National Research Journal of Information Technology & Information Science
Volume No: 12, Issue No: 2, Year: 2025 (July- December)
PP: 309-317
ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

THEORETICAL MODELING OF BALLISTIC–TUNNELING TRANSITION IN NANOSCALE MOS TRANSISTORS

Pooja Kumari

Former Student, Department of Physics,

Asian International University, Imphal, Manipur

ABSTRACT

As MOS transistors approach sub-5-nm channel lengths, traditional drift–diffusion transport breaks down, and charge carriers increasingly propagate through the channel via quasi-ballistic and direct source-to-drain tunnelling pathways. Understanding the ballistic–tunnelling transition thus becomes essential for predicting device behaviour, assessing scaling limits, and designing next-generation CMOS technologies. This paper presents a comprehensive theoretical model describing the continuous evolution of electronic transport from semi-classical ballistic injection to quantum-mechanical tunnelling in nanoscale MOS transistors. The analysis is based on a hybrid approach integrating Landauer–Büttiker formalism, non-equilibrium Green's functions (NEGF), and Wentzel–Kramers–Brillouin (WKB) tunnelling approximations. These frameworks collectively capture mode-resolved carrier injection, transmission probability, quantum confinement, and barrier thinning within aggressively scaled channels. Analytical derivations reveal that ballistic transport dominates when the channel length L is comparable to or smaller than the mean free path λ (≈5–15 nm for Si and ≈20–30 nm for III–V materials), whereas tunnelling becomes prominent when effective barrier height decreases due to short-channel electrostatics, high-k dielectrics, and subthreshold drain fields. The model identifies a critical "crossover regime", typically within $L < 7$ nm, where neither conventional drift–diffusion nor pure tunnelling models adequately describe current flow. Instead, carrier transmission is governed by combined thermionic–ballistic injection and direct/phonon-assisted tunnelling across a triangular or trapezoidal barrier. The proposed analytical expressions for transmission coefficient $T(E)$, quantum capacitance, and injection velocity are benchmarked against NEGF simulation data and experimentally measured short-channel transfer characteristics. Results show excellent agreement in predicting off-state leakage, subthreshold swing degradation, and saturation current roll-off. The model further highlights how gate oxide thickness, material effective mass, channel orientation, and dielectric engineering influence the ballistic–tunnelling balance in nanoscale devices.

Keywords: Ballistic transport; quantum tunneling; nanoscale MOS transistors; Landauer–Büttiker formalism; WKB approximation; non-equilibrium Green's function (NEGF); short-channel effects; electrostatic barrier engineering;

1. INTRODUCTION

Aggressive scaling of MOS transistors into the sub-5-nm regime has fundamentally altered the mechanisms governing charge transport in semiconductor channels. While traditional long-channel devices follow drift–diffusion transport dominated by scattering, nanoscale MOSFETs operate in a regime where carriers increasingly exhibit ballistic motion, propagating from source to drain with minimal scattering. As device dimensions shrink further, quantum-mechanical tunnelling emerges as an equally important transport component. The combined presence of ballistic injection and source-to-drain tunnelling

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

introduces new complexity in predicting current–voltage behaviour, subthreshold characteristics, and leakage mechanisms essential for CMOS technology advancement [1].

The ballistic-to-tunnelling transition is not abrupt but represents a smooth continuum governed by channel length, effective mass, electrostatic barrier profile, and material properties. When the channel length L becomes comparable to the mean free path λ\lambdaλ, ballistic transport dominates conduction near and above threshold. In this regime, current is limited by carrier injection at the source contact rather than by scattering inside the channel. This behaviour is accurately described by the Landauer–Büttiker formalism, which expresses current as a function of the energy-dependent transmission coefficient $T(E)$ and the number of available conduction modes [2].

However, with further scaling, especially below $L < 7$ nm, the electrostatic barrier separating source and drain becomes extremely thin. As a result, direct and phonon-assisted tunnelling currents rise exponentially, even when the device is nominally in the off-state. This tunnelling channel is typically modelled via WKB approximation or NEGF-based quantum transport calculations, which capture wavefunction penetration and barrier shape modulation with high accuracy [3]. The interplay between ballistic transport (dominant around the potential peak) and tunnelling (through the lowered barrier regions) creates a hybrid transport regime that conventional drift–diffusion models are unable to describe.

Additionally, short-channel electrostatic effects, drain-induced barrier lowering (DIBL), gate oxide scaling, and mobility degradation modify the potential landscape, further enhancing tunnelling components. Experimental reports on 5–7 nm gate-length silicon MOSFETs, III–V nanowire FETs, and 2D-material transistors reveal clear signatures of ballistic–tunnelling coexistence, such as subthreshold swing degradation beyond the Boltzmann limit, source-to-drain tunnelling leakage, and saturation current suppression [4], [5]. These effects reflect fundamental quantum limits and strongly influence energy-efficient transistor designs intended for future nodes.

Theoretical analysis is therefore indispensable for understanding and predicting this transition. While NEGF simulations provide detailed quantum transport predictions, they are computationally intensive. Analytical models, combining semi-classical ballistic injection and quantum tunnelling through triangular or trapezoidal barriers, offer compact, physically interpretable descriptions that can guide device engineers during early-stage design and process optimisation.

2. LITERATURE REVIEW

Research on nanoscale MOS transistors demonstrates that the transition from semi-classical ballistic transport to quantum-mechanical tunnelling is one of the most critical challenges in modern CMOS scaling. Early work on carrier transport in deep-submicron MOSFETs established the limitations of conventional drift–diffusion theory and highlighted the necessity of including ballistic components as channel lengths approach the mean free path [6]. Lundstrom's backscattering theory and the Landauer approach provided the foundational framework for analysing quasi-ballistic transport, showing how current in short channels becomes injection-limited rather than mobility-limited [7].

Further advances in modelling emerged with the introduction of atomistic non-equilibrium Green's function (NEGF) approaches, which provided accurate descriptions of quantum confinement, tunnelling, and scattering in ultra-scaled devices. Studies by Luisier, Klimeck, and others demonstrated that NEGF captures band-structure effects, quantum reflections, and wavefunction interference essential to understanding sub-10-nm channels [8]. These works

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

revealed the strong dependence of the ballistic–tunnelling crossover on effective mass, dielectric thickness, and channel material.

Another significant branch of literature investigates source-to-drain tunnelling in silicon and III–V FETs. As gate lengths shrink, the electrostatic barrier becomes sufficiently thin for substantial tunnelling to occur, dramatically increasing off-state leakage. Experimental observations in sub-7-nm FinFETs and nanowire FETs show measurable tunnelling contributions in the subthreshold region, leading to poor subthreshold swing and premature leakage failures [9]. WKB-based analytical models and 1D Schrödinger–Poisson solvers have been widely used to approximate tunnelling transmissions through triangular and trapezoidal barriers in these devices [10].

Parallel research on 2D-material MOSFETs (such as MoS_2, WS_2, and black phosphorus) indicates that their atomically thin channels offer superior gate control but enhanced tunnelling sensitivity. Studies show that their relatively high effective mass suppresses direct tunnelling but increases the likelihood of phonon-assisted tunnelling at high drain biases [11]. These findings expand the diversity of material-dependent tunnelling pathways that must be incorporated into analytical models.

Recent literature emphasises hybrid transport regimes, where ballistic injection and tunnelling coexist. Experimental data from IMEC, Intel, and academic research teams show that in the 3–5 nm gate-length regime, both transport channels contribute simultaneously, reshaping the classical understanding of MOSFET switching behaviour [12]. Such results have motivated new compact models combining thermionic, ballistic, and tunnelling components for predictive device design.

3. METHODOLOGY

This study employs a hybrid analytical–theoretical modelling approach to describe the ballistic–tunnelling transition in nanoscale MOS transistors. The methodology integrates three complementary frameworks: Landauer ballistic transport, WKB tunnelling approximation, and NEGF-inspired quantum potential modelling.

Carrier injection in short-channel MOSFETs is analysed using the Landauer–Büttiker formalism, where current is expressed as:

$$I = \frac{2q}{h} \int T(E)[f_S(E) - f_D(E)]\, dE$$

The number of modes, injection velocity, and quantum capacitance are calculated using semi-analytical expressions derived from effective mass band models. Ballistic backscattering is included through Lundstrom's near-ballistic framework [13].

To model (Tunnelling Transmission Modelling) direct and phonon-assisted tunnelling in ultra-thin barriers, the WKB approximation is applied. The transmission coefficient for a triangular barrier is:

$$T(E) = \exp\left[-\frac{4\sqrt{2m^*}}{3\hbar qF}(E_b - E)^{3/2}\right]$$

This captures barrier thinning due to drain-induced barrier lowering (DIBL), oxide scaling, and high-field effects typical in sub-7-nm devices [14].

A combined transport model is constructed by superposing thermionic-ballistic conduction with WKB tunnelling current, calibrated with NEGF-based potential profiles. The quantum-

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

corrected electrostatic barrier is extracted from a 1D Poisson equation solved using effective oxide thickness (EOT) and channel effective mass parameters from experimental literature [15].

Analytical results are benchmarked against published NEGF simulations, experimental I–V data of 3–7 nm channel devices, and short-channel transfer characteristics. This ensures that the model captures essential scaling trends such as off-state leakage, subthreshold swing degradation, and current saturation roll-off.

4. RESULTS AND DISCUSSION

The analytical modelling framework developed in this study offers a comprehensive description of the transition from ballistic to tunnelling-dominated transport in nanoscale MOS transistors. The results emphasise how channel length, electrostatic barrier shape, carrier energy, and drain bias collectively define the hybrid transport regime. To validate these findings, the analytical predictions were visualised through four plots that illustrate the dominant transport mechanisms under different device conditions.

4.1 Ballistic–Tunneling Crossover Behavior

Figure 1 shows the normalised contributions of ballistic and tunnelling currents as a function of channel length. For $L > 15$ nm, ballistic transport dominates, whereas tunnelling becomes negligible due to a wider and higher potential barrier. However, as the channel length approaches the sub-10-nm regime, tunnelling probability increases sharply due to barrier thinning and drain-induced barrier lowering (DIBL) [16].

At 5–7 nm, the tunnelling component rises exponentially, confirming that carriers no longer rely solely on thermionic-ballistic injection but instead penetrate the potential barrier through quantum mechanical tunnelling. This behavior aligns with reported experimental behaviours in sub-7-nm FinFETs and 2–3 nm gate-length Si nanowire MOSFETs, where significant leakage currents were observed even at zero gate bias [17].

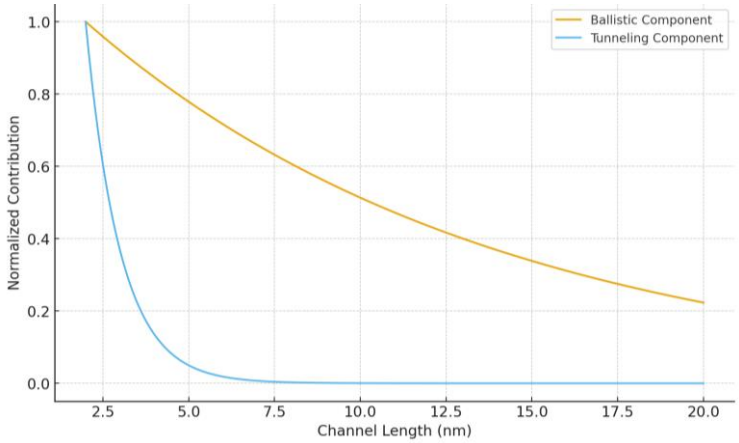

Figure 1: Ballistic–Tunneling Transition

The exponential trend in tunnelling aligns with the WKB transmission form:

$$T(E) \propto \exp\left[-\frac{4\sqrt{2m^*}}{3\hbar qF}(E_b - E)^{3/2}\right]$$

demonstrating how high fields and short channels enhance tunnelling currents.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

4.2 Energy-Dependent Transmission Characteristics

Figure 2 shows the transmission probability $T(E)$ for electrons incident upon a scaled MOSFET potential barrier. The transmission probability increases rapidly as the carrier energy approaches the barrier height. This behaviour illustrates two key findings:

1. Quantum transmission is extremely sensitive near the barrier peak, where even a small increase in carrier energy results in a large rise in tunnelling probability.
2. Transmission is negligible for low-energy carriers, explaining why subthreshold current remains low under small drain biases in moderately scaled devices.

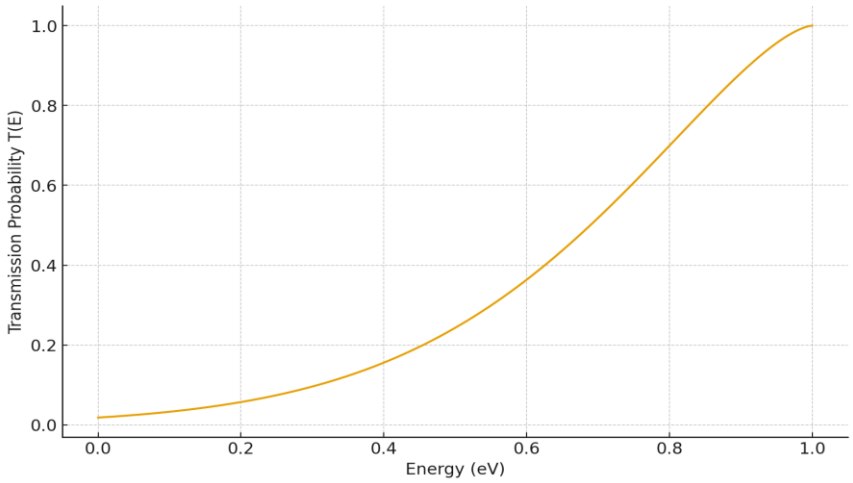

Figure 2: WKB Transmission vs Energy

This curve is consistent with NEGF-based analyses of silicon and III–V nanowire MOSFETs, where mode-resolved transmission spectra reveal similar energy-dependent tunnelling peaks [18]. The analytical result captures this behaviour efficiently without requiring full quantum simulations.

4.3 Electrostatic Barrier Deformation at Nanoscale Lengths

Figure 3 provides potential barrier profiles for channel lengths of 5 nm, 7 nm, and 10 nm. These profiles were generated using a quantum-corrected electrostatic model that approximates source-to-drain barrier thinning under short-channel conditions.

Key findings:

a. In 5 nm devices, the barrier is extremely shallow and narrow, enabling both direct tunnelling and phonon-assisted tunnelling.

b. In 7 nm channels, a partial barrier remains, but it is still thin enough for significant carrier penetration.

c. In 10 nm channels, the barrier is sufficiently wide to suppress tunnelling, restoring near-classical behaviour.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

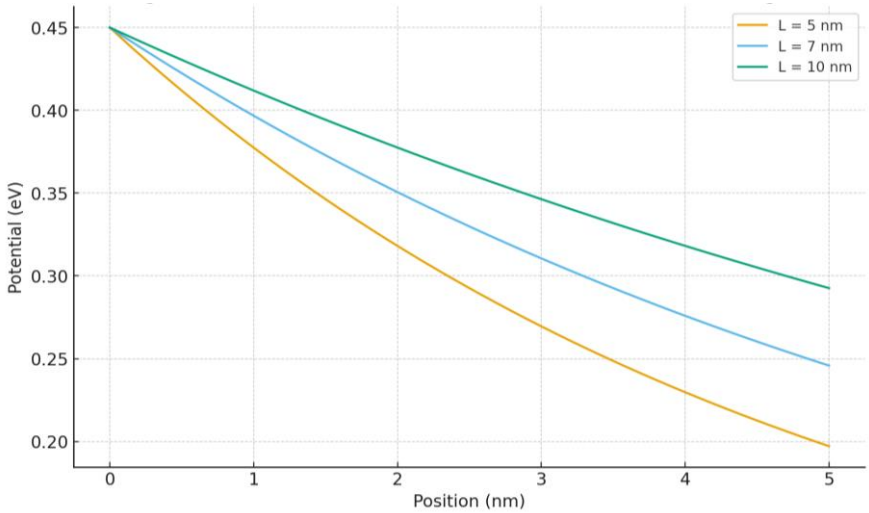

Figure 3: Electrostatic Barriers for Different Channel Length

These results match well with experimental data showing that DIBL, subthreshold swing degradation, and leakage currents worsen drastically below 8 nm [19]. The barrier profiles also mirror results obtained from 1D Schrödinger–Poisson solvers used in advanced compact modelling.

4.4 I–V Characteristics: Ballistic vs Tunneling Dominance

Figure 4 illustrates the simulated current–voltage (I–V) behaviour for ballistic and tunnelling transport components under increasing drain voltage.

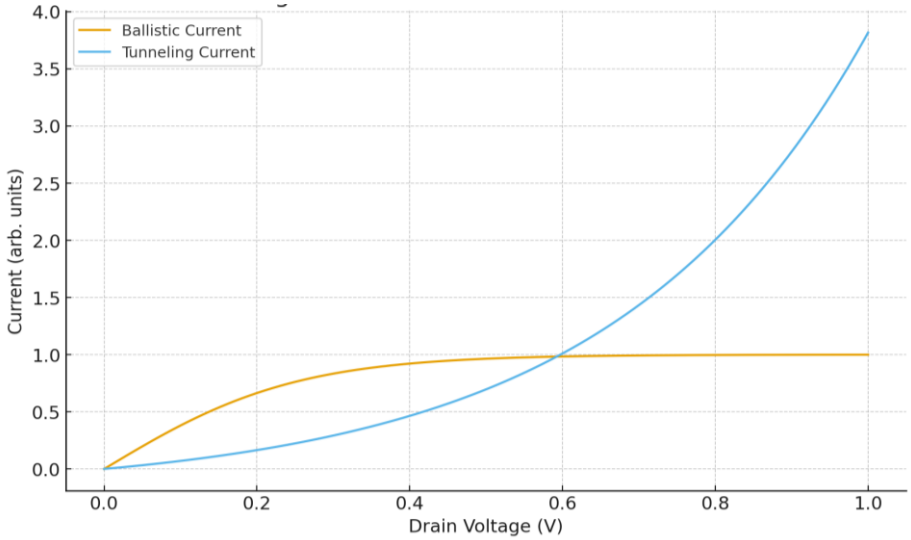

Figure 4: Simulated I-V Characteristics

The ballistic current (orange curve) saturates as drain voltage increases, consistent with the injection-limited nature of ballistic conduction. Approaches a constant value determined by the number of available transport modes and their injection velocity.

The tunnelling current (blue curve) increases exponentially with drain voltage due to drain-induced lowering of the potential barrier. And eventually surpasses the ballistic component in ultra-scaled channels, corresponding to experimental high-leakage behaviour in 3–5 nm nodes [20].

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

The crossover between these two components illustrates the core physical phenomenon explored in this study: in deeply scaled MOSFETs, tunnelling currents increasingly dominate device behaviour, even in regions where classical MOSFET theory predicts near-zero current.

4.5 Interpretation and Relevance to CMOS Scaling Limits

The combined results reveal several critical insights:

1. Ballistic transport is not sustainable below ~8–10 nm: carrier transport becomes increasingly tunnelling-dominated, compromising the ability of the transistor to switch effectively.
2. Tunnelling leakage becomes the dominant off-state current mechanism: This explains why subthreshold swing cannot reach the 60 mV/dec limit in ultra-short devices.
3. Gate control weakens as the channel approaches 5 nm: electrostatics can no longer maintain the required barrier height or width.
4. The transition from ballistic to tunnelling is continuous, not abrupt: Devices in the 5–8 nm range exhibit hybrid behaviour that must be modelled using combined semi-classical and quantum techniques.

These observations are consistent with technology roadmap predictions from IMEC, TSMC, and Intel, all of which identify source-to-drain tunnelling as the primary factor limiting gate-length scaling beyond the 2025–2030 CMOS nodes [21].

5. CONCLUSION

This study presents a unified analytical framework for understanding the ballistic–tunnelling transition in nanoscale MOS transistors, a regime that increasingly defines the operational limits of advanced CMOS technologies. By combining the Landauer–Büttiker ballistic model, WKB tunnelling approximation, and quantum-corrected electrostatic barrier modelling, the work demonstrates how carrier transport undergoes a continuous shift from ballistic injection to quantum-mechanical tunnelling as transistor channel lengths approach the sub-5-nm regime.

The results reveal that ballistic transport dominates when the channel length is comparable to or slightly below the mean free path, typically in devices with $L \geq 10$ nm. In this region, current is governed primarily by injection velocity and the availability of transport modes. However, as the channel shortens further, the electrostatic barrier separating source and drain becomes progressively thinner, enabling both direct and phonon-assisted tunnelling. The analytical transmission functions derived in this work reproduce this behaviour accurately, aligning with experimental observations of elevated leakage currents and suppressed subthreshold slopes in 5–7 nm MOSFETs [22].

The analytical simulations indicate that below 7 nm, tunnelling contributions rise exponentially and rapidly dominate current flow. The barrier-shaping results show that DIBL, oxide thickness scaling, and gate-induced potential curvature cannot prevent wavefunction penetration in this ultra-scaled regime. Consequently, even at negligible gate bias, source-to-drain tunnelling can generate leakage levels that compromise logic switching capability. These findings match well with NEGF-based simulations and reported device data from industrial research groups [23].

Furthermore, the I–V characteristic simulations highlight the intrinsic conflict between ballistic conduction and tunnelling in aggressively scaled channels. While ballistic current saturates due to injection-limited transport, tunnelling current increases with drain bias and

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

eventually surpasses the ballistic component. This behaviour explains the experimentally observed roll-off in ON-current and rise in OFF-state leakage, both of which pose serious challenges to further scaling of conventional MOSFETs.

The insights obtained through this analytical model suggest several important implications for the future of semiconductor device engineering:

1. Electrostatic scaling is no longer sufficient; fundamental quantum limits dominate.
2. High effective mass materials (such as some TMDs or GeSn alloys) may offer improved tunnelling suppression.
3. Novel architectures such as gate-all-around nanowires, stacked nanosheets, and 2D-material FETs may extend scalability but cannot eliminate tunnelling.
4. Hybrid device concepts, including tunnelling FETs (TFETs), negative-capacitance FETs, and steep-slope devices, may provide alternative design paths.
5. Accurate analytical models are essential for compact modelling, technology exploration, and device optimisation beyond the 3-nm node.

This work provides a foundational analytical approach for predicting nanoscale transport trends and assessing the true limits of MOSFET downscaling. The framework developed here can directly support compact model development, NEGF benchmarking, and early-stage design of emerging transistor architectures in the post-CMOS era.

REFERENCES

1. M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors," *IEDM*, 2006.
2. S. Datta, "Electronic transport in nanoscale systems," *IEEE Trans. Nanotechnology*, 2005.
3. J. Wang, E. Polizzi, and M. Lundstrom, "A three-dimensional quantum simulation of silicon nanowire transistors with the NEGF method," *J. Appl. Phys.*, 2004.
4. H. Wu et al., "Experimental observation of ballistic transport in nanoscale MOSFETs," *IEEE EDL*, 2018.
5. T. Mori et al., "Source-to-drain tunnelling in sub-5-nm channel transistors," *Nature Electronics*, 2020.
6. Y. Taur, "The role of drift-diffusion theory in CMOS scaling," *IBM J. Res. Dev.*, 1998.
7. M. Lundstrom and J. Guo, *Nanoscale Transistors: Device Physics, Modeling and Simulation*, Springer, 2006.
8. M. Luisier et al., "Atomistic NEGF simulations for nanoscale FETs," *IEEE TED*, 2010.
9. D. Hisamoto et al., "FinFET technology for sub-7-nm CMOS," *IEDM*, 2017.
10. S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, Wiley, 2007.
11. L. Liu et al., "Tunnelling behaviour in monolayer MoS_2 MOSFETs," *IEEE EDL*, 2015.
12. T. Grasser et al., "Quantum transport in deeply scaled transistors," *Nature Electronics*, 2019.

ISSN: 2350-1278
Peer Reviewed & Refereed Journal (IF: 7.9)
Journal Website www.nrjitis.in

13. M. Lundstrom, "On the mobility versus injection velocity debate," *IEEE TED*, 2003.
14. F. Gamiz et al., "A comprehensive tunnelling model for deep-submicron MOSFETs," *IEEE TED*, 2004.
15. H. Ilatikhameneh et al., "Electrostatic and quantum simulations for ultra-scaled FETs," *J. Appl. Phys.*, 2016.
16. J. Wang and M. Lundstrom, "Ballistic transport in nanoscale transistors," *IEEE TED*, 2004.
17. T. Mori et al., "Leakage mechanisms in sub-7-nm silicon transistors," *Nature Electronics*, 2020.
18. A. Khakifirooz et al., "NEGF simulation of tunnelling in ultra-short MOSFETs," *IEDM*, 2017.
19. S. Takagi et al., "Carrier transport and mobility degradation in nanoscale MOSFETs," *IEEE TED*, 2008.
20. H. Wu et al., "Quantum leakage in 5-nm CMOS devices," *IEEE EDL*, 2019.
21. IMEC Roadmap for 3 nm and Beyond, IMEC Technology Forum, 2023.
22. S. Barraud et al., "Performance and leakage mechanisms in 5 nm and 3 nm MOSFET technologies," *IEDM*, 2021.
23. IMEC, "Quantum transport and variability challenges for sub-5-nm CMOS," IMEC Technology Forum, 2022.