Thickness Scaling and Roughness-Induced Mobility Cliff in Ultra-Thin-Body MOSFETs

Mir Md Fahimul Islam, In Huh, Jian-Yu Lin, *Graduate Student Member, IEEE*, Seokyeon Shin, Youngjoon Lee, Changwook Jeong, *Member, IEEE*, Peide D. Ye, *Fellow, IEEE*, and Muhammad Ashraful Alam, *Fellow, IEEE*

***Abstract*— For future technology generations, Ultra-thin-body (UTB) MOSFETs are expected to replace classical bulk MOSFETs owing to their superior electrostatic scaling. However, process-control variability, especially under cryogenic operation, has heightened concerns regarding the channel/gate oxide interface. In this regime, interface-induced surface-roughness (SR) scattering becomes increasingly dominant, leading to a sharp roll-off in mobility. In this letter, we present (i) a universal thickness-scaling principle for long-channel mobility (μ_c), validated across multiple transistor technologies, that captures the non-ideal mobility roll-off exponent (n_{eff}) by explicitly accounting for finite potential-well type and confinement, and (ii) a closed-form analytical model that directly extracts the critical thickness (T_{crit}) and the associated critical-mobility trade-off. Overall, this compact and scalable approach quantitatively highlights the need for stringent thickness uniformity to preserve mobility control in next-generation complementary field-effect transistors (CFETs) technology.**

***Index Terms*— Universality, analytical model, mobility, ultra-thin semiconductors, surface roughness.**

I. Introduction

IN THE pursuit of next-generation transistors, the semiconductor industry is exploring a range of reduced-dimensional channel materials, including Si/SiGe nanosheet and forksheet transistors, MoS_2, WS_2, In_2O_3, Te, and others [1]. These technologies are highly diverse: channels are prepared using chemical vapor deposition (CVD), atomic layer deposition (ALD), layer transfer, or sputtering; they target applications across front-end, back-end, and display technologies; and they exhibit varying levels of technological maturity.

Received 10 March 2026; revised 6 April 2026; accepted 13 April 2026. Date of publication 28 April 2026; date of current version 25 May 2026. This work was supported by Samsung Electronics under Grant IO250514-12841-01 and Grant IO250311-12331-01. The review of this letter was arranged by Editor S. Cea. *(Corresponding author: Mir Md Fahimul Islam.)*

Mir Md Fahimul Islam, In Huh, Jian-Yu Lin, Peide D. Ye, and Muhammad Ashraful Alam are with Elmore Family School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: islam93@purdue.edu).

Seokyeon Shin, Youngjoon Lee, and Changwook Jeong are with the Graduate School of Semiconductor Materials and Devices Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, Republic of Korea.

Digital Object Identifier 10.1109/LED.2026.3685176

Despite this diversity, differences in process controllability across technologies lead to variability issues in modern transistors. For example, carrier mobility can become non-uniform due to surface roughness (SR), and unavoidable surface dangling bonds. Such non-idealities strongly scatter charge carriers and become increasingly severe as T_{ch} decreases, leading to a substantial degradation of carrier mobility (μ_c). Given this scenario, a key question arises: if the underlying physics is the same, can transistor on-state variability be understood within a non-dimensionalized [2] universal scaling framework? After all, off-state metrics such as the subthreshold slope are universal and serve as indicators of sub-bandgap defects [3].

In this letter, we demonstrate that (i) scaling theory can indeed extract the universal features of thickness-dependent long-channel mobility, and (ii) it provides a framework to calculate the critical channel thickness (T_{crit}), thereby guiding technology roadmap development using quantifiable metrics.

II. Experimental Study: μ–T_{CH} Characteristics

Fig. 1 plots the long-channel mobility (μ_c) as a function of channel thickness (T_{ch}) for multiple channel materials, including Si, Ge, Te, InAs, In_2O_3, and IGZO [4], [5], [6], [7], [8], [10], [11], [12]. When the channel thickness is scaled below a certain thickness, μ_c drops by orders of magnitude (i.e., a mobility cliff) across multiple MOSFET technologies. Variability around the mobility cliff would make robust design logic and memory ICs impossible. For SOI-Silicon transistors, the mobility cliff is described by the sixth-power thickness dependence, arising from SR-induced scattering [4] of an infinite quantum well, as described in (1):

$$\mu_{SR} \propto \left|\frac{\partial T_{\text{ch}}}{\partial E_0}\right|^2 \propto \frac{m_z^2}{m_x \Lambda^2 \Delta_{\text{rms}}^2} T_{\text{ch}}^6. \tag{1}$$

Since it is derived from an idealized infinite potential well, one wonders about the robustness of the exponent. Moreover, it is unclear whether the scaling hypothesis generalizes to non-Si MOSFETs.

III. Universality of $\mu(T_{\text{CH}})$ and T_{CRIT} Extraction

Total carrier mobility in a MOSFET is controlled by multiple (nominally) uncorrelated scattering mechanisms

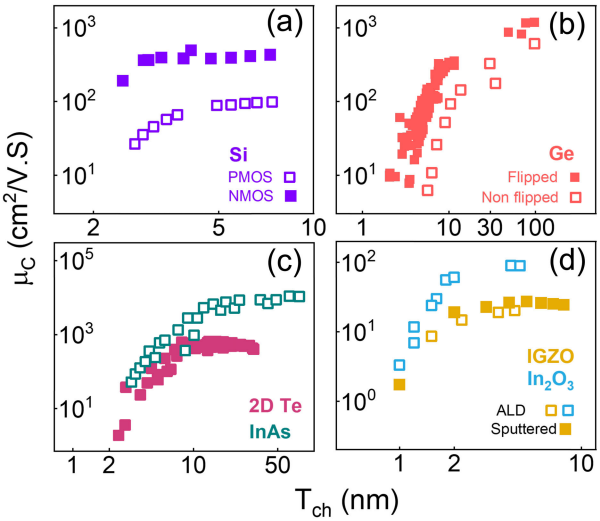

Fig. 1. (a)–(d) Room-temperature carrier mobility measured for various channel materials, including Si [4], Ge [5], [6], InAs [7], [8], 2D Te [9], IGZO [10], and In_2O_3 [11], [12].

(e.g., phonon, impurity, and surface-roughness scattering) and is given by Matthiessen's rule:

$$\mu_{\text{tot}}^{-1} = \mu_0^{-1} + \mu_{\text{AC}}^{-1}(T_{\text{ch}}) + \mu_{\text{CO}}^{-1}(T_{\text{ch}}) + \mu_{\text{SR}}^{-1}(T_{\text{ch}}). \quad (2)$$

Importantly, only a subset of scattering mechanisms exhibits a strong dependence on channel thickness. Confinement-related scattering typically strengthens rapidly as T_{ch} is reduced [13], [14], [15], whereas optical deformation potential and polar optical phonon (ODP/POP) limited mobility is comparatively weakly thickness-dependent over the same range [13], [16]. Accordingly, the total mobility can be expressed as

$$\frac{1}{\mu_{\text{tot}}} = \frac{1}{\mu_0} + \frac{\alpha_{\text{SR}}}{T_{\text{ch}}^{n_{\text{eff}}}} + \frac{\beta_{\text{CO}}}{T_{\text{ch}}} + \frac{\gamma_{\text{AC}}}{T_{\text{ch}}}, \quad (3)$$

where μ_0 represents the *thickness-independent* bulk saturation mobility arising from phonon and impurity scattering mechanisms, and the remaining terms capture *thickness-dependent* scattering contributions from surface roughness (SR), Coulomb (CO), and acoustic phonon (AC). By fitting measured μ_{tot}–T_{ch} trends using (3), the dominant scattering mechanisms can be quantitatively decomposed. Fig. 2(a) illustrates the principle of decomposition and the corresponding experimental fits for four back-gated long-channel amorphous In_2O_3 transistors with different channel thicknesses (1.2, 1.6, 1.8, and 2 nm).

This decomposition technique is subsequently applied to the remaining experimental datasets in Fig. 1(a)–(d). The extracted effective SR exponent (n_{eff}) is summarized in Fig. 2(b). Most devices exhibit $n < 6$, typically falling within the range $4 \le n \le 6$.

IV. The Origin of the "Anomalous" Exponents

The channel confinement potential in all MOSFETs is finite, and the carrier wavefunction penetrates into the surrounding barrier by an amount determined by the effective mass and barrier height. To elucidate this behavior, we consider multiple

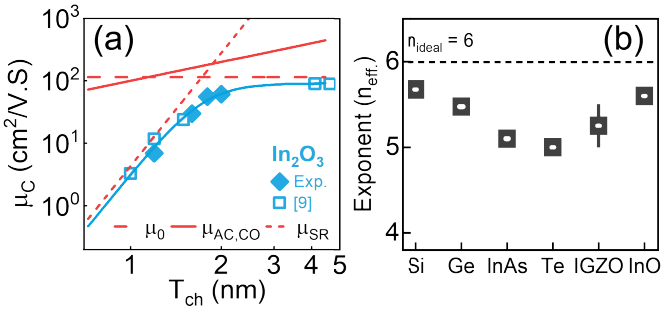

Fig. 2. (a) Mobility decomposition, separating the thickness-dependent and independent components. (b) Extracted mobility roll-off exponent (n_{eff}) for different thin-body transistor categories.

quantum well hetero-structure, where the semiconductor channel is confined by an insulator with finite potential barriers at both front and back interfaces. As T_{ch} is continuously scaled, quantum confinement increases the lowest sub-band energy, effectively widening the bandgap.

For an infinite potential-well approximation, the ground-state energy shift as a function of T_{ch} is given by [17], [21], [22]

$$E_0 \propto T_{\text{ch}}^{-2} = T_{\text{ch}}^{-l_{\text{ideal}}}, \quad \frac{\partial \ln |\partial T_{\text{ch}}/\partial E_0|^2}{\partial \ln T_{\text{ch}}} = 2(l_{\text{ideal}} + 1) = 6. \quad (4)$$

In contrast, for a finite potential well, wavefunction penetration into the gate oxide softens the confinement, modifying this dependence to $E_0 \propto T_{\text{ch}}^{-l_{\text{eff}}}$, as illustrated in Fig. 3(a). The origin of the power exponent can be rationalized by a simple analytical model, in which wavefunction penetration into the oxide effectively softens the confining walls. For a double-gate FET, this effect can be captured by a symmetric monomial family of potentials:

$$V(z) = V_0 \cdot |z/(T_{\text{ch}}/2)|^{2/k} = V_0^*(T_{\text{ch}})\,|z|^{2/k}, \quad (5)$$

where z is the confinement coordinate and $k \in (0, 1]$ controls the *softness* of confinement: $k \to 0$ approaches the square-well limit, whereas $k = 1$ yields a smooth parabolic potential. V_0 is a prefactor, and $V_0^* = V_0 \cdot (2/T_{\text{ch}})^{2/k}$ collects the physical parameters into a single coefficient. The corresponding one-dimensional Schrödinger equation with (5) is given by

$$-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z^2}\psi(z) + V_0^*(T_{\text{ch}})|z|^{2/k}\psi(z) = E\psi(z). \quad (6)$$

In (6), the only parameters carrying physical units are $\hbar$, m^*, and V_0^*. It follows that each eigenenergy E_n must take the form of

$$E_n = C_n(k)\hbar^{\alpha}(m^*)^{\beta}(V_0^*)^{\gamma}, \quad (7)$$

where $C_n(k)$ is a dimensionless function. Using $[E_n] = \text{kg}\cdot\text{m}^2\cdot\text{s}^{-2}$, $[\hbar] = \text{kg}\cdot\text{m}^2\cdot\text{s}^{-1}$, $[m^*] = \text{kg}$, and $[V_0^*] = \text{kg}\cdot\text{m}^{2-2/k}\cdot\text{s}^{-2}$, the law of *dimensional homogeneity* requires

$$[E_0] = [\hbar^{\alpha}(m^*)^{\beta}(V_0^*)^{\gamma}] = \text{kg}^{\alpha+\beta+\gamma}\cdot\text{m}^{2\alpha+(2-2/k)\gamma}\cdot\text{s}^{-\alpha-2\gamma}. \quad (8)$$

Matching exponents gives

$$\alpha + \beta + \gamma = 1,$$
$$2\alpha + \left(2 - \frac{2}{k}\right)\gamma = 2,$$

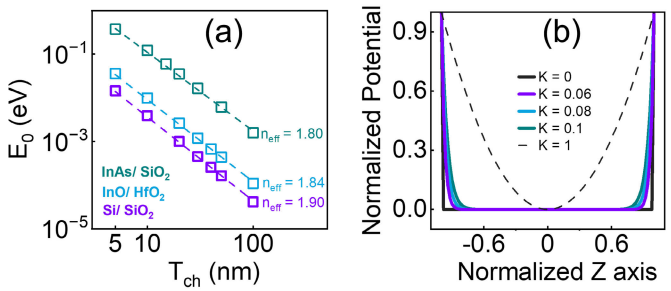

Fig. 3. (a) Ground-state energy (E_0) as a function of channel thickness (T_{ch}), showing $1/T_{ch}^{n_{eff}}$ trend. (b) Analytical interpretation of potential-well type and finite-well confinement.

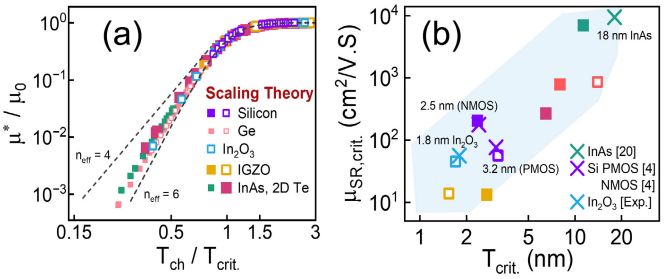

Fig. 4. (a) Universal μ_c–T_{ch} curve for thin-body devices, showing a power-law mobility roll-off below T_{crit} and saturation above T_{crit}. (b) Extracted T_{crit} and $\mu_{SR,\,crit}$ from the universal model.

$$-\alpha - 2\gamma = -2, \tag{9}$$

hence $\gamma = \frac{k}{1+k}$. Consequently, E_0 exhibits the following compact power-law scaling with T_{ch}:

$$E_0(T_{ch}) \propto V_0^*(T_{ch})^{\frac{k}{1+k}} \propto (T_{ch}^{-2/k})^{\frac{k}{1+k}} = T_{ch}^{-\frac{2}{1+k}} = T_{ch}^{-l_{eff}}. \tag{10}$$

This smoothly interpolates between the hard-wall limit ($k = 0$) and soft parabolic confinement ($k = 1$): for $k \to 0$, $V(z)$ approaches a box potential and the familiar infinite-well scaling $l_{eff} = l_{ideal} = 2$ is recovered. When $k \to 1$, $V(z)$ becomes smooth parabolic, yielding $l_{eff} = 1$. More generally, any $0 < k < 1$ corresponds to a softened confinement with $1 < l_{eff} < 2$. For example, in Fig. 3(b), a near-square yet slightly softened profile $k \approx 0.1$ gives $l_{eff} \approx 1.8$. Finally, the *softened scaling law* $E_0 \propto T_{ch}^{-l_{eff}}$ implies $n_{eff} = 2\,(l_{eff} + 1) < 6$, which explains the observed deviation from the ideal scaling relation of $n_{ideal} = 2(l_{ideal} + 1) = 6$. This relation implies that the practical SR limit across diverse transistor technologies, with the limiting exponent bounded between 4 (parabolic well, $l_{eff} = 1$) and 6 (infinite well, $l_{eff} = 2$). Exponents outside this range may indicate contributions beyond SR-induced scattering. The extracted exponent becomes fragile at thick T_{ch}, where both the increased channel thickness itself and mobility-extraction artifacts may affect the apparent exponent.

Next, we apply the finite-well concept to the mobility decomposition model in (3), enabling separate extraction of thickness-independent and SR contributions. From this, we can derive an analytical expression for the critical thickness (T_{crit}). Specifically, $1/\mu^* = 1/\mu_0 + 1/\mu_{SR}$,

with the critical thickness expressed as

$$T_{crit} = \left(\alpha_0 \frac{m_x\,\Lambda^2\,\Delta_{rms}^2}{m_z^2}\,\mu_0\right)^{\frac{1}{n_{eff}}}. \tag{11}$$

Temperature dependence is incorporated through $\alpha_0(T)$, the effective masses $m^*(T)$, and the phonon-limited bulk saturation mobility $\mu_0(T)$, which collectively account for thermal effects on screening, band structure, and scattering within the model. Understanding the temperature dependence of the mobility scaling remains an important topic for future research, particularly for cryogenic and high-temperature applications. The effects of channel material crystallinity are captured through effective mass and the surface roughness parameters (Δ_{rms} and Λ). Additionally, strain and channel orientation modify the band structure and associated transport parameters (e.g., effective mass), and are thus incorporated implicitly within the model parameters without changing its general form [6], [18], [19]. By scaling the T_{ch} with T_{crit} and normalizing the carrier mobility by its bulk saturation value (μ_0), we obtain a universal thickness-dependent mobility model, as shown in Fig. 4(a), and can be expressed as

$$\frac{\mu^*}{\mu_0} = \frac{1}{1 + \left(\frac{T_{crit}}{T_{ch}}\right)^{n_{eff}}}. \tag{12}$$

Using this universal model, we can also calculate the SR limited critical mobility, $\mu_{SR,crit}$, at T_{crit} and plot it as a function of T_{crit}. The resulting trend provides an important design guideline for next-generation transistors (GAA FETs, CFETs, FSFETs). Fig. 4(b) predicts an essential thickness limit for each technology, below which mobility degradation is dominated by SR [4], [20]. This introduces a trade-off between maximizing mobility and further scaling the T_{ch}. While the proposed scaling framework indicates that channels with $T_{ch} < T_{crit}$ can achieve higher mobility, the strong thickness dependence in this regime also implies increased sensitivity to process-induced variations in T_{ch}. This leads to enhanced mobility fluctuations, which may limit the practical scalability and reliability of such devices in large-scale circuit applications.

V. Conclusion

In this letter, we demonstrate that a universal scaling principle unifies the extreme thickness sensitivity of channel mobility across diverse transistor technologies. Within this framework, we show that (i) the mobility roll-off exponent (n_{eff}) can be systematically lower than the infinite potential well prediction ($n = 6$), with an explicit dependence on finite-well shape and wavefunction confinement, and (ii) a universal closed-form model can quantitatively extract the critical thickness and critical mobility for each technology, validated using experimental mobility data reported in the literature. Overall, the proposed framework suggests that, in the context of a potential mobility cliff, CFET technology may remain viable down to a Si channel thickness of $T_{ch} \sim 2.5 - 3.2$ nm, while the thickness scaling potential of other MOSFETs will likely depend on the ability to control surface roughness.

REFERENCES

[1] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature Electron.*, vol. 1, no. 8, pp. 442–450, Aug. 2018, doi: 10.1038/s41928-018-0117-x.

[2] B. Zohuri, *Dimensional Analysis Beyond the Pi Theorem*, 1st ed., Cham, Switzerland: Springer, 2017, doi: 10.1007/978-3-319-45726-0.

[3] M. M. F. Islam, C.-A. Shih, S. Lee, J.-Y. Lin, Z. Lin, C. Niu, M. Hong, D. Ha, P. D. Ye, and M. A. Alam, "Universality of subthreshold slope and defect characterization in thin body transistors," *IEEE Electron Device Lett.*, vol. 47, no. 2, pp. 379–382, Feb. 2026, doi: 10.1109/LED.2025.3612366.

[4] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental study on carrier transport mechanism in ultrathin-body SOI NAND p-MOSFETs with SOI thickness less than 5 nm," in *IEDM Tech. Dig.*, 2002, pp. 47–50, doi: 10.1109/IEDM.2002.1175776.

[5] X. Han, C.-T. Chen, K. Sumita, K. Toprasertpong, M. Takenaka, and S. Takagi, "Electron mobility enhancement of (111)-oriented extremely thin body Ge-on-insulator nMOSFETs by flipped smart-cut substrates," *IEEE Trans. Electron Devices*, vol. 71, no. 9, pp. 5198–5204, Sep. 2024, doi: 10.1109/TED.2024.3434782.

[6] S. Takagi, K. Sumita, X. Han, Z. Jin, Y. Chen, M.-S. Kang, M. Takenaka, and K. Toprasertpong, "Electron transport properties in nanosheet MOSFETs—From assessment of optimum channel system to operation at cryogenic temperatures," in *IEDM Tech. Dig.*, Dec. 2025, pp. 1–4, doi: 10.1109/IEDM50572.2025.11353778.

[7] K. Sumita, K. Toprasertpong, M. Takenaka, and S. Takagi, "Subband engineering by combination of channel thickness scaling and (111) surface orientation in InAs-on-insulator nMOSFETs," in *IEDM Tech. Dig.*, Dec. 2020, pp. 2.5.1–2.5.4, doi: 10.1109/IEDM13553.2020.9371995.

[8] S. P. Le, T. Ui, and T.-K. Suzuki, "Low-frequency noise in InAs films bonded on low-k flexible substrates," *Appl. Phys. Lett.*, vol. 107, no. 19, Nov. 2015, Art. no. 192103, doi: 10.1063/1.4935458.

[9] Y. Wang, G. Qiu, R. Wang, S. Huang, Q. Wang, Y. Liu, Y. Du, W. A. Goddard, M. J. Kim, X. Xu, P. D. Ye, and W. Wu, "Field-effect transistors made from solution-grown two-dimensional tellurene," *Nature Electron.*, vol. 1, no. 4, pp. 228–236, Apr. 2018, doi: 10.1038/s41928-018-0058-4.

[10] M. J. Kim, H. J. Park, S. Yoo, M. H. Cho, and J. K. Jeong, "Effect of channel thickness on performance of ultra-thin body IGZO field-effect transistors," *IEEE Trans. Electron Devices*, vol. 69, no. 5, pp. 2409–2416, May 2022, doi: 10.1109/TED.2022.3156961.

[11] M. Si, Z. Lin, Z. Chen, X. Sun, H. Wang, and P. D. Ye, "Scaled indium oxide transistors fabricated using atomic layer deposition," *Nature Electron.*, vol. 5, no. 3, pp. 164–170, Feb. 2022, doi: 10.1038/s41928-022-00718-w.

[12] M. Si, Y. Hu, Z. Lin, X. Sun, A. Charnas, D. Zheng, X. Lyu, H. Wang, K. Cho, and P. D. Ye, "Why In_2O_3 can make 0.7 nm atomic layer thin transistors," *Nano Lett.*, vol. 21, no. 1, pp. 500–506, Jan. 2021, doi: 10.1021/acs.nanolett.0c03967.

[13] M. Poljak, V. Jovanovic, D. Grgec, and T. Suligoj, "Assessment of electron mobility in ultrathin-body InGaAs-on-insulator MOSFETs using physics-based modeling," *IEEE Trans. Electron Devices*, vol. 59, no. 6, pp. 1636–1643, Jun. 2012, doi: 10.1109/TED.2012.2189217.

[14] C. Thanh Nguyen, H.-A. Shih, M. Akabori, and T.-K. Suzuki, "Electron distribution and scattering in InAs films on low-k flexible substrates," *Appl. Phys. Lett.*, vol. 100, no. 23, Jun. 2012, Art. no. 232103, doi: 10.1063/1.4722798.

[15] S. Jin, M. V. Fischetti, and T.-W. Tang, "Modeling of surface-roughness scattering in ultrathin-body SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2191–2203, Sep. 2007, doi: 10.1109/TED.2007.902712.

[16] Y. Go, R. Dutt, and N. Neophytou, "Theory of quasistatically screened electron-polar optical phonon scattering," *Phys. Rev. B, Condens. Matter*, vol. 111, no. 19, May 2025, Art. no. 195211, doi: 10.1103/physrevb.111.195211.

[17] K. Uchida and S.-I. Takagi, "Carrier scattering induced by thickness fluctuation of silicon-on-insulator film in ultrathin-body metal–oxide–semiconductor field-effect transistors," *Appl. Phys. Lett.*, vol. 82, no. 17, pp. 2916–2918, Apr. 2003, doi: 10.1063/1.1571227.

[18] I. M. Datye, A. Daus, R. W. Grady, K. Brenner, S. Vaziri, and E. Pop, "Strain-enhanced mobility of monolayer MoS_2," *Nano Lett.*, vol. 22, no. 20, pp. 8052–8059, Oct. 2022, doi: 10.1021/acs.nanolett.2c01707.

[19] J. Wang and M. Lundstrom, "Ballistic transport in high electron mobility transistors," *IEEE Trans. Electron Devices*, vol. 50, no. 7, pp. 1604–1609, Jul. 2003, doi: 10.1109/TED.2003.814980.

[20] A. C. Ford, J. C. Ho, Y.-L. Chueh, Y.-C. Tseng, Z. Fan, J. Guo, J. Bokor, and A. Javey, "Diameter-dependent electron mobility of InAs nanowires," *Nano Lett.*, vol. 9, no. 1, pp. 360–365, Jan. 2009, doi: 10.1021/nl803154m.

[21] H. Sakaki, T. Noda, K. Hirakawa, M. Tanaka, and T. Matsusue, "Interface roughness scattering in GaAs/AlAs quantum wells," *Appl. Phys. Lett.*, vol. 51, no. 23, pp. 1934–1936, Dec. 1987, doi: 10.1063/1.98305.

[22] R. K. Jana and D. Jena, "Stark-effect scattering in rough quantum wells," *Appl. Phys. Lett.*, vol. 99, no. 1, Jul. 2011, Art. no. 012104, doi: 10.1063/1.3607485.