines above the channel/drain interface quickly drop to the channel, whereas when $\epsilon_{\parallel}^{(ox)}$ = 50, the contour lines extend farther into the channel. Therefore, when $\epsilon_{\parallel}^{(ox)}$ is large, the drain exerts greater control at regions closer to the center of the channel, which is the fundamental reason why gate control decreases as $\epsilon_{\parallel}^{(ox)}$ increases. For example, consider the contour line that begins 1 nm above the channel/drain interface, highlighted with a red

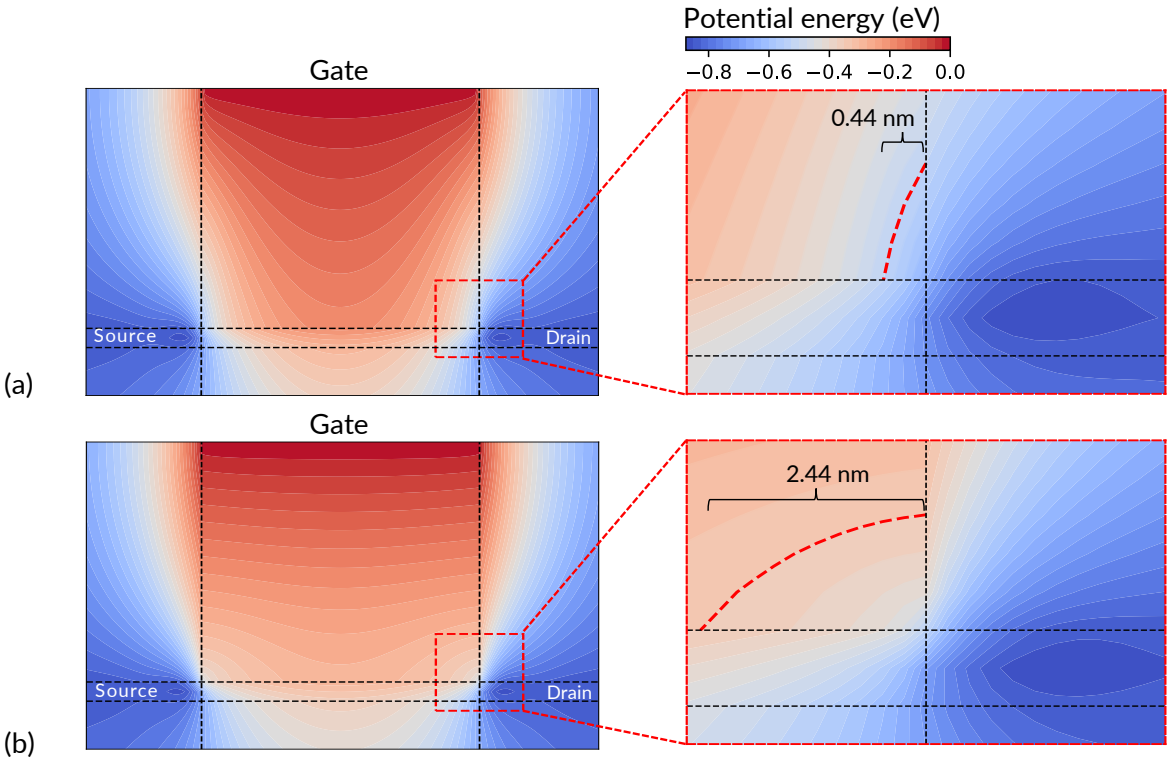

Figure 4.3: Contour plots showing the potential energy in devices with anisotropic gate oxides for **(a)** $\epsilon_{\parallel}^{(ox)}$ = 2 and **(b)** $\epsilon_{\parallel}^{(ox)}$ = 50. The plots on the right show the potential energy contour lines of the region around the channel/drain interface enclosed in the red dashed box.

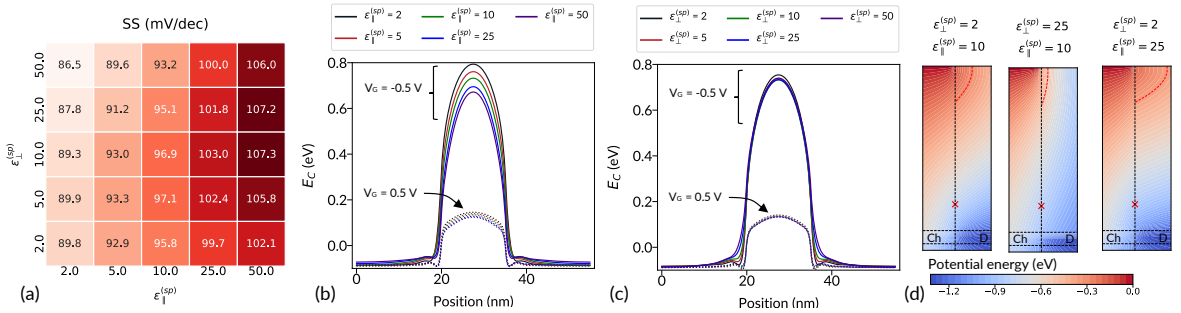

Figure 4.4: **(a)** Heat map showing SS as a function of the permittivities of the spacer, $\epsilon_{\perp}^{(sp)}$ and $\epsilon_{\parallel}^{(sp)}$ (with $\epsilon_{\parallel}^{(ox)} = \epsilon_{\perp}^{(ox)} = 10$). Conduction band profiles of devices at V_G = -0.5 and 0.5 V ($V_D = V_S = 0$ V) **(b)** for various $\epsilon_{\parallel}^{(sp)}$ (with $\epsilon_{\perp}^{(sp)} = \epsilon_{\parallel}^{(ox)} = \epsilon_{\perp}^{(ox)} = 10$) and **(c)** for various $\epsilon_{\perp}^{(sp)}$ (with $\epsilon_{\parallel}^{(sp)} = \epsilon_{\parallel}^{(ox)} = \epsilon_{\perp}^{(ox)} = 10$). The inset shows a close-up of the region towards the top of the barriers enclosed by a red dashed rectangle [35]. © 2021 IEEE.

dashed line. This contour line extends 0.44 nm into the channel when $\epsilon_{\parallel}^{(ox)} = 2$, and 2.44 nm into the channel when $\epsilon_{\parallel}^{(ox)} = 50$.

Next, I investigate how the anisotropic permittivities of the spacers affect the subthreshold performance of MOSFETs. SS is shown as a function of $\epsilon_{\parallel}^{(sp)}$ and $\epsilon_{\perp}^{(sp)}$ for MOSFETs with $\epsilon_{\parallel}^{(ox)} = \epsilon_{\perp}^{(ox)} = 10$ in Fig. 4.4(a), which shows that SS decreases as $\epsilon_{\parallel}^{(sp)}$ decreases. This is consistent with the previous discussion, as ψ_{ox} is coupled to ${\psi_{sp}}^{S}$ and ${\psi_{sp}}^{D}$ through $C_{\parallel}^{(S-sp)}$ and $C_{\parallel}^{(D-sp)}$, so decreasing $\epsilon_{\parallel}^{(sp)}$ provides a similar effect to decreasing $\epsilon_{\parallel}^{(ox)}$. However, SS is non-monotonic with respect to $\epsilon_{\perp}^{(sp)}$, initially increasing and then decreasing as $\epsilon_{\perp}^{(sp)}$ increases. These trends are reflected in the conduction band profiles, where the barrier at V_G = -0.5 V constantly decreases and the barrier at V_G = 0.5 V is mostly unchanged as $\epsilon_{\parallel}^{(sp)}$ increases [Fig. 4.4(b)]. Meanwhile, the height of the barrier at V_G = -0.5 V is non-monotonic with respect to $\epsilon_{\perp}^{(sp)}$, converging again to similar values at V_G = 0.5 V [Fig. 4.4(c)].

From the capacitance network shown in Fig. 4.1(a), ${\psi_{sp}}^{S}$ is coupled to the source through $C_{\perp}^{(S-sp)}$, which increases as $\epsilon_{\perp}^{(sp)}$ increases. Since ${\psi_{sp}}^{S}$ is coupled laterally to ψ_{ox}, which is in turn coupled to ψ_{ch}, increasing $\epsilon_{\perp}^{(sp)}$ increases the influence of the source on the channel's potential, thereby decreasing gate control. However, ${\psi_{sp}}^{S}$ is also coupled to the gate through $C_{\parallel+\perp}^{(S-sp)}$, which increases as $\epsilon_{\perp}^{(sp)}$ increases. Increasing $\epsilon_{\perp}^{(sp)}$ thus increases the influence of the gate on ${\psi_{sp}}^{S}$, which increases the influence of the gate on ψ_{ch} through the same coupling path as before. The same effect occurs on the drain side using the analogous capacitors and potentials. Therefore, increasing $\epsilon_{\perp}^{(sp)}$ increases both the influence of the source/drain and the gate on ψ_{ch}, which creates competing effects that lead to the non-monotonic behaviour observed in Figs. 4.4(a,c).

To confirm the above discussion, I have plotted the potential energy contours around the channel/drain interfaces in Fig. 4.4(d) (this is the same region enclosed in the red dashed lines in Fig. 4.3, where the top has been extended to the gate) for devices with $\epsilon_{\perp}^{(sp)}/\epsilon_{\parallel}^{(sp)}$ = 10/2, 25/2, and 25/10. As $\epsilon_{\perp}^{(sp)}$ decreased from 25 to 10 (comparing the first and second plots in Fig. 4.4(d) with $\epsilon_{\parallel}^{(sp)} = 10$), the potential emanating from the gate became tighter, which indicates that gate control increased [53]. For example, the contour line located 1.7 nm below the gate at the channel/drain interface (highlighted with a dashed red line) extended 1.2 and 0.62 nm into the spacer when $\epsilon_{\perp}^{(sp)}$ was 10 and 25, respectively. However, the potential at the drain/channel interface

1 nm above the channel (marked with red crosses in the plots in Fig. 4.4(d)) also decreased from -0.646 to -0.805 eV, indicating that the drain exerted greater control along the channel/drain interface. This confirms that these two competing effects are the origin of the non-monotonic trends of Figs. 4.4(a,c) As $\epsilon_{\parallel}^{(sp)}$ increased from 2 to 25 (comparing the first and third plots in Fig. 4.4(d) with $\epsilon_{\perp}^{(sp)} = 2$), however, the fringing fields became broader (the same highlighted contour line extended farther to 2.2 nm laterally) and the potential 1 nm above the channel/drain interface further decreased to -0.866 eV. These work co-operatively to decrease gate control, which is why monotonic trends were observed for $\epsilon_{\parallel}^{(sp)}$.

While these simulations assumed that gates were 2-D by neglecting their heights, realistic devices will usually have three-dimensional (3-D) gates. As the height of the gate can affect fringing fields in the spacers, the role of the anisotropic permittivities of the spacers may also change depending on the gate's height. Therefore, while changing the height of the gate will not significantly change the qualitative nature of the trends observed when varying the spacers' anisotropic permittivities, future works may benefit from a quantitative analysis of spacers with anisotropic permittivities in MOSFETs with 3-D gates.

Fringing effects are often minimized in MOSFETs by using spacers with low isotropic permittivities. Air spacers are ideal for this but add complexity to fabrication and have not been successfully implemented at the 3 nm node [54]. However, the previous results show high out-of-plane permittivities and low in-plane permittivities are desirable in both the gate insulators and the spacers. Therefore, anisotropic insulators could obviate the need for low-κ spacers, since a single anisotropic insulator could provide the universally required low in-plane and high out-of-plane permittivities.

To explore this, I simulated FETs following the architecture shown in Fig. 4.5(a), where the gate insulator and spacers are replaced with a single anisotropic insulator. The insulator's out-of-plane permittivity was fixed at $\epsilon_{\perp}^{(ox+sp)} = 25$, and I measured the SS as a function of the insulator's in-plane permittivity $\epsilon_{\parallel}^{(ox+sp)}$. As shown in Fig. 4.5(b), SS improves from 81 to 71 as $\epsilon_{\parallel}^{(ox+sp)}$ decreases from 25 [equivalent to hafnium dioxide (HfO_2)] to 2. For comparison, devices with HfO_2 gate insulators ($\epsilon_{\parallel}^{(ox)} = \epsilon_{\perp}^{(ox)} = 25$) with SiO_2 and air spacers have SS's of 75.6 and 73.8 mV/dec, respectively [marked by the green and red dashed lines in Fig. 4.5(b)]. The device with a single anisotropic insulator has these SS values when $\epsilon_{\parallel}^{(sp)}$ is 13 and 9. Furthermore, this structure can offer an improved intrinsic delay τ, which decreases from 4.53 to 1.26 ps as $\epsilon_{\parallel}^{(ox+sp)}$ decreases from 25 to 2 [Fig. 4.5(c)]. I attribute this decrease in τ to a reduction in lateral fringing parasitic capacitances, which decrease as $\epsilon_{\parallel}^{(ox+sp)}$

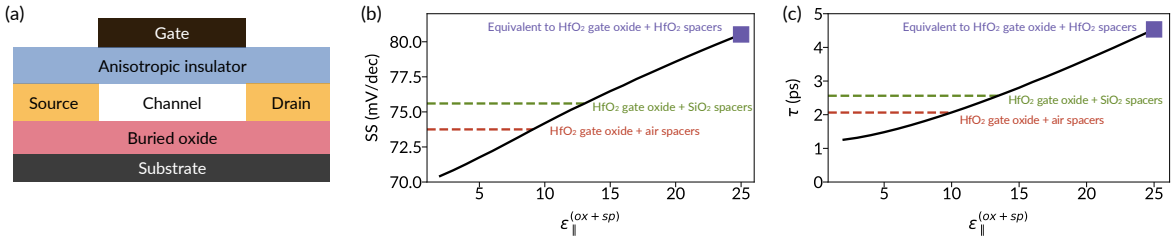

Figure 4.5: **(a)** Revised MOSFET architecture where the gate insulator and spacers are replaced by a single anisotropic insulator. **(b)** SS and **(c)** τ as a function of $\epsilon_{\parallel}^{(ox+sp)}$ (with $\epsilon_{\perp}^{(ox+sp)} = 25$) for the device architecture shown in (a) [35]. © 2021 IEEE.

decreases. Devices with HfO_2 gate insulators with SiO_2 and air spacers have τ = 2.56 and 2.06 ps, whereas the device with a single anisotropic insulator obtains these values of τ at $\epsilon_{\parallel}^{(sp)}$ = 13 and 10 [marked by the green and red dashed lines in Fig. 4.5(c)]. These results suggest that this new architecture could provide enhanced I_D-V_G characteristics while simultaneously offering an improved circuit-level performance.

4.2.2 Tunnel FETs

Fig. 4.6(a) shows I_D-V_G characteristics of TFETs with various values of $\epsilon_{\parallel}^{(ox)}$ with $\epsilon_{\perp}^{(ox)}$ = 10. SS is minimized and I_{on} is maximized when $\epsilon_{\parallel}^{(ox)}$ = 2, though the trends for SS and I_{on} are non-monotonic, with SS and I_{on} initially worsening and then improving as $\epsilon_{\parallel}^{(ox)}$ increases from 2 to 50. The values of SS and I_{on} for TFETs with various values of $\epsilon_{\parallel}^{(ox)}$ and $\epsilon_{\perp}^{(ox)}$ are summarized in the heat maps shown in Figs. 4.6(b,c), which demonstrate that the performance of TFETs can be improved drastically by implementing anisotropic insulators. For example, an isotropic gate insulator with $\epsilon_{\perp}^{(ox)} = \epsilon_{\parallel}^{(ox)}$ = 10 (EOT = 2.5 nm) provides an SS of 43.4 mV/dec and an I_{on} of 6.6 μA/μm. The SS can be reduced by 34% and I_{on} can be more than tripled by instead using an insulator with the same EOT (i.e. $\epsilon_{\perp}^{(ox)}$ = 10) and $\epsilon_{\parallel}^{(ox)}$ = 2, which offers an SS of 28.7 mV/dec and an I_{on} of 20.2 μA/μm.

To understand these non-monotonic trends, we may consider the band profiles plotted in Fig. 4.6(d) for devices with $\epsilon_{\perp}^{(ox)}$ = 10 and $\epsilon_{\parallel}^{(ox)}$ = 2, 10, and 50 at a common V_G such that $V_G \approx V_{ON}$ for each device. The source/channel and channel/drain interfaces are marked with vertical lines, and the energy at which maximum tunnelling occurs is marked with a horizontal dashed line. As $\epsilon_{\parallel}^{(ox)}$ increases, the valence band edge (E_V) to the left of the source/channel interface is pushed downwards. Decreasing E_V in this region increases the width of the tunnelling barrier, assuming the conduction band edge (E_C) to the right of this interface is unchanged. However, increasing $\epsilon_{\parallel}^{(ox)}$ also decreases E_C to the immediate right of the source/channel interface, which decreases the width of the tunnelling barrier. This suggests that these competing effects could be the origin of the non-monotonic trends observed throughout Fig. 4.6.

To better understand how changing $\epsilon_{\parallel}^{(ox)}$ affects the band profiles, we may consider how relevant points in the channel are coupled together. Electrons tunnel from the left of the source/channel interface to the right edge

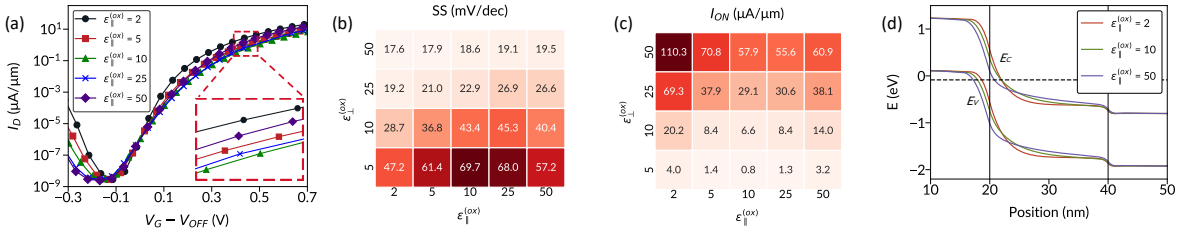

Figure 4.6: **(a)** I_D-V_G characteristics of TFETs while varying $\epsilon_{\parallel}^{(ox)}$ (with $\epsilon_{\perp}^{(ox)}$ = 10). Inset: a close-up of the area enclosed by the red dashed rectangle. Heat maps showing **(b)** SS and **(c)** I_{on} as a function of $\epsilon_{\perp}^{(ox)}$ and $\epsilon_{\parallel}^{(ox)}$ for TFETs. **(d)** Band profiles for TFETs with various $\epsilon_{\parallel}^{(ox)}$ (with $\epsilon_{\perp}^{(ox)}$ = 10) at $V_G \approx V_{ON}$. The source/channel and channel/drain interfaces are marked with vertical lines, and the horizontal dashed line shows the energy at which maximum tunnelling occurs [35]. © 2021 IEEE.

of the tunnel barrier. Let us consider the potential at the source/channel interface, $\psi_{S/C}$, the potential slightly to the right of the tunnel barrier, ψ_{TB-R}, and the potential of channel/drain interface, $\psi_{C/D}$. $\psi_{S/C}$ and ψ_{TB-R} are coupled together laterally through both the semiconductor and gate insulator. To simplify the analysis, I lump these lateral capacitances into a single capacitor, $C_{\parallel}^{L}$. ψ_{TB-R} and $\psi_{C/D}$ are coupled together laterally through similar capacitors, which I lump into $C_{\parallel}^{R}$. This allows us to establish the simplified capacitor model based on the device architecture shown in Fig. 4.7(a) and overlain with the band profiles in Fig. 4.7(b), which shows the physical locations of relevant potentials and capacitors in these devices (I discuss this new architecture and the plotted band profiles in detail later). In this model, $C_{S}^{(\parallel)}$ is the capacitance between the source and the source/channel interface, $C_{D}^{(\parallel)}$ is the capacitance between the drain and drain/channel interface, and $C_{\perp}^{TB-R}$ is the capacitance between the point to the right of the tunnelling barrier and the gate. Applying a capacitive voltage divider to this model, we find:

$$\psi_{S/C} = \frac{C_{S}^{(\parallel)} V_S + C_{\parallel}^{L} \psi_{TB-R}}{C_{S}^{(\parallel)} + C_{\parallel}^{L}}, \tag{4.1}$$

$$\psi_{TB-R} = \frac{C_{\parallel}^{L} \psi_{S/C} + C_{\perp}^{TB-R} V_G + C_{\parallel}^{R} \psi_{C/D}}{C_{\parallel}^{L} + C_{\perp}^{TB-R} + C_{\parallel}^{R}}. \tag{4.2}$$

From these equations:
(i) $\psi_{S/C}$ approaches V_S as $C_{\parallel}^{L}$ decreases,
(ii) $\psi_{S/C}$ and ψ_{TB-R} approach one another as $C_{\parallel}^{L}$ increases,
(iii) ψ_{TB-R} approaches $\psi_{C/D}$ as $C_{\parallel}^{R}$ increases.

To minimize the width of the tunnelling barrier, $\psi_{S/C}$ and ψ_{TB-R} should differ as much as possible. This will create a sharp drop in the potential energy near the tunnelling barrier that causes the conduction band to approach the valence band, thus minimizing the width of the tunnelling barrier [20]. This can be achieved by having $\psi_{S/C}$ approach V_S and having ψ_{TB-R} approach $\psi_{C/D}$. Therefore, using this design principle and the limits noted in (i) – (iii), $C_{\parallel}^{L}$ should be minimized and $C_{\parallel}^{R}$ should be maximized to minimize the width of the tunnelling barrier. However, $C_{\parallel}^{L}$ and $C_{\parallel}^{R}$ both contain capacitors which act laterally within the insulator, and will hence both increase as $\epsilon_{\parallel}^{(ox)}$ increases. This suggests that a competition between $C_{\parallel}^{L}$ and $C_{\parallel}^{R}$ could be the origin of the non-monotonic behaviour with respect to $\epsilon_{\parallel}^{(ox)}$ throughout Fig. 4.6. This would imply that the non-monotonic trends described above can be decomposed into two opposing monotonic trends: From (i) and (ii), decreasing $C_{\parallel}^{L}$ by decreasing $\epsilon_{\parallel}^{(ox)}$ on the left side of the device should cause $\psi_{S/C}$ to approach V_S while increasing the difference between $\psi_{S/C}$ and ψ_{TB-R}, thereby decreasing tunnelling distance and increasing I_{on}. Increasing $C_{\parallel}^{R}$ by increasing $\epsilon_{\parallel}^{(ox)}$ on the right side of the device should cause ψ_{TB-R} to approach $\psi_{C/D}$, which would decrease tunnelling distance and increase I_{on}.

To verify this discussion, I used a hetero-gate dielectric architecture where the gate insulator was broken into two portions so that $\epsilon_{\parallel}^{(ox\text{-}L)}$ and $\epsilon_{\parallel}^{(ox\text{-}R)}$ could be varied individually, as illustrated in Fig. 4.7(a). The left portion had length L = 5 nm [as labelled in Fig. 4.7(a)] and in-plane/out-of-plane permittivities of $\epsilon_{\parallel}^{(ox\text{-}L)}$

and $\epsilon_{\perp}^{(\text{ox-L})}$. The right portion had length $L_{ch} - L = 15$ nm and in-plane/out-of-plane permittivities of $\epsilon_{\parallel}^{(\text{ox-R})}$ and $\epsilon_{\perp}^{(\text{ox-R})}$. As shown in Fig. 4.7(b), decreasing $\epsilon_{\parallel}^{(\text{ox-L})}$ (while holding $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\parallel}^{(\text{ox-R})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$) causes the potential energy at the source/channel interface to approach the potential energy of the source [as predicted in (i)], while the potential energy to the right of the tunnel barrier is pushed farther away from that of the source/channel interface [as predicted in (ii)]. Both of these effects cause the width of the tunnelling barrier to decrease monotonically as $C_{\parallel}^{L}$ decreases. Increasing $\epsilon_{\parallel}^{(\text{ox-R})}$ (while holding $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\parallel}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$) does not affect the potential energy of the source/channel interface but causes the potential energy to the right of this interface to approach the potential energy of the drain [as predicted in (iii)], making the overall length of the tunnelling barrier decrease monotonically, as shown in Fig. 4.7(c) [note that the band profiles in Fig. 4.7(b) and Fig. 4.7(c) were measured at a common V_G such that $V_G \approx V_{ON}$]. This further suggests that the overall non-monotonic of varying $\epsilon_{\parallel}^{(\text{ox})}$ can be decomposed into two opposite monotonic trends that arise due to competitions between lateral capacitances in the left and right sides of the gate insulator.

If L is very large, then the left portion of the gate insulator (with $\epsilon_{\parallel}^{(\text{ox})} = \epsilon_{\parallel}^{(\text{ox-L})}$) will extend far laterally into the channel past ψ_{TB-R}. Since $\epsilon_{\parallel}^{(\text{ox-L})} < \epsilon_{\parallel}^{(\text{ox-R})}$, this will reduce $C_{\parallel}^{R}$, thereby reducing I_{on} for the reasons discussed above [see point (iii)]. Similarly, if L is very small, then the right portion of the gate insulator (with $\epsilon_{\parallel}^{(\text{ox})} = \epsilon_{\parallel}^{(\text{ox-R})}$) will impinge upon the region of the channel to the left of ψ_{TB-R}. This will increase $C_{\parallel}^{L}$, which will also decrease I_{on} for the reasons discussed above [see points (i) and (ii)]. Consequently, the ideal value of L that maximizes I_{on} should lie between these two extremes. I found this value of L for a model device with $\epsilon_{\parallel}^{(\text{ox-L})} = 2$, $\epsilon_{\parallel}^{(\text{ox-R})} = 10$, and $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$ by plotting I_{on} as a function of L, as shown in Fig. 4.7(d). Based on these results, the optimal value is $L = 2.2$ nm, while using a longer or shorter value of L decreases I_{on} for the reasons just discussed. However, note that I_{on} once again begins to increase as L increases beyond 10.4 nm. To understand why this is the case, I have also plotted V_{OFF} as a function of L on the same axis in Fig. 4.7(d). V_{OFF} decreases continuously until this same value of $L = 10.4$ nm, after which V_{OFF} increases as L increases. This increase in V_{OFF} is important to note because I_{on} is measured at $V_G = V_{ON} = V_{OFF} + 0.7$ V. Therefore, this increased V_{OFF} meant that the corresponding I_{on} values were measured at higher values of V_G, which is why I_{on} increases beyond $L = 10.4$ nm. Nevertheless, I_{on} is still clearly maximized at $L = 2.2$ nm, indicating that this is the ideal L to use for this device.

The trend for V_{OFF} in Fig. 4.7(d) can be understood on a similar basis as the trend for I_{on}. A high V_{OFF} implies that there is a small leakage current in the OFF-state, as this indicates that a higher V_G is needed to achieve a specified I_{off}. Just as before, when L is very large, then $C_{\parallel}^{R}$ is lowered. This leads to a smaller leakage current and hence a higher V_{OFF}. Likewise, if L is very small then $C_{\parallel}^{L}$ will increase. This also decreases leakage current and leads to a higher V_{OFF}. $L = 10.4$ nm therefore represents the point at which the leakage current is maximized in the OFF-state, which is why V_{OFF} is minimized. Note that because the TFET is in the OFF-state, the right edge of the tunnel barrier extends farther into the channel laterally, and hence ψ_{TB-R} is pushed closer to the drain (as ψ_{TB-R} is defined to be slightly to the right of the tunnel barrier). This is why this minimum V_{OFF} occurs at a relatively large value of L.

Using the optimized device with $L = 2.2$ nm, I then varied $\epsilon_{\parallel}^{(\text{ox-L})}$ (with $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\parallel}^{(\text{ox-R})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$) and $\epsilon_{\parallel}^{(\text{ox-R})}$ (with $\epsilon_{\perp}^{(\text{ox-R})} = \epsilon_{\parallel}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-L})} = 10$), with I_D-V_G characteristics plotted in Fig. 4.7(e) and Fig. 4.7(f). These results confirm that I_{on} indeed becomes monotonic with respect to $\epsilon_{\parallel}^{(\text{ox-L})}$ and $\epsilon_{\parallel}^{(\text{ox-R})}$ for this optimized device, which

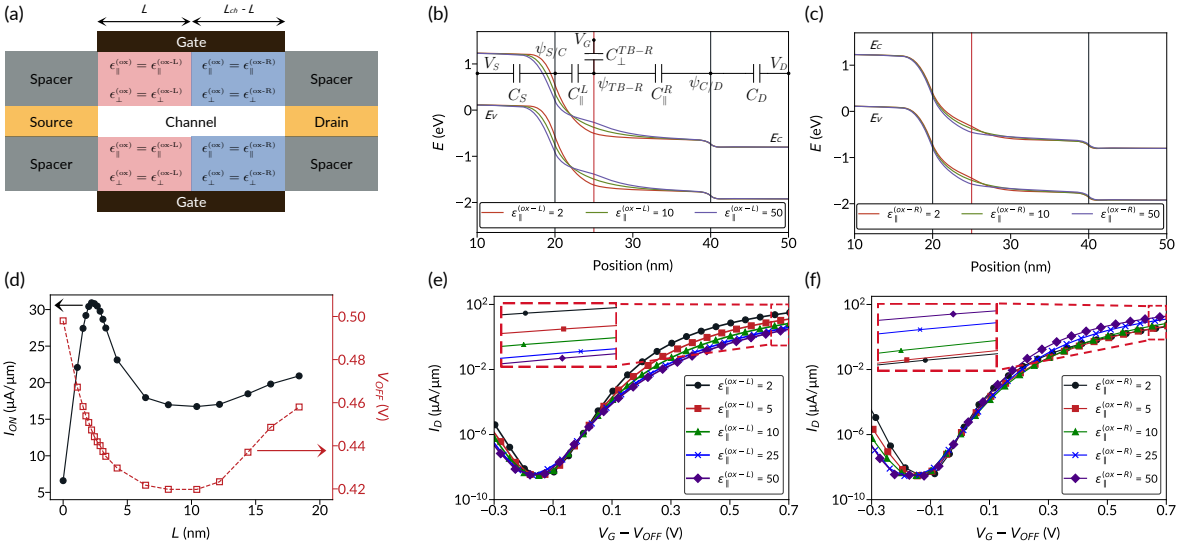

Figure 4.7: **(a)** Hetero-gate dielectric architecture that allows for the permittivities of the left and right portions of the gate insulator to be varied separately. **(b)** Conduction and valence band profiles for various $\epsilon_{\parallel}^{(\text{ox-L})}$ ($V_G \approx V_{ON}$, $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\parallel}^{(\text{ox-R})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$). The simplified capacitor model used in the discussion to analyse the potential energy profiles is shown at the top of the plot to illustrate the physical locations of relevant potentials and capacitors. The black vertical lines mark the locations of the source/channel and channel/drain interfaces, and the red vertical line marks the location in the insulator where the permittivities changes from $\epsilon_{\perp}^{(\text{ox-L})}/\epsilon_{\parallel}^{(\text{ox-L})}$ to $\epsilon_{\perp}^{(\text{ox-R})}/\epsilon_{\parallel}^{(\text{ox-R})}$. For simplicity, we assume here that ψ_{TB-R} is located near the interface between the left and right portions of the gate insulator, though this is not true in general for all values of L. **(c)** The same, for various $\epsilon_{\parallel}^{(\text{ox-R})}$ (while holding $\epsilon_{\perp}^{(\text{ox-R})} = \epsilon_{\parallel}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-L})} = 10$). **(d)** I_{on} (black circles/solid lines) and V_{OFF} (red squares/dashed lines) as functions of L for a device with $\epsilon_{\parallel}^{(\text{ox-L})} = 2$, $\epsilon_{\parallel}^{(\text{ox-R})} = 10$, and $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$. I_D-V_G characteristics for devices with $L = 2.2$ nm for **(e)** various $\epsilon_{\parallel}^{(\text{ox-L})}$ (while holding $\epsilon_{\perp}^{(\text{ox-L})} = \epsilon_{\parallel}^{(\text{ox-R})} = \epsilon_{\perp}^{(\text{ox-R})} = 10$) and **(f)** various $\epsilon_{\parallel}^{(\text{ox-R})}$ (while holding $\epsilon_{\perp}^{(\text{ox-R})} = \epsilon_{\parallel}^{(\text{ox-L})} = \epsilon_{\perp}^{(\text{ox-L})} = 10$). The insets in (e) and (f) show close-ups of the areas enclosed by the red dashed rectangles [35]. © 2021 IEEE.

offers additional evidence that the non-monotonic trends observed when varying $\epsilon_{\parallel}^{(\text{ox})}$ can be decomposed into two competing monotonic trends. I measured the intrinsic delay of a TFET that used an isotropic gate insulator with $\epsilon_{\parallel}^{(\text{ox})} = \epsilon_{\perp}^{(\text{ox})} = 10$ to be $\tau = 307.1$ fs, which improves to 89.8 fs for the optimized structure in Fig. 4.7(e) with $\epsilon_{\parallel}^{(\text{ox-L})} = 2$, and 83.3 fs for the optimized structure in Fig. 4.7(f) with $\epsilon_{\parallel}^{(\text{ox-R})} = 50$. These results indicate that implementing the hetero-gate dielectric architecture proposed in Fig. 4.7(a) with either a low $\epsilon_{\parallel}^{(\text{ox-L})}$ or a high $\epsilon_{\parallel}^{(\text{ox-R})}$ could also enhance the circuit-level performance of TFETs compared to a TFET with a uniform isotropic gate insulator. Future works may benefit from a thorough circuit-level analysis on this hetero-gate dielectric architecture, including studies to see how this architecture can be optimized in terms of circuit-level metrics.

Past reports on TFETs with hetero-gate dielectics have used high-κ and low-κ dielectrics near the source/channel and channel/drain interfaces, respectively, to induce high electric fields that enhance tunnelling [16, 55]. Meanwhile, the hetero-gate dielectric used in this chapter used insulators with low in-plane and high in-plane permittivities, respectively, at these same interfaces, while considering a fixed out-of-plane permittivity. This comparison suggests that while the gate insulator's in-plane permittivity plays

an important role in hetero-gate dielectric-based TFETs, it is overshadowed by the role of the out-of-plane permittivity when using isotropic insulators. Further work on engineering TFETs with hetero-gate dielectrics could focus on studying how the in-plane and out-of-plane permittivities can be manipulated separately to achieve devices with truly optimized electrostatics.

Finally, while the work documented in this chapter studied how dielectric anisotropy within a TFET's gate insulator affects device performance, other works have also investigated how the electrostatics of TFETs can be optimized by engineering their spacers, and have introduced unique architectures featuring symmetric dual-κ [56] and asymmetric dual-κ spacers [57] to accomplish this. While a detailed analysis of how the spacers' anisotropic permittivities affect TFET performance is beyond the scope of this thesis, such an analysis may be an interesting topic for future works. In particular, given the asymmetric nature of TFETs highlighted in previous articles as well as this one, it may be necessary for such studies to examine the role of anisotropic permittivities in the source-side and drain-side spacers separately.

4.3 Conclusions

MOSFETs can be improved by using anisotropic insulators. Both the gate insulator and spacers should have low in-plane permittivities to minimize fringing effects and high out-of-plane permittivities to maximize gate control. A single global insulator can be used as both the gate insulator and spacers to provide the benefits of a high-κ gate oxide with low-κ spacers, while simultaneously reducing the number of features that need to be patterned.

A TFET's SS and I_{on} are non-monotonic with respect to the insulator's in-plane permittivity, with best performance obtained when the in-plane permittivity is very small. A TFET's performance can be optimized by using a hetero-gate dielectric where the left portion of the gate dielectric has a low in-plane permittivity and the right portion has a high in-plane permittivity. This could be combined with other recent developments in oxide engineering to further increase the performance of TFETs by allowing the position of the conduction/valence band edges throughout the device to be engineered more precisely and in new ways.

As FETs with anisotropic insulators have never before been fabricated (with the exception of hBN), a crucial next step will be to establish experimental protocols to develop FETs with anisotropic insulators. As the study reported in this Chapter was the first of its kind, I explored FETs with ideal anisotropic insulators whose dielectric constants were perfectly uniform in each direction. When implemented in real devices, anisotropic insulators may not have perfect crystal structures and anisotropic nanocomposites could contain defects, both of which could affect their permittivities. Further research that analyses the extent to which these defects may impact device performance would therefore benefit future works on engineering the electrostatics of FETs.

Finally, while the work presented in this chapter considered FETs with 2-D semiconductors, devices that use thicker semiconductors will have more significant coupling through the semiconductor, which may

de-emphasize the effects of lateral coupling through the insulators. As fringing fields through MOSFETs' gate insulators have been shown to affect the performance of 50 nm-thick silicon-based MOSFETs with EOTs of 1 nm [58], the trends reported in this Chapter will likely be qualitatively similar in FETs with thicker semiconductors. Nevertheless, quantitative studies on the degree to which anisotropic insulators can improve FETs with thicker semiconductors will be an important step towards understanding the utility of anisotropic insulators.

5 Exploiting Fringing Fields Created by High-κ Gate Insulators

The work documented in this chapter was published in July 2021 and is © 2021 IEEE (R. K. A. Bennett and Y. Yoon, "Exploiting Fringing Fields Created by High-κ Gate Insulators to Enhance the Performance of Ultrascaled 2-D-Material-Based Transistors," *IEEE Transactions on Electron Devices*, July 2021, doi: 10.1109/TED.2021.3096178).

5.1 Introduction

As metal-oxide-semiconductor field-effect transistors' (MOSFETs') channel lengths are scaled down, their gate capacitances must be increased to ensure that the gate electrode maintains its electrostatic control over the channel. Although gate capacitance may be increased by decreasing the gate insulator's thickness, gate leakage will drastically interfere with device performance if the gate insulator is made too thin. Instead, device designers often use high-κ gate insulators to achieve high gate capacitances while maintaining insulator thicknesses sufficient to prevent gate leakage [59]. Unfortunately, replacing low-κ gate insulators with thicker high-κ insulators at identical equivalent oxide thicknesses (EOTs) strengthens lateral electric fields throughout MOSFETs. These lateral electric fields (also known as *fringing fields*) couple the source and drain more strongly to the center of the channel, thereby weakening the gate's control over the channel's electrostatic potential. This phenomenon, known as fringe-induced barrier lowering (FIBL), has been studied thoroughly over the last two decades and is a well-known drawback of using high-κ insulators [50, 60, 61, 62]. Consequently, when implementing high-κ insulators, device designers are forced to make a trade-off where a device's ideal source-to-drain transfer characteristics are sacrificed to suppress gate leakage [15].

In this Chapter, I will demonstrate that FIBL caused by implementing high-κ gate insulators does not need to be suppressed in ultra-scaled MOSFETs based on two-dimensional (2D) materials and can actually be exploited to improve the performance of devices that suffer from source-to-drain tunneling leakage in the OFF state. Specifically, I find that at extremely short channel lengths, FIBL caused by implementing high-κ gate insulators improves, rather than deteriorates, a MOSFET's ideal source-to-drain transfer characteristics by (a) reducing the ratio of the source-to-drain tunneling current to the total driving current and (b) making the source-to-drain tunneling current more responsive to changes in the applied gate bias. Both of these effects cooperatively decrease the device's overall subthreshold swing (SS) and boost ON currents (I_{on}'s) at fixed OFF currents (I_{off}'s) targeted by the *IEEE International Roadmap for Devices and Systems* (IRDS) for logic-core applications [63]. I investigate the physics behind the observed performance boost by studying the electrostatics of FIBL in ultra-scaled MOSFETs, and I conclude this Chapter by studying how FIBL impacts the performance of ultra-scaled MOSFETs based on three different 2D semiconductors. The results

presented in this Chapter demonstrate benefits of implementing high-κ gate insulators in MOSFETs that I anticipate will help device designers scale MOSFETs into the sub-10 nm regime by overcoming limitations presented by source-to-drain tunneling currents.

5.2 Methodology

MOSFETs were simulated using the non-equilibrium Green's function method as previously described in Chapter 2. MOSFETs followed the double-gated architecture shown in Fig. 5.1(a), where a monolayer of black phosphorous (BP) served as the semiconductor. The electronic structure of BP was described using previously reported tight-binding parameters [44]. BP's in-plane and out-of-plane permittivities were set to $4.56\epsilon_0$ and $1.36\epsilon_0$ (where ϵ_0 is the permittivity of free space) based on a previous study [24]. All transport was assumed to be ballistic and occurred along the armchair direction of BP. BP-based MOSFETs' sources and drains were 15 nm long and were n-doped to concentrations of 1.5×10^{13} cm^{-2}.

Here, I considered the following gate insulators: SiO_2 ($\epsilon_{ox} = 3.9\epsilon_0$), Al_2O_3 ($\epsilon_{ox} = 10\epsilon_0$), HfO_2($\epsilon_{ox} = 25\epsilon_0$), and TiO_2 ($\epsilon_{ox} = 85\epsilon_0$). I considered only the electrostatic effects of using different gate insulators by neglecting gate leakage in my simulations, which allowed me to directly measure a MOSFET's ideal source-to-drain transfer characteristics.

Simulations were performed at $V_D - V_S = 0.7$ V, where V_D and V_S are the drain and source voltages. The OFF state voltage V_{OFF} is the gate voltage V_G that yields a drain current I_D equal to a specified I_{off}. I measured a MOSFET's I_{on} at the ON state voltage $V_{ON} = V_{OFF} + V_{DD}$, where $V_{DD} = 0.7$ V is the power supply voltage.

Unless stated otherwise, all MOSFET simulations performed in this Chapter use monolayer BP as the semiconductor. However, at the end of this chapter, I also study MOSFETs that use GeSe-, HfS_2-, and GeH as the semiconductor. For these simulations, I used tight-binding-like Hamiltonians to capture the electronic structure of each semiconductor. To ensure numerical convergence for each semiconductor, I n-doped the sources and drains of GeSe-, HfS_2-, and GeH-based devices with donor concentrations of 1.25×10^{13}, 5×10^{13}, and 5×10^{12} cm^{-2}, respectively. Transport occurred along the zigzag direction for GeSe (as defined in [64]) and the $\Gamma \rightarrow$ X and $\Gamma \rightarrow$ M directions for HfS_2 and GeH, respectively. Permittivities of GeSe, HfS_2, and GeH were taken from [9], [19], and [65], respectively. Except for where otherwise noted, computational details for MOSFETs based on these semiconductors are identical to those of MOSFETs based on monolayer BP.

The tight-binding-like Hamiltonians for GeH and HfS_2 were identical to those reported in previous works [66, 1]. The tight-binding-like Hamiltonian for GeSe was computed based on the plane-wave approximation using density functional theory (DFT) with QUANTUM ESPRESSO software [67] using a rectangular supercell containing 4 Ge and 4 Se atoms. I used projector augmented wave pseudopotentials with a wave function cutoff of 50 Ry and a charge density cutoff of 400 Ry, and I inserted a vacuum layer

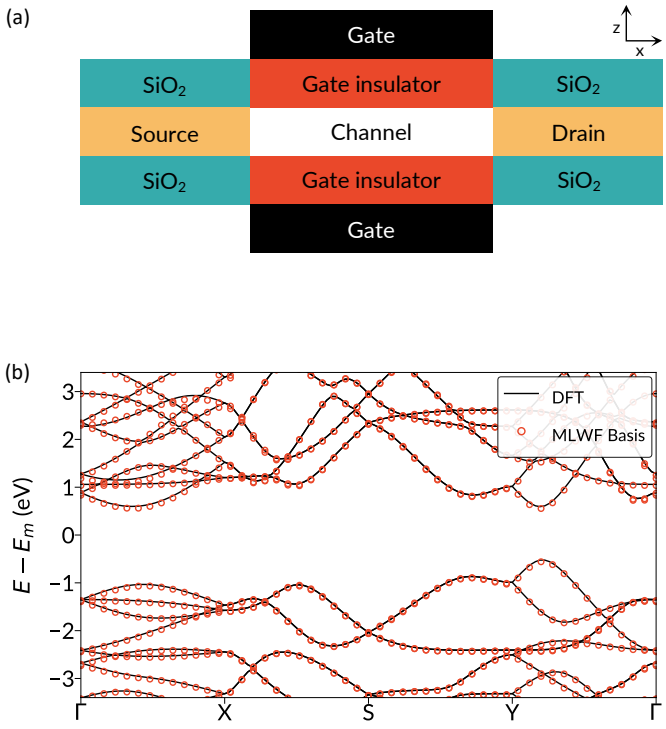

Figure 5.1: **(a)** Structure of MOSFETs. **(b)** Band structure of the GeSe supercell as computed by DFT and after conversion to an MLWF basis. Here, E_m is the mid-gap energy [69]. © 2021 IEEE.

of 20 Å to avoid interactions with periodic replicas. Afterwards, I used Wannier90 to transform the results from QUANTUM ESPRESSO into a maximally localized Wannier function (MLWF) basis [68]. I used a $21 \times 21 \times 1$ k-point grid for all computations with QUANTUM ESPRESSO and Wannier90. As shown in Fig. 5.1(b), the band structures from DFT and from the MLWF basis match one another well.

5.3 Results and Discussion

5.3.1 Performance Boost from FIBL

We begin by examining how the gate insulator's dielectric constant influences a MOSFET's I_{on} when the gate insulator's EOT is held constant by varying t_{ox}. I_{on} is plotted as a function of L_{ch} in Fig. Fig. 5.2 for both SiO_2- and HfO_2-based MOSFETs, where I_{off} is fixed at 10^{-4} µA/µm, which is the I_{off} targeted by the *IRDS* for low-power logic applications [63]. Here, the EOT is 1.25 nm for MOSFETs with both SiO_2 and HfO_2gate insulators at L_{ch} = 15 nm; as L_{ch} decreases, the EOT decreases such that the ratio L_{ch}/EOT remains constant, which ensures that MOSFETs have similar levels of electrostatic control across L_{ch}'s [70].

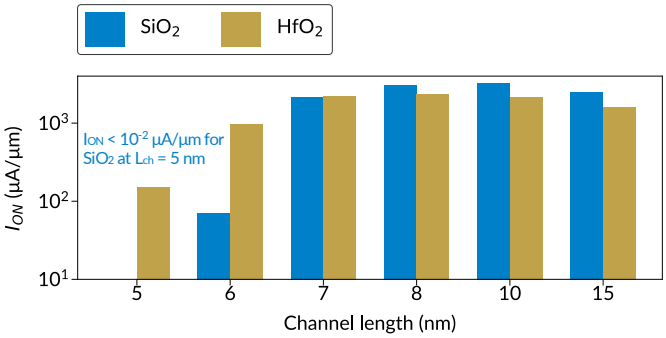

Figure 5.2: I_{on} at I_{off} = 10^{-4} μA/μm for MOSFETs with SiO_2 and HfO_2 gate oxides at various L_{ch}'s, where EOT is varied such that the ratio L_{ch}/EOT is held constant to ensure similar levels of electrostatic control across L_{ch}'s [69]. © 2021 IEEE.

At $L_{ch} \geq 8$ nm, I_{on}'s for MOSFETs with SiO_2 gate insulators exceed those of MOSFETs with HfO_2gate insulators. This result can be attributed to an increased SS for HfO_2-based devices resulting from FIBL and is consistent with previous works [15]. However, the performance boost offered by thin SiO_2 (relative to thick HfO_2at the same EOT) reverses at L_{ch} = 7 nm, at which point the I_{on} of a MOSFET with an HfO_2gate insulator is slightly greater than that of a MOSFET with an SiO_2 gate insulator. This disparity in I_{on} grows significantly at L_{ch} = 6 nm, with the HfO_2-based device offering an I_{on} more than an order of magnitude greater than that of the SiO_2-based device. At L_{ch} = 5 nm, the leakage current of the SiO_2-based MOSFET grows substantially and requires an extremely negative gate bias to enter the OFF state, resulting in an extremely low $I_{on} < 10^{-2}$ μA/μm, whereas the HfO_2-based device is still able to achieve $I_{on}/I_{off} > 10^6$ at this L_{ch}. Note that the EOT is the same for SiO_2- and HfO_2-based devices at each L_{ch} in Fig. 1(b). Therefore, the improved performance offered by high-κ insulators at $L_{ch} \leq 7$ nm cannot be explained based on changes to gate capacitance. Additionally, because these simulations neglect tunneling from the gate electrode, this result also cannot be attributed to changes in gate leakage.

I_D-V_G characteristics at $I_{off} = 10^{-4}$ μA/μm for MOSFETs with L_{ch} = 6 nm and various gate insulators (where EOT is held constant at 0.5 nm by varying t_{ox}) are plotted in Fig. 5.3(a), which shows that I_{on} is maximized and SS is minimized when $HfO_2$and TiO_2 are used as gate insulators. I_{on} is plotted as a function of I_{off} in Fig. 5.3(b), which shows that the improvement to I_{on} is most noticeable at low I_{off}'s, with $HfO_2$and TiO_2 offering an increase of nearly five orders of magnitude to I_{on} compared to SiO_2 at $I_{off} = 10^{-5}$ μA/μm. As I_{off} increases to 10^{-2} μA/μm, devices with different gate insulators converge to similar I_{on}'s, with TiO_2 offering the lowest I_{on}. At I_{off} = 10^{-4} μA/μm, I_{on} is more than an order of magnitude greater for devices that use HfO_2gate insulators compared to devices that use SiO_2. At this I_{off}, the *IRDS* targets I_{on}'s ranging from 861 to 1336 μA/μm for logic-core applications between 2020 and 2030 [63], which is within the range of I_{on}'s offered by devices with HfO_2gate insulators at L_{ch} = 6 nm.

As previously noted, the use of high-κ insulators appears to suppress leakage currents in the OFF state (I will examine this reduction of leakage current in greater detail in Chapter 5.3.2). This observation is also reflected

in Fig. 2(c), which shows V_{OFF} plotted as a function of L_{ch} for HfO_2-and SiO_2-based devices. V_{OFF} decays quickly as L_{ch} decreases for SiO_2-based devices, whereas it decreases more gradually when $HfO_2$is used as the gate insulator instead. This result suggests that using high-κ gate insulators may also reduce OFF state voltage variations in short-channel devices caused by fabrication errors resulting in non-uniform channel lengths, even at channel lengths where FIBL still deteriorates the source-to-drain transfer characteristics (e.g. at L_{ch} = 8 nm).

5.3.2 Underlying Physics

Next, I investigate the physical reasons behind the benefits of FIBL for ultra-scaled devices by decomposing the I_D-V_G curves in Fig. 5.3(a) into their source-to-drain tunneling currents I_{tunnel} [Fig. 5.4(a)] and thermionic currents I_{therm} [Fig. 5.4(b)] following the approach used in [71]. These results show that the SS of the tunneling component [$SS_{tunnel} = \partial V_G/\partial(\log_{10} I_{tunnel})$] decreases when high-$\kappa$ insulators are used, whereas the SS of the thermionic component [$SS_{therm} = \partial V_G/\partial(\log_{10} I_{therm})$] increases. This increase to SS_{therm} agrees with previous studies on FIBL [60, 15]. However, at L_{ch} = 6 nm, $I_{tunnel} > I_{therm}$ for V_G's around V_{OFF}. When I_D is composed of both I_{therm} and I_{tunnel}, the total SS is given by [72]

$$SS = \left(\frac{r_t}{SS_{tunnel}} + \frac{1 - r_t}{SS_{therm}}\right)^{-1}, \tag{5.1}$$

where r_t is the ratio of tunneling current to total drain current, i.e. $r_t = I_{tunnel}/I_D$. From (5.1), SS decreases when r_t decreases (as $SS_{tunnel} > SS_{therm}$) and/or SS_{tunnel} decreases. As shown in Fig. 5.4(c), r_t is smaller for devices with high-κ gate insulators. Consequently, the decrease to both SS_{tunnel} and r_t [Figs. 5.4(a,c)]

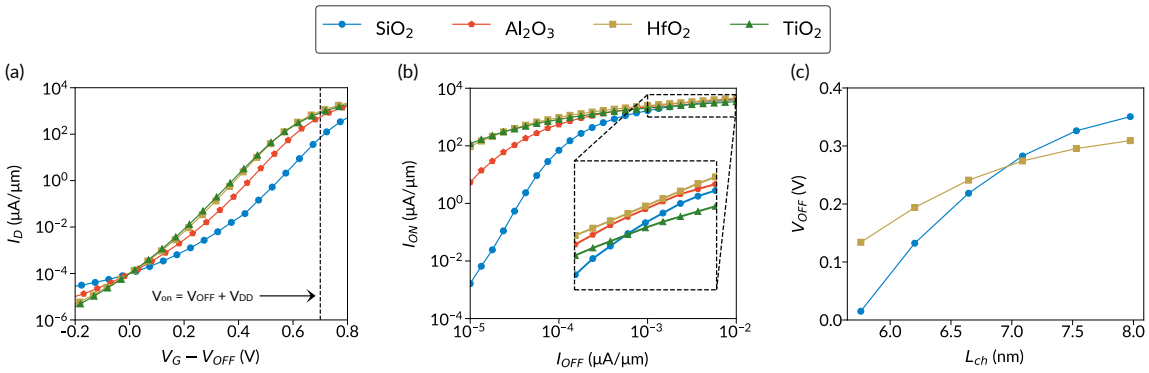

Figure 5.3: **(a)** I_D-V_G curves at I_{off} = 10^{-4} μA/μm and **(b)** I_{on} vs. I_{off} curves for MOSFETs with L_{ch} = 6 nm and various gate oxides, where oxide thickness is varied to maintain EOT = 0.5 nm. Inset: a close-up of the region encased in the dashed rectangle. **(c)** V_{OFF} as a function of L_{ch} for SiO_2-and HfO_2-based MOSFETs [69]. © 2021 IEEE.

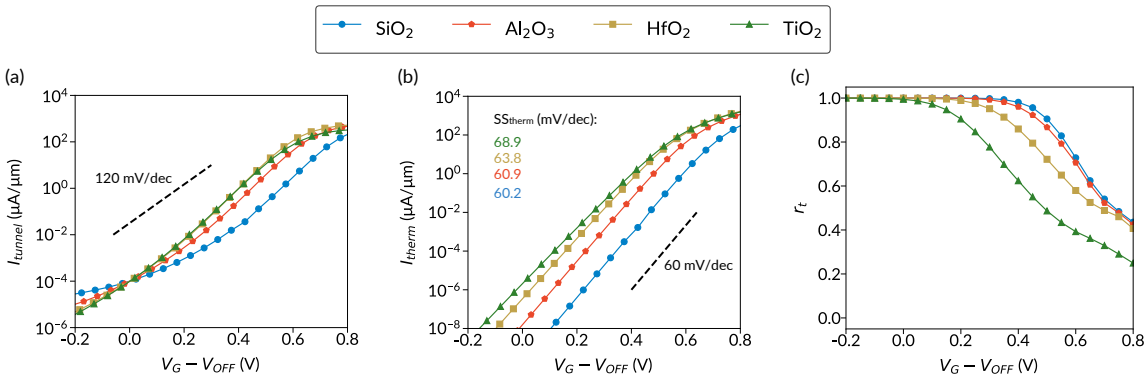

Figure 5.4: **(a)** Tunneling and **(b)** thermionic components of the I_D-V_G curves plotted in Fig. 5.3(a). **(c)** The ratio of tunneling current to total drain current r_t as a function of $V_G - V_{OFF}$ for MOSFETs with various gate oxides [69]. © 2021 IEEE.

offsets the slight increase to SS_{therm} [Fig. 5.4(b)], causing the overall SS of ultra-scaled MOSFETs to decrease when using high-κ gate insulators. Because I consider a fixed $I_{off} = 10^{-4}$ μA/μm, this decrease in SS translates directly to an increased I_{on}.

To understand why r_t decreases when using high-κ insulators, I have plotted the conduction band (E_C) profiles and current spectra of MOSFETs with SiO_2 and HfO_2gate insulators (L_{ch} = 6 nm, EOT = 0.5 nm) at $V_G - V_{OFF}$ = 0.25 V in Fig. 5.5(a). I_D is composed entirely of I_{tunnel} for the device with an SiO_2 gate insulator, whereas there is a noticeable I_{therm} component when an HfO_2gate insulator is used because the barrier height h_B is reduced due to FIBL [73]. However, note that the width of the potential energy barrier w_B (i.e. the lateral distance across the barrier, which I measure at E_C = 0 eV for consistency) is slightly greater for the device with an HfO_2gate insulator. $I_{tunnel} \propto \exp(-w_B\sqrt{h_B})$ [71], whereas $I_{therm} \propto \exp(\text{-}h_B)$ [2]. Based on these proportionalities, decreasing h_B while maintaining or increasing w_B causes I_{therm} to increase more quickly than I_{tunnel}, which is why r_t decreases as ϵ_{ox} increases at a fixed EOT.

Next, I consider the reason why using high-κ gate insulators decreases SS_{tunnel}. As $I_{tunnel} \propto \exp(-w_B\sqrt{h_B})$, a decrease to SS_{tunnel} implies that (i) $|\partial h_B/\partial V_G|$ increases (i.e. h_B becomes more responsive to V_G) and/or (ii) $|\partial w_B/\partial V_G|$ increases (i.e. w_B becomes more responsive to V_G). To investigate these possibilities, I have plotted E_C profiles for MOSFETs with SiO_2 and HfO_2gate insulators at $V_G - V_{OFF}$ = 0.25 V (solid lines) and $V_G - V_{OFF}$ = 0.35 V (dotted lines) in Fig. 5.5(b). h_B decreases by 0.1 and 0.09 eV across this ΔV_G for devices with SiO_2 and HfO_2gate insulators, respectively, indicating that h_B is slightly less responsive to changes in V_G when high-κ insulators are used, ruling out possibility (i). However, w_B changes by 0.245 nm and 0.424 nm across this same ΔV_G for MOSFETs with SiO_2 and HfO_2gate insulators, respectively [see also Fig. 5.5(b) inset], confirming that (ii) is the reason why using high-κ insulators decreases SS_{tunnel}.

To understand why increasing ϵ_{ox} increases w_B, I have plotted the potential energy contours around the channel/drain interface plotted in Fig. 5.5(c) for SiO_2 and HfO_2-based MOSFETs (L_{ch} = 6 nm,

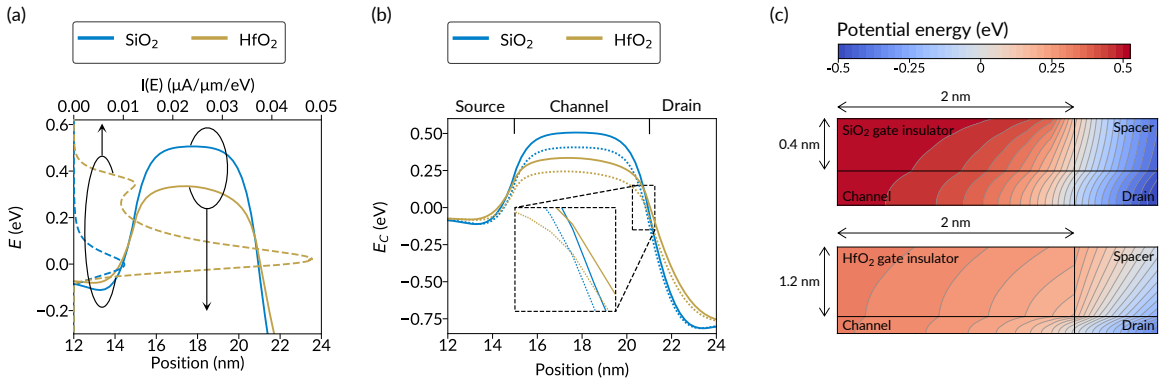

Figure 5.5: **(a)** E_C profiles (solid lines, bottom x axis) and current spectra $I(E)$ (dashed lines, top x axis) for MOSFETs with SiO_2 and HfO_2gate oxides at $V_G - V_{OFF}$ = 0.25 V. **(b)** E_C for MOSFETs with SiO_2 and HfO_2gate oxides (L_{ch} = 6 nm, EOT = 0.5 nm) at $V_G - V_{OFF}$ = 0.25 V (solid lines) and $V_G - V_{OFF}$ = 0.35 V (dotted lines). Inset: a close-up of the region in the dashed rectangle. **(c)** Contour plots of the potential energy near the channel/drain interface for an SiO_2-based MOSFET (top) and an HfO_2-based MOSFET (bottom), where L_{ch} = 6 nm, EOT = 0.5 nm, and $V_G - V_{OFF}$ = 0.25 V. Note that the reference potential energy in (c) is shifted so that the potential energy in the semiconductor aligns with with E_C, which facilitates comparisons between (b) and (c) [69]. © 2021 IEEE.

EOT = 0.5 nm) at $V_G - V_{OFF}$ = 0.25 V. Here, the potential energy contour lines near the channel/drain interface have a more significant horizontal component when $HfO_2$is used as a gate insulator compared to when SiO_2 is used as a gate insulator, which allows these contour lines to penetrate far laterally into both the channel and drain. This in turn lowers the potential energy in the semiconductor to the left of the channel/drain interface and raises the potential energy in the semiconductor to the right of the channel/drain interface. Both of these effects cooperatively increase w_B. These elongated potential energy contour lines in the HfO_2-based device are the origin of FIBL and are known to arise because of the strengthened lateral electrostatic coupling when using high-κ gate insulators at the same EOTs as their lower-κ counterparts [60, 15].

Next, to understand why $|\partial w_B/\partial V_G|$ increases as ϵ_{ox} increases, consider the slope of E_C, i.e. $|\partial E_C/\partial x|$, near the source/channel and channel/drain interfaces. From Fig. 4(b), the slope of E_C is steep near the source/channel and channel/drain interfaces for the SiO_2-based MOSFET, whereas this slope is much gentler for the HfO_2-based MOSFET. Note that when the slope of E_C is large at these interfaces, $|\partial w_B/\partial V_G|$ will naturally decrease, whereas $|\partial w_B/\partial V_G|$ increases when the slope of E_C near these interfaces decreases. Consequently, this difference in $|\partial E_C/\partial x|$ in these regions is directly responsible for the increase in $|\partial w_B/\partial V_G|$ observed when using high-κ insulators.

The physical reason why $|\partial E_C/\partial x|$ is larger near the source/channel and channel/drain interfaces when SiO_2 is used as a gate insulator can also be explained by the potential energy contours shown in Fig. 5.5(c). As the potential energy contour lines near the channel/drain interface do not extend far laterally into the SiO_2 gate insulator before reaching the gate insulator/channel interface, the potential energy in the channel even a short distance away from the channel/drain interface is strongly influenced by the gate electrode. Similarly,

because the potential energy contour lines near the channel/drain interface do not extend far laterally into the drain, the potential energy in the drain is minimally influenced by the gate electrode, allowing the potential energy in the drain to drop to its minimum value a short distance away from the channel/drain interface. As a result, E_C changes rapidly near the channel/drain interface for the SiO_2-based MOSFET, resulting in a large $|\partial E_C/\partial x|$. However, as these contour lines extend far laterally into both the channel and drain for the HfO_2-based MOSFET, E_C changes gradually in this region, thereby lowering $|\partial E_C/\partial x|$. Here, note that $\partial E_C/\partial x$ is directly proportional to the lateral component of the electric field in the semiconductor [$E_x^{(SC)} = (1/q)(\partial E_C/\partial x)$, where q is the elementary charge]. The above discussion is therefore consistent with previous works that have observed a decrease in $E_x^{(SC)}$ at the channel/drain interface as a by-product of FIBL when using MOSFETs with high-κ gate insulators (compared to MOSFETs with low-κ insulators at the same EOTs) [74, 75, 76].

Finally, lateral capacitances and electric fields are screened out as the number of charge carriers in the channel increases [41]. Therefore, the aforementioned effects that influence both w_B and $|\partial w_B/\partial V_G|$ become less important at higher I_{off}'s, which is why the performance boost from FIBL is most noticeable at low I_{off}'s [Fig. 5.3(b)].

5.3.3 Applicability to Other 2D-Material-Based MOSFETs

While the qualitative nature of the trends described at the end of Chapter 5.3.2 are not unique to BP-based MOSFETs, the quantitative nature of source-to-drain tunneling can vary drastically between different semiconductors and MOSFET configurations. Therefore, I conclude this Chapter by studying the degree to which FIBL caused by implementing high-κ gate insulators can enhance the performance of MOSFETs based upon three different monolayer semiconductors: germanium selenide (GeSe), hafnium disulfide (HfS_2), and germanane (GeH). I selected these three semiconductors for two main reasons: first, like BP, each of these materials offer strong performance when used as the semiconductor in MOSFETs [9, 66]. Second, as I shall verify shortly, each of these materials exhibit different degrees of source-to-drain tunneling when used in ultra-scaled field-effect transistors, making the results of this Chapter more applicable to the wide range of 2D semiconductors currently available. Because of GeSe's smaller band gap (E_g = 1.1 eV), I performed these simulations at a slightly reduced $V_D = V_{DD}$ = 0.6 V. Additional computational details for simulating MOSFETs made from these semiconductors are described in the Appendix.

I_D-V_G curves for GeSe-, HfS_2-, and GeH-based MOSFETs with SiO_2 and HfO_2gate insulators are plotted in Figs. 5.6(a–c), where I_{off} is fixed at 10^{-4} μA/μm. Here, I set L_{ch} = 6 nm for GeSe and HfS_2-based MOSFETs. However, I found that at L_{ch} = 6 nm, extensive source-to-drain tunneling prevented GeH-based MOSFETs with SiO_2 gate insulators from achieving $I_{off} = 10^{-4}$ μA/μm, making it impossible to compare I_{on}'s at a fixed I_{off}. Therefore, I instead used L_{ch} = 7 nm for GeH-based MOSFETs in Fig. 5.6(c). The EOTs of GeSe- and HfS_2- based MOSFETs were fixed at 0.5 nm, whereas GeH-based MOSFETs had EOTs of 0.583 nm to ensure that the ratio of L_{ch}/EOT was constant for every configuration in Fig. 5.6.

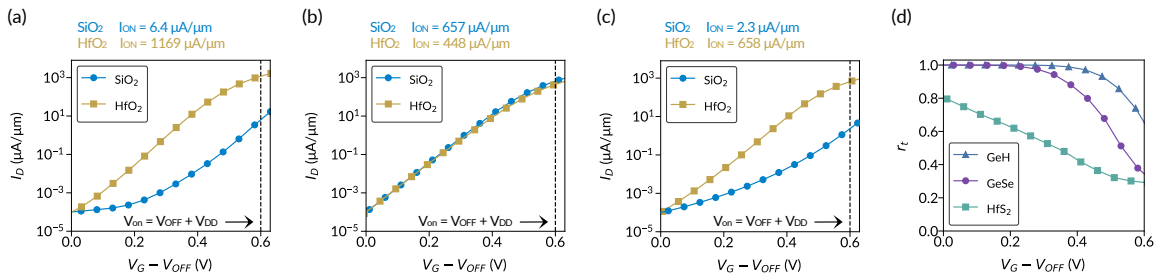

Figure 5.6: I_D-V_G curves for MOSFETs that use SiO_2 and HfO_2gate insulators and **(a)** GeSe, **(b)** HfS_2, and **(c)** GeH as the semiconductor. L_{ch} = 6 nm for (a) and (b) and 7 nm for (c), and EOT = 0.5 nm in (a) and (b) and 0.583 nm in (c). **(d)** r_t as a function of $V_G - V_{OFF}$ for MOSFETs where SiO_2 serves as the gate insulator and GeSe, HfS_2, and GeH serve as semiconductors. L_{ch}'s and EOTs are the same as in (a) – (c) for MOSFETs based upon each semiconductor [69]. © 2021 IEEE.

From these results, replacing an SiO_2 gate insulator with an HfO_2gate insulator at the same EOT boosts the I_{on}'s of GeSe- and GeH-based MOSFETs by more than 2 orders of magnitude, but slightly degrades the I_{on} of HfS_2-based MOSFETs. To understand these behaviors, I have plotted r_t for each semiconductor with SiO_2 gate insulators in Fig. 5.6(d). Here, $r_t \approx 1$ for GeSe- and GeH-based MOSFETs for much of the V_G window considered, indicating that tunneling dominates the total current when SiO_2 is used as a gate insulator. Consequently, FIBL improves the performance of these GeSe- and GeH-based MOSFETs for the same reasons that FIBL benefits monolayer BP-based devices with $L_{ch} \leq 7$ nm [as previously observed in Fig. 5.1(b)]. However, r_t is much lower for HfS_2-based devices with SiO_2 gate insulators, indicating that I_{tunnel} is less significant for the HfS_2-based MOSFETs considered here [relative to I_{tunnel} for the GeSe- and GeH-based MOSFETs in Figs. 5.6(a,c)]. As a result, the effect of decreasing SS_{tunnel} and I_{tunnel} becomes less significant [see the discussion surrounding (1)], which is why FIBL deteriorates, rather than improves, the performance of the HfS_2-based MOSFETs shown in Fig. 5(b).

The results presented in Fig. 5.6 confirm that FIBL can be exploited to improve the performance of MOSFETs made from a variety of semiconductors. However, the degree to which FIBL may improve an ultra-scaled MOSFET's performance depends strongly on the configuration and transfer characteristics of a nominal device. Therefore, while device designers should consider inducing FIBL to improve the ideal source-to-drain transfer characteristics of ultra-scaled MOSFETs, they should do so with caution to ensure that the performance enhancement from decreasing r_t and SS_{tunnel} is able to overcome the detrimental effect of increasing SS_{therm}.

5.4 Conclusions

FIBL enhances the performance of ultra-scaled MOSFETs by increasing the potential energy barrier's width while decreasing its height (which increases the ratio of thermionic current to total current) and by making the barrier's width more responsive to changes in gate voltages (which decreases the SS of the source-to-drain

tunneling current). Therefore, as 2D-material-based MOSFETs are scaled into the sub-10 nm regime, I anticipate that device designers will be able to implement high-κ gate insulators while enhancing, rather than deteriorating, a MOSFET's ideal source-to-drain transfer characteristics even before considering the effect of gate leakage. However, I have also observed instances where FIBL deteriorates the performance of sub-10 nm devices. Therefore, device designers should be cautious when intentionally introducing FIBL to short-channel devices to determine whether or not they are operating in the regime where FIBL will improve, rather than deteriorate, their nominal device's performance.

In this Chapter, I considered only MOSFETs based upon monolayer semiconductors, although MOSFETs based upon few-layer semiconductors are also a topic of interest in current research. As FIBL has been observed in MOSFETs based on bulk semiconductors [58], I anticipate that the performance boost offered by FIBL reported in this Chapter could likewise be used to improve the performance of ultra-scaled MOSFETs based on few-layer semiconductors. However, FIBL becomes less significant as the semiconductor's thickness increases [77], which in turn may also diminish the performance boost reported in this Chapter. I therefore recommend that future works investigate how the performance boost offered by FIBL in ultra-scaled MOSFETs may change with the number of layers in a MOSFET's 2D semiconductor.

Recent studies have used hexagonal boron nitride or other low-κ insulators as intermediate "buffer layers" between 2D semiconductors and high-κ insulators. These buffer layers improve gate control by weakening the lateral electric field near the semiconductor/gate insulator interface (thereby reducing FIBL), while the high-κ layer ensures that the entire gate insulator is sufficiently thick in order to prevent gate leakage [58, 15]. However, the results of this Chapter suggest that implementing these "gate-stack architectures" may worsen source-to-drain tunneling currents in ultra-scaled MOSFETs. Consequently, future studies may focus upon optimizing gate-stack architectures in ultra-scaled MOSFETs by simultaneously considering source-to-drain tunneling, gate leakage, and thermionic emission. Additionally, these low-κ buffer layers have also previously been used to reduce the number of interfacial defects between 2D semiconductors and high-κ insulators [78, 79]. As this approach may be unideal for ultra-scaled devices, I anticipate that the results presented in this Chapter may further motivate work on developing high-κ insulators that are natively compatible with 2D semiconductors, a topic that has been receiving increased attention in recent years [80, 81, 82].

6 Conclusions and Future Work

In this thesis, I carried out three studies which provide new insights towards the electrostatic operation of nanoscale FETs and offer novel design strategies for optimizing these devices. In my first study, I demonstrated that device designers should generally avoid implementing the isotropic approximation in their simulations, especially at short channel lengths and/or when the MOSFET's EOT is extremely thin, as the role of the out-of-plane permittivity becomes more important under these conditions. In my second study, I demonstrated that replacing MOSFETs' isotropic gate insulators with anisotropic insulators with small in-plane and large out-of-plane permittivities can improve devices' ON-currents and lower their SS's, and used this to demonstrate a novel device architecture that uses a singular anisotropic insulator to span the entirety of the source, channel, and drain. I also demonstrated that the performance of TFETs responds non-monotonically to the gate insulator's in-plane permittivity, with highest ON-currents/smallest SS's obtained when the insulator's in-plane permittivity is either extremely small or extremely large. This non-monotonic behaviour arose because of competitions between coupling on the source and drain side of the device, which I took advantage of to propose a novel hetero-gate dielectric for TFETs. Finally, in my third study, I demonstrated that FIBL can be exploited to improve the performance of MOSFETs with extremely short channel lengths by lowering the subthreshold swing and quantity of tunneling current around the OFF-state, resulting in a dramatic improvement to the ON-currents of ultra-scaled devices.

Future work may continue along this direction to develop a deeper understanding of the electrostatic principles and design strategies I have presented in this work. Below, I propose several specific research topics to further develop the research topics explored in this thesis.

- Developing a natural length scale to estimate the electrostatic integrity of MOSFETs that considers a semiconductor's and/or insulator's anisotropic permittivities
- Exploring the role of the semiconductor's out-of-plane permittivity in few-layer devices, where vertical coupling may become more significant
- Experimentally developing MOSFETs and/or TFETs that take advantage of the novel architectures proposed in Chapter 4
- Exploring how dielectric anisotropy in the semiconductor and/or insulator affects the performance of nanoscale FETs whose electrostatic operation differs from that of conventional MOSFETs/TFETs, e.g. ferroelectric and negative capacitance FETs
- Optimizing gate insulators of ultra-scaled devices based on source-to-drain from FIBL *and* gate leakage currents simultaneously

Letter of Copyright Permission

The studies presented in Chapters 3, 4, and 5 were previously published in *IEEE* journals. The content in these chapters is reused in accordance with *IEEE*'s copyright guidelines, which require the following statement be made in my thesis:

In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of the University of Waterloo's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink. If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

References

[1] R. K. A. Bennett, D. Yin, and Y. Yoon, "Assessing the role of a semiconductor's anisotropic permittivity in hafnium disulfide monolayer field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 67, pp. 2607–2613, June 2020. doi: 10.1109/ted.2020.2985023.

[2] M. S. Lundstrom, *Fundamentals of Nanotransistors (Lessons from Nanoscience: A Lecture Notes)*. WSPC, jul 2017.

[3] H. Liu, S. Datta, and V. Narayanan, "Steep switching tunnel FET: A promise to extend the energy efficient roadmap for post-CMOS digital and analog/RF applications," in *International Symposium on Low Power Electronics and Design (ISLPED)*, IEEE, Sept. 2013. doi: 10.1109/islped.2013.6629285.

[4] X. P. A. Gao, G. Zheng, and C. M. Lieber, "Subthreshold regime has the optimal sensitivity for nanowire FET biosensors," *Nano Letters*, vol. 10, pp. 547–552, Nov. 2009. doi: 10.1021/nl9034219.

[5] R.-H. Yan, A. Ourmazd, and K. Lee, "Scaling the si MOSFET: from bulk to SOI to bulk," *IEEE Transactions on Electron Devices*, vol. 39, pp. 1704–1710, July 1992. DOI: 10.1109/16.141237.

[6] H. Liu, A. T. Neal, and P. D. Ye, "Channel length scaling of MoS2 MOSFETs," *ACS Nano*, vol. 6, pp. 8563–8569, Sept. 2012. DOI: 10.1021/nn303513c.

[7] J. Wang and M. Lundstrom, "Does source-to-drain tunneling limit the ultimate scaling of MOSFETs?," in *Digest. International Electron Devices Meeting,*, IEEE. doi: 10.1109/iedm.2002.1175936.

[8] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically ThinMoS2: A new direct-gap semiconductor," *Physical Review Letters*, vol. 105, Sept. 2010. doi: 10.1103/physrevlett.105.136805.

[9] C. Klinkert, Á. Szabó, C. Stieger, D. Campi, N. Marzari, and M. Luisier, "2-d materials for ultrascaled field-effect transistors: One hundred candidates under the ab initio microscope," *ACS Nano*, vol. 14, pp. 8605–8615, June 2020. doi: 10.1021/acsnano.0c02983.

[10] S.-H. Chen, S.-W. Lian, T. R. Wu, T.-R. Chang, J.-M. Liou, D. D. Lu, K.-H. Kao, N.-Y. Chen, W.-J. Lee, and J.-H. Tsai, "Impact of semiconductor permittivity reduction on electrical characteristics of nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 66, pp. 2509–2512, June 2019. DOI: 10.1109/ted.2019.2912058.

[11] A. K. A. Lu, G. Pourtois, M. Luisier, I. P. Radu, and M. Houssa, "On the electrostatic control achieved in transistors based on multilayered MoS2: A first-principles study," *Journal of Applied Physics*, vol. 121, p. 044505, Jan. 2017. DOI: 10.1063/1.4974960.

[12] A. E. Maniadaki, G. Kopidakis, and I. N. Remediakis, "Strain engineering of electronic properties of transition metal dichalcogenide monolayers," *Solid State Communications*, vol. 227, pp. 33–39, Feb.

2016. doi: 10.1016/j.ssc.2015.11.017.

[13] B. Amin, T. P. Kaloni, and U. Schwingenschlögl, "Strain engineering of WS2, WSe2, and WTe2," *RSC Advances*, vol. 4, p. 34561, Aug. 2014. doi: 10.1039/c4ra06378c.

[14] Y. Y. Illarionov, T. Knobloch, M. Jech, M. Lanza, D. Akinwande, M. I. Vexler, T. Mueller, M. C. Lemme, G. Fiori, F. Schwierz, and T. Grasser, "Insulators for 2d nanoelectronics: the gap to bridge," *Nature Communications*, vol. 11, July 2020. doi: 10.1038/s41467-020-16640-8.

[15] M. Salmani-Jelodar, H. Ilatikhameneh, S. Kim, K. Ng, P. Sarangapani, and G. Klimeck, "Optimum high-k oxide for the best performance of ultra-scaled double-gate MOSFETs," *IEEE Transactions on Nanotechnology*, vol. 15, pp. 904–910, Nov. 2016. doi: 10.1109/tnano.2016.2583411.

[16] W. Y. Choi and W. Lee, "Hetero-gate-dielectric tunneling field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 57, pp. 2317–2319, Sept. 2010. doi: 10.1109/ted.2010.2052167.

[17] J. Nagel, *Solving the Generalized Poisson Equation Using theFinite-Difference Method (FDM)*, ECE 6340 lecture notes, feb. 2012. [Accessed 01 May 2021] Available: https://my.ece.utah.edu/~ece6340/LECTURES/Feb1/Nagel%202012%20-%20Solving%20the%20Generalized%20Poisson%20Equation%20using%20FDM.pdf.

[18] Z. Ren, *NANOSCALE MOSFETS: PHYSICS, SIMULATION AND DESIGN*. PhD Thesis, *Purdue University,* 2001.

[19] A. Laturia, M. L. V. de Put, and W. G. Vandenberghe, "Dielectric properties of hexagonal boron nitride and transition metal dichalcogenides: from monolayer to bulk," *npj 2D Materials and Applications*, vol. 2, Mar. 2018. DOI: 10.1038/s41699-018-0050-x.

[20] H. Ilatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, and J. Appenzeller, "Tunnel field-effect transistors in 2-d transition metal dichalcogenide materials," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 1, pp. 12–18, Dec. 2015. DOI: 10.1109/jxcdc.2015.2423096.

[21] R. K. Ghosh and S. Mahapatra, "Monolayer transition metal dichalcogenide channel-based tunnel transistor," *IEEE Journal of the Electron Devices Society*, vol. 1, pp. 175–180, Oct. 2013. DOI: 10.1109/jeds.2013.2292799.

[22] T. A. Ameen, H. Ilatikhameneh, G. Klimeck, and R. Rahman, "Few-layer phosphorene: An ideal 2d material for tunnel transistors," *Scientific Reports*, vol. 6, June 2016. DOI: 10.1038/srep28515.

[23] M. Brahma, A. Kabiraj, D. Saha, and S. Mahapatra, "Scalability assessment of group-IV mono-chalcogenide based tunnel FET," *Scientific Reports*, vol. 8, Apr. 2018. DOI: 10.1038/s41598-018-24209-1.

[24] F. W. Chen, H. Ilatikhameneh, T. A. Ameen, G. Klimeck, and R. Rahman, "Thickness engineered tunnel field-effect transistors based on phosphorene," *IEEE Electron Device Letters*, vol. 38, pp. 130–133, Jan. 2017. DOI: 10.1109/led.2016.2627538.

[25] P. Singh and B. P. Pandey, "Performance analysis of n-channel MoS2 FET with variation of oxide thickness and channel length," in *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, IEEE, Nov. 2018. DOI: 10.1109/upcon.2018.8597143.

[26] S. P. Kumar, P. Sandeep, and S. Choudhary, "Changes in transconductance(gm) and ion/ioff with high-k dielectrics in MX2 monolayer 10 nm channel double gate n-MOSFET," *Superlattices and Microstructures*, vol. 111, pp. 642–648, Nov. 2017. DOI: 10.1016/j.spmi.2017.07.021.

[27] I. Huh, S. Park, M. Shin, and W. Y. Choi, "An accurate drain current model of monolayer transition-metal dichalcogenide tunnel FETs," *IEEE Transactions on Electron Devices*, vol. 64, pp. 3502–3507, Aug. 2017. DOI: 10.1109/ted.2017.2716339.

[28] T. Finge, F. Riederer, M. R. Mueller, T. Grap, K. Kallis, and J. Knoch, "Investigations on field-effect transistors based on two-dimensional materials," *Annalen der Physik*, vol. 529, p. 1700087, July 2017. DOI: 10.1002/andp.201700087.

[29] S. Luo, K.-T. Lam, B. Wang, C.-H. Hsu, W. Huang, L.-Z. Yao, A. Bansil, H. Lin, and G. Liang, "Effects of contact placement and intra/interlayer interaction in current distribution of black phosphorus sub-10-nm FET," *IEEE Transactions on Electron Devices*, vol. 64, pp. 579–586, Feb. 2017. DOI: 10.1109/ted.2016.2635690.

[30] A. Mukhopadhyay, S. Bhattacharya, P. S. Gupta, L. Banerjee, A. Sengupta, and H. Rahaman, "A hybrid atomistic - semi-analytical modeling on schottky barrier au-MoS2-au MOSFETs," in *2018 IEEE Electron Devices Kolkata Conference (EDKCON)*, IEEE, Nov. 2018. DOI: 10.1109/edkcon.2018.8770224.

[31] J. Lin, D. A. Antoniadis, and J. A. del Alamo, "Impact of intrinsic channel scaling on InGaAs quantum-well MOSFETs," *IEEE Transactions on Electron Devices*, vol. 62, pp. 3470–3476, Nov. 2015. DOI: 10.1109/ted.2015.2444835.

[32] K. Suzuki and S. Pidin, "Short-channel single-gate soi mosfet model," *IEEE Transactions on Electron Devices*, vol. 50, pp. 1297–1305, May 2003. DOI: 10.1109/ted.2003.813450.

[33] A. Szabo and M. Luisier, "Under-the-barrier model: An extension of the top-of-the-barrier model to efficiently and accurately simulate ultrascaled nanowire transistors," *IEEE Transactions on Electron Devices*, vol. 60, pp. 2353–2360, July 2013. DOI: 10.1109/ted.2013.2263386.

[34] J.-P. Colinge, "Multiple-gate SOI MOSFETs," *Solid-State Electronics*, vol. 48, pp. 897–905, June 2004. DOI: 10.1016/j.sse.2003.12.020.

[35] R. K. A. Bennett and Y. Yoon, "Using anisotropic insulators to engineer the electrostatics of conventional and tunnel field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 68, pp. 865–872, Feb. 2021. doi: 10.1109/ted.2020.3044559.

[36] A. Singh, J. Hänisch, V. Matias, F. Ronning, N. Mara, D. Pohl, B. Rellinghaus, and D. Reagor, "Transforming insulating rutile single crystal into a fully ordered nanometer-thick transparent semiconductor," *Nanotechnology*, vol. 21, p. 415303, Sept. 2010. doi: 10.1088/0957-4484/21/41/415303.

[37] X. Zhao and D. Vanderbilt, "Phonons and lattice dielectric properties of zirconia," *Physical Review B*, vol. 65, Jan. 2002. doi: 10.1103/physrevb.65.075105.

[38] S. Coh, T. Heeg, J. H. Haeni, M. D. Biegalski, J. Lettieri, L. F. Edge, K. E. O'Brien, M. Bernhagen, P. Reiche, R. Uecker, S. Trolier-McKinstry, D. G. Schlom, and D. Vanderbilt, "Si-compatible candidates for high-k dielectrics with the *Pbnm* perovskite structure," *Physical Review B*, vol. 82, Aug. 2010. doi:10.1103/physrevb.82.064101.

[39] Z. N. Wing, J. W. Halloran, X. Gong, W. H. She, E. Hoppenjans, and W. J. Chappell, "Fabrication and properties of an anisotropic TiO2 dielectric composite," *Journal of the American Ceramic Society*, vol. 89, pp. 2812–2815, June 2006. doi: 10.1111/j.1551-2916.2006.01134.x.

[40] L. Zhang, Y. Wang, D. He, and Y. Deng, "Poly(vinylidene fluoride)-based nanocomposite employing oriented bi2s3 nanorods with double-shell structure for high dielectric performance and loss suppression," *Composites Science and Technology*, vol. 171, pp. 118–126, Feb. 2019. doi: 10.1016/j.compscitech.2018.12.020.

[41] S.-H. Kim, J. Fossum, and J.-W. Yang, "Modeling and significance of fringe capacitance in nonclassical CMOS devices with gate–source/drain underlap," *IEEE Transactions on Electron Devices*, vol. 53, pp. 2143–2150, Sept. 2006. doi: 10.1109/ted.2006.880369.

[42] D. Yin and Y. Yoon, "Design strategy of two-dimensional material field-effect transistors: Engineering the number of layers in phosphorene FETs," *Journal of Applied Physics*, vol. 119, p. 214312, June 2016. doi: 10.1063/1.4953256.

[43] D. Yin and Y. Yoon, "Can bilayer black phosphorus outperform monolayer in field-effect transistors?," in *2015 73rd Annual Device Research Conference (DRC)*, IEEE, June 2015. doi: 10.1109/drc.2015.7175614.

[44] A. N. Rudenko and M. I. Katsnelson, "Quasiparticle band structure and tight-binding model for single- and bilayer black phosphorus," *Physical Review B*, vol. 89, May 2014. doi: 10.1103/physrevb.89.201408.

[45] W. Zhu, S. Park, M. N. Yogeesh, K. M. McNicholas, S. R. Bank, and D. Akinwande, "Black phosphorus flexible thin film transistors at gighertz frequencies," *Nano Letters*, vol. 16, pp. 2301–2306, Mar. 2016. doi:10.1021/acs.nanolett.5b04768.

[46] D. Yin, A. AlMutairi, and Y. Yoon, "Assessment of high-frequency performance limit of black phosphorus field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 64, pp. 2984–2991, July 2017. doi:10.1109/ted.2017.2699969.

[47] M. Brahma, A. Kabiraj, and S. Mahapatra, "Insights on anisotropic dissipative quantum transport in n-type phosphorene MOSFET," in *2019 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID)*, IEEE, Jan. 2019. doi:10.1109/vlsid.2019.00050.

[48] X. Cao and J. Guo, "Simulation of phosphorene field-effect transistor at the scaling limit," *IEEE Transactions on Electron Devices*, vol. 62, pp. 659–665, Feb. 2015. doi:10.1109/ted.2014.2377632.

[49] S. O. Koswatta, M. S. Lundstrom, and D. E. Nikonov, "Performance comparison between p-i-n tunneling transistors and conventional MOSFETs," *IEEE Transactions on Electron Devices*, vol. 56, pp. 456–465, Mar. 2009. doi:10.1109/ted.2008.2011934.

[50] Q. Chen, L. Wang, and J. D. Meindl, "Fringe-induced barrier lowering (FIBL) included threshold voltage model for double-gate MOSFETs," *Solid-State Electronics*, vol. 49, pp. 271–274, Feb. 2005. doi:10.1016/j.sse.2004.08.008.

[51] E. J. Mele, "Screening of a point charge by an anisotropic medium: Anamorphoses in the method of images," *American Journal of Physics*, vol. 69, pp. 557–562, May 2001. doi: 10.1119/1.1341252.

[52] H.-Y. Deng, "Electrostatic responses of anisotropic dielectric films," *European Journal of Physics*, vol. 41, p. 035203, Apr. 2020. doi: 10.1088/1361-6404/ab806c.

[53] T. Mizuno, T. Kobori, Y. Saitoh, S. Sawada, and T. Tanaka, "Gate-fringing field effects on high performance in high dielectric LDD spacer MOSFETs," *IEEE Transactions on Electron Devices*, vol. 39, pp. 982–989, Apr. 1992. doi: 10.1109/16.127491.

[54] A. Pal, S. Natarajan, B. Ayyagari, S. Mittal, E. M. Bazizi, A. Sachid, M. Saremi, B. Colombeau, G. Thareja, S. Lin, and B. Alexander, "Impact of MOL/BEOL air-spacer on parasitic capacitance and circuit performance at 3 nm node," in *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, IEEE, Sept. 2019. doi: 10.1109/sispad.2019.8870410.

[55] J. Madan and R. Chaujar, "Gate drain-overlapped-asymmetric gate dielectric-GAA-TFET: a solution for suppressed ambipolarity and enhanced ON state behavior," *Applied Physics A*, vol. 122, Oct. 2016. doi: 10.1007/s00339-016-0510-0.

[56] H. G. Virani, R. B. R. Adari, and A. Kottantharayil, "Dual-k spacer device architecture for the improvement of performance of silicon n-channel tunnel FETs," *IEEE Transactions on Electron Devices*, vol. 57, pp. 2410–2417, Oct. 2010. doi:10.1109/ted.2010.2057195.

[57] M. A. Raushan, N. Alam, M. W. Akram, and M. J. Siddiqui, "Impact of asymmetric dual-k spacers on tunnel field effect transistors," *Journal of Computational Electronics*, vol. 17, pp. 756–765, Feb. 2018.

doi:10.1007/s10825-018-1129-5.

[58] B.-Y. Tsui and L.-F. Chin, "A comprehensive study on the FIBL of nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 51, pp. 1733–1735, Oct. 2004. doi:10.1109/ted.2004.835022.

[59] Q. Chen and J. D. Meindl, "Nanoscale metal–oxide–semiconductor field-effect transistors: scaling limits and opportunities," *Nanotechnology*, vol. 15, pp. S549–S555, July 2004. doi: 10.1088/0957-4484/15/10/009.

[60] E. Gnani, S. Reggiani, M. Rudan, and G. Baccarani, "Effects of high-κ hfo2 gate dielectrics in double-gate and cylindrical-nanowire FETs scaled to the ultimate technology nodes," *IEEE Transactions On Nanotechnology*, vol. 6, pp. 90–96, Jan. 2007. doi: 10.1109/tnano.2006.888547.

[61] M. Charmi, H. R. Mashayekhi, and A. A. Orouji, "The impact of high-k gate dielectric and FIBL on performance of nano DG-MOSFETs with underlapped source/drain regions," *Journal of Computational Electronics*, vol. 13, pp. 307–312, Nov. 2013. doi: 10.1007/s10825-013-0528-x.

[62] D. ming Ke, D. Wu, J. Meng, F. Yang, L. xu Wan, J. guo Yang, and H. Chang, "Subthreshold characteristics analysis and modeling of fully depleted silicon-on-insulator MOSFETs with high-k SiO2 stacked gate structure," *Japanese Journal of Applied Physics*, vol. 57, p. 094201, July 2018. doi: 10.7567/jjap.57.094201.

[63] IEEE (2020). "More Moore – IEEE International Roadmap for Devices and Systems (IRDS)". [Online white paper]. Available: https://irds.ieee.org/editions/2020. [Accessed: 01 October 2020].

[64] Y. Guo, F. Pan, G. Zhao, Y. Ren, B. Yao, H. Li, and J. Lu, "Sub-5 nm monolayer germanium selenide (GeSe) MOSFETs: towards a high performance and stable device," *Nanoscale*, vol. 12, no. 28, pp. 15443–15452, 2020. doi: 10.1039/d0nr02170a.

[65] R. K. Ghosh, M. Brahma, and S. Mahapatra, "Germanane: A low effective mass and high bandgap 2-d channel material for future FETs," *IEEE Transactions on Electron Devices*, vol. 61, pp. 2309–2315, July 2014. doi: 10.1109/ted.2014.2325136.

[66] A. AlMutairi, Y. Zhao, D. Yin, and Y. Yoon, "Performance limit projection of germanane field-effect transistors," *IEEE Electron Device Letters*, vol. 38, pp. 673–676, May 2017. doi: 10.1109/led.2017.2681579.

[67] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials," *Journal of Physics: Condensed Matter*, vol. 21, p. 395502, Sept. 2009. DOI: 10.1088/0953-8984/21/39/395502.

[68] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, "wannier90: A tool for obtaining maximally-localised wannier functions," *Computer Physics Communications*, vol. 178, pp. 685–699, May 2008. doi: 10.1016/j.cpc.2007.11.016.

[69] R. K. A. Bennett and Y. Yoon, "Exploiting fringing fields created by high-κ gate insulators to enhance the performance of ultrascaled 2-D-material-based transistors," *IEEE Transactions on Electron Devices*, jul 2021. doi: 10.1109/TED.2021.3096178 (in press).

[70] D. Frank, Y. Taur, and H.-S. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Letters*, vol. 19, pp. 385–387, Oct. 1998. doi: 10.1109/55.720194.

[71] J. Yang, R. Quhe, Q. Li, S. Liu, L. Xu, Y. Pan, H. Zhang, X. Zhang, J. Li, J. Yan, B. Shi, H. Pang, L. Xu, Z. Zhang, J. Lu, and J. Yang, "Sub 10 nm bilayer bi2o2se transistors," *Advanced Electronic Materials*, p. 1800720, Feb. 2019. doi: 10.1002/aelm.201800720.

[72] R. Quhe, Q. Li, Q. Zhang, Y. Wang, H. Zhang, J. Li, X. Zhang, D. Chen, K. Liu, Y. Ye, L. Dai, F. Pan, M. Lei, and J. Lu, "Simulations of quantum transport in sub-5-nm monolayer phosphorene transistors," *Physical Review Applied*, vol. 10, Aug. 2018. doi: 10.1103/physrevapplied.10.024022.

[73] Q. Chen, L. Wang, and J. D. Meindl, "Fringe-induced barrier lowering (FIBL) included threshold voltage model for double-gate MOSFETs," *Solid-State Electronics*, vol. 49, pp. 271–274, Feb. 2005. doi: 10.1016/j.sse.2004.08.008.

[74] X. Liu, S. Lou, Z. Xia, D. Guo, H. Zhu, J. Kang, and R. Han, "Characteristics of different structure sub-100nm MOSFETs with high-k gate dielectrics," in *2001 6th International Conference on Solid-State and Integrated Circuit Technology. Proceedings (Cat. No.01EX443)*, IEEE. doi: 10.1109/icsict.2001.981488.

[75] N. Mohapatra, M. Desai, and V. Rao, "Detailed analysis of FIBL in MOS transistors with high-k gate dielectrics," in *16th International Conference on VLSI Design, 2003. Proceedings.*, IEEE Comput. Soc. doi: 10.1109/icvd.2003.1183121.

[76] E. Jang, S. Shin, J. W. Jung, and K. R. Kim, "Gate induced drain leakage reduction with analysis of gate fringing field effect on high-κ metal gate CMOS technology," *Japanese Journal of Applied Physics*, vol. 54, p. 06FG10, May 2015. doi: 10.7567/jjap.54.06fg10.

[77] S. Tayal and A. Nandi, "Effect of FIBL in-conjunction with channel parameters on analog and RF FOM of FinFET," *Superlattices and Microstructures*, vol. 105, pp. 152–162, May 2017. doi: 10.1016/j.spmi.2017.03.018.

[78] B. Terres, F. H. L. Koppens, H. Agarwal, L. Orsini, A. Montanaro, V. Sorianello, D. van Thourhout, K. Watanabe, T. Taniguchi, and M. Romagnoli, "2d-3d integration of high-κ dielectric with 2d heterostructures for opto-electronic applications," in *2019 IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2019. doi: 10.1109/iedm19573.2019.8993546.

[79] W. Li, J. Zhou, S. Cai, Z. Yu, J. Zhang, N. Fang, T. Li, Y. Wu, T. Chen, X. Xie, H. Ma, K. Yan, N. Dai, X. Wu, H. Zhao, Z. Wang, D. He, L. Pan, Y. Shi, P. Wang, W. Chen, K. Nagashio, X. Duan, and X. Wang, "Uniform and ultrathin high-κ gate dielectrics for two-dimensional electronic devices," *Nature Electronics*, vol. 2, pp. 563–571, Dec. 2019. doi: 10.1038/s41928-019-0334-y.

[80] A. AlMutairi and Y. Yoon, "Device performance assessment of monolayer $HfSe_2$: A new layered material compatible with high-κ HfO_2," *IEEE Electron Device Letters*, vol. 39, pp. 1772–1775, Nov. 2018. doi: 10.1109/led.2018.2867957.

[81] Y. Y. Illarionov, A. G. Banshchikov, D. K. Polyushkin, S. Wachter, T. Knobloch, M. Thesberg, M. I. Vexler, M. Waltl, M. Lanza, N. S. Sokolov, T. Mueller, and T. Grasser, "Reliability of scalable mos2 fets with 2 nm crystalline caf2 insulators," *2D Materials*, vol. 6, p. 045004, July 2019. doi: 10.1088/2053-1583/ab28f2.

[82] T. Knobloch, Y. Y. Illarionov, F. Ducry, C. Schleich, S. Wachter, K. Watanabe, T. Taniguchi, T. Mueller, M. Waltl, M. Lanza, M. I. Vexler, M. Luisier, and T. Grasser, "The performance limits of hexagonal boron nitride as an insulator for scaled CMOS devices based on two-dimensional materials," *Nature Electronics*, vol. 4, pp. 98–108, Feb. 2021. doi: 10.1038/s41928-020-00529-x.