III-V-based Hetero Tunnel FETs: A Simulation Study with Focus on Non-ideality Effects

(Invited Paper)

Andreas Schenk*, Saurabh Sant*, Kirsten Moselund†, and Heike Riel†
*Integrated Systems Laboratory, ETH Zurich, CH-8092 Zürich, Switzerland
†IBM Research-Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland

Email: schenk@iis.ee.ethz.ch

I. Abstract

We present semi-classical simulations of Gate-overlapped-Source Tunnel Field Effect Transistors (GoS-TFETs) taking into account the effects of trap-assisted tunneling, channel quantization, surface roughness, and density-of-state tails.

II. Introduction

In quest of new (”post-CMOS”) switches, in particular for ultra-low power application, the Tunnel Field Effect Transistor (TFET) [1] raised a lot of attention. The working principle of this device is the generation of electron-hole pairs by band-to-band tunneling (BTBT) between the valence band (VB) and the conduction band (CB). Thus, contrary to the common MOSFET, the source of current in a TFET is not thermionic injection, but inter-band generation.The inter-band tunnel transitions can be either direct (at the Γ-point) or indirect (e.g. from the Γ-point to the k-point of the minimum of the CB valleys as in silicon). In the latter case a phonon is needed for momentum conservation. A TFET is essentially a gated, reversed-biased *pin* diode, and the width of the tunnel barrier is modulated by the gate voltage. It is even possible to turn on the BTBT generation quickly as a consequence of the gate-induced energetic alignment between CB and VB, which eventually results in an average slope (averaged over 3 - 4 decades in the current after the onset of BTBT) steeper than 60 mV/dec. That’s why TFETs belong to the so-called ”steep-slope devices”.

III-V/Si hetero junctions have been proposed for an improved on-current as compared to Si TFETs [2]. The BTBT rate can be increased by using small-gap semiconductors like InAs or $In_{0.53}Ga_{0.47}As$ as source material, while ambipolar leakage is reduced by the wide band gap of the channel/drain materials Si or InP. Nanowires (NWs) are advantageous in terms of gate control and strain relaxation when their diameter is scaled down [3]. Borg et al. integrated individual InAs/Si hetero-structure NW tunnel diodes onto Si using selective epitaxy in nanotube templates [4], an approach which allows to start with Si substrates of any crystalline orientation and to scale the diameter of the NWs down to reasonable limits, thereby improving previous techniques of nanometer-scale hetero epitaxy [5]–[8]. Temperature-dependent *IV*-measurements of p-type TFETs fabricated by this technology indicated that the sub-threshold swing (SS) is limited by traps [9]. The present TCAD simulation study partly takes advantage of the geometry (see Fig. 1), the measured *IV*-characteristics, and the limited electrical characterization of these devices. In this paper, all simulations are performed with the semi-classical device simulator Sentaurus-Device (S-Device) [10] using the nonlocal BTBT model with calibrated parameters [11]. Non-ideality effects as trap-assisted tunneling (TAT) via interface and bulk traps, channel quantization, roughness at material interfaces, and density-of-state (DOS) tails induced by random doping fluctuations (RDF), are modeled in the frame of the so-called Physical Model Interface (PMI) of S-Device. The developed models for the non-idealities and their impact on the transfer characteristics are presented in Section IV.

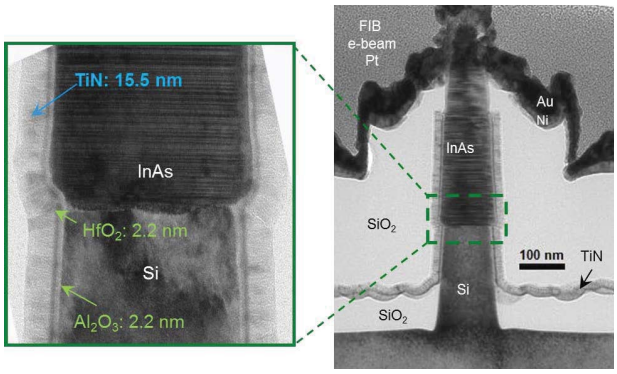

Fig. 1: TEM image of an InAs/Si NW p-TFET showing the geometrical dimensions and a close-up of the InAs/Si interface which is not smooth across the entire junction area and has a high density of defects. Also visible are stacking faults caused by the change in crystal structure of InAs between wurtzite and zincblende. Figure taken from [9].

III. Experimental data of p-type InAs/Si NW TFETs

InAs nanowires with diameters of few tens of nanometers have been grown on Si substrates by IBM Research Zurich [12]. This material configuration provides a small effective tunnel gap (low tunnel barrier) while allowing for integration on a Si platform which maintains the advantages of Si for the channel and drain regions. Although the TFETs based on this technology (shown in Fig. 1) exhibit a good electrostatic control due to their gate-all-around configuration, they can still be considered as bulk-like, i.e. geometrical confinement is negligible. The measured transfer characteristics are shown in Fig. 2 for two different temperatures. The measured on-current at room temperature is 1.62 μA ($\sim$ 5.2 μA/μm) at $|V_{GS}|$ = 1.0 V, $|V_{DS}|$ = 0.5 V and the I_{on}/I_{off} ratio is $\sim$ 10^6. The

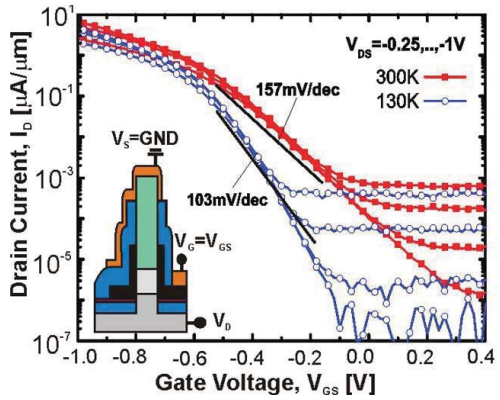

Fig. 2: Measured $I_{\mathrm{D}}V_{\mathrm{GS}}$-characteristics of the InAs/Si NW p-TFET of Fig. 1 for two temperatures and different values of V_{DS}. Figure taken from [9].

steeper SS obtained at 130 K clearly proves the presence of traps, which limit sub-60 mV/dec operation.

IV. Impact of Non-ideality Effects on TFET Performance

The ideal slope of the TFET characteristics (neglecting all non-idealities) is determined by the details of the onset of BTBT. In GoS-TFETs, BTBT takes place along two kinds of tunnel paths - almost parallel to the gate and across the pn-junction ("point tunneling" paths) and almost perpendicular to the gate and within the source ("line tunneling" paths). Often point tunneling sets in first, since line tunneling only starts after the energetic alignment of VB and CB (inversion conditions in the source channel). A sub-thermal SS of point tunneling is hard to achieve because a sudden and strong 3D-3D DOS matching is only possible when the band edge profile is "step-like" which requires (i) a perfect gate control, (ii) a very steep doping gradient, and (iii) preferably the gate edge exactly at the pn-junction (which, however would kill line tunneling completely). A sub-thermal SS of line tunneling is easier to achieve thanks to the cut of the semiconductor band gap at the oxide interface and the 2D-3D DOS matching. One of the design goals is, therefore, to delay the onset of point tunneling. Two means can be used for this: the band offsets of the hetero-structure and pocket implants [19]. All non-idealities discussed in the following act contrary to a sub-thermal slope.

A. Trap-assisted Tunneling

TAT is an additional electron-hole pair generation mechanism in TFETs which sets in prior to BTBT due to the lower tunnel barrier. Being a multi-phonon, field-assisted process, it requires efficient generation-recombination (G-R) centers (as usual called "traps" in the following). Three kinds of traps may contribute to the TAT current: bulk traps, traps at the material interface of the hetero-structure, and traps at the oxide interface. Note, that CV experiments to extract D_{it} at interfaces do not measure the density-of-states of G-R centers active in the multi-phonon process. The latter is only a subset of the measured D_{it} levels. The S-Device model for interface TAT needs an additional parameter "interaction volume of the trap" which is not only a scaling factor but serves to distinguish

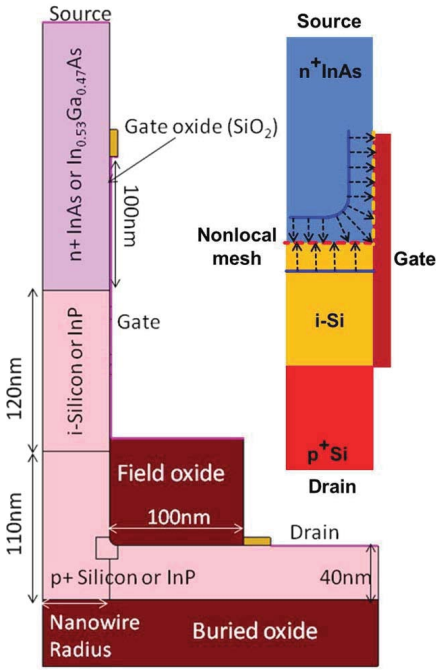

Fig. 3: Simulation domain of the GoS InAs/Si NW TFET (left) and nonlocal mesh for TAT tunnel paths from traps at hetero-interface (dashed red) and traps from gate oxide interface (dashed yellow).

between the charging of the trap levels and the electron-phonon coupling of the trap states. The physics at interfaces is highly self-consistent: already the occupation of the D_{it} changes the electrostatics and, thus, the onset voltage and the tunnel rates (including the BTBT rate). Surface SRH generation without tunneling causes a leakage current which appears to be gate-bias-dependent due to the downward move of the quasi-Fermi levels in the CB which gradually empties the final states for the generation process. We found that in InAs/Si NW TFETs, bulk traps cannot explain the measured weak slope [20]. In GoS InAs/Si NW TFETs, both gate-oxide and material interface traps are expected to contribute to TAT. Fig. 3 illustrates the simulation domain which adapts the geometry of the fabricated device shown in Fig. 1. Appropriate internal region interfaces must be defined to connect the traps with source and drain on a nonlocal mesh which supports the actual tunnel paths. As an example we show in Fig. 4 the different generation contributions that yield the total $I_{\mathrm{D}}V_{\mathrm{GS}}$-curve for the case of gate-oxide interface traps. Note, that the initial branch is solely determined by surface SRH generation! When the electron quasi Fermi level crosses the CB edge, the contribution of surface SRH generation remains constant and the out-tunneling from trap levels into the CB starts to become effective. This provides a" short-cut" to the thermal electron emission step. With increasing gate bias the tunnel probability increases and the most probable transition energy shifts towards the VB edge which finally yields a current comparable to the pure BTBT current.

B. Channel Quantization

Line tunneling starts when the transistor channel becomes strongly inverted. Under such conditions, the triangular-like

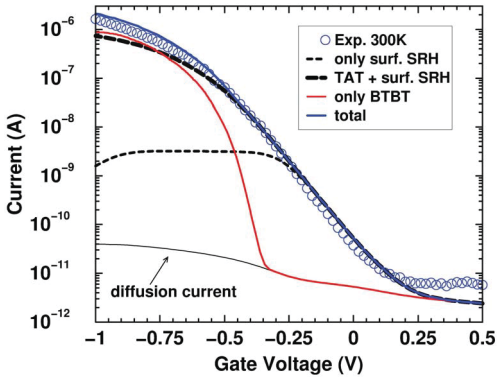

Fig. 4: The different generation contributions to the drain current of the InAs/Si NW p-TFET at 300 K shown in the left part of Fig. 3: only surface SRH generation (short-dashed), TAT included (long-dashed), only BTBT (solid red), total (solid blue). Parameters: $V_{\mathrm{DS}} = 0.5\mathrm{V}$, $DOS_{\mathrm{InAs}} = 3.4 \cdot 10^{17}\,\mathrm{cm}^{-3}$, $WF = 4.8\,\mathrm{eV}$, $\sigma_{\mathrm{n,p}} = 10^{-14}\,\mathrm{cm}^2$, $D_{\mathrm{it}} = 10^{13}\,\mathrm{cm}^{-2}\mathrm{eV}^{-1}$ [14], $V_{\mathrm{trap}} = 10\,\mathrm{Å}^3$, $S\hbar\omega = 120\,\mathrm{meV}$.

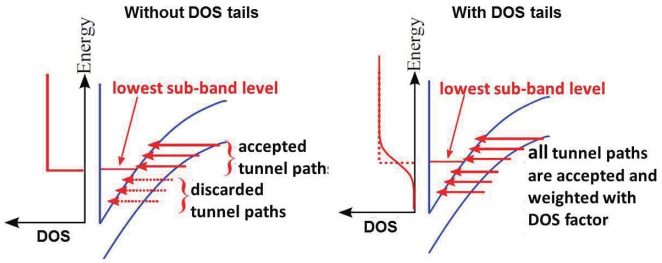

Fig. 5: Modeling the effect of channel quantization without (left) and with 2D DOS tails (right).

potential well of the channel quantizes the electronic states and a 2DEG forms. This shifts the onset voltage of line tunneling to a higher value when compared with the 3DEG case [13]. Besides, the nature of the wave functions changes from Airy functions to "quantized" Airy functions, which modifies the tunnel generation rate. These effects are not covered by the default "dynamic nonlocal path BTBT model" based on Kane's treatment of BTBT [15], available in S-Device [10]. A method to model the quantization effect within a semi-classical framework has been proposed by Vandenberghe et al. [16]: Tunneling paths with energies above the lowest sub-band level are accepted while those with a lower energy are rejected, as illustrated in the left part of Fig. 5. We incorporated such a model in S-Device by developing our own code for the "dynamic non-local path BTBT model" using the PMI for "nonlocal recombination". The implemented algorithm detects a line tunneling path at a proper energy by checking whether the extension of this path intersects the oxide interface. If not, it is not considered to be affected by channel quantization. As an alternative to this rather involved approach, a much simpler method was developed making use of the Quantum Potential Correction (QPC) available in S-Device. In the electrical quantum limit [21] (constant field) the QPC for the splitting of the lowest sub-band is given by

$$E_{\mathrm{QM}} = \begin{cases} \hbar\Theta_c \cdot a_1 - d_{\mathrm{ox}}F & \text{if } E_{\mathrm{QM}} > 0 \\ 0 & \text{otherwise,} \end{cases}$$

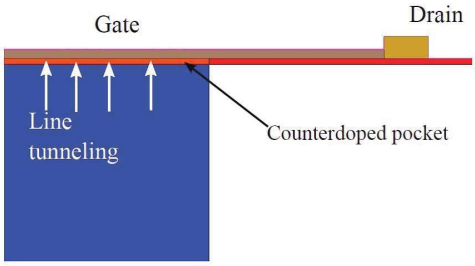

Fig. 6: InGaAs TFET with counter-doped pocket. The special geometry favors line tunneling and is used to analyze the impact of surface roughness on line tunneling.

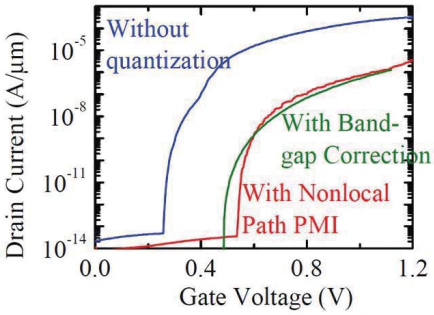

Fig. 7: Comparison between the nonlocal PMI model which employs the path rejection method (red), the PMI model that modifies the band gap (green), and the case without channel quantization (blue).

where $\hbar\Theta_c = (e^2\hbar^2F^2/2m_c)^{1/3}$ and a_1 denotes the first zero of the Aity function. The resulting effective CB edge is then computed by $E_{\mathrm{CB}}^{\mathrm{eff}} = E_{\mathrm{CB}} + E_{\mathrm{QM}}$. A special TFET geometry which favors vertical tunneling (see Fig. 6) was simulated to compare both quantization correction methods. The results are plotted in Fig. 7. The good agreement between the two models suggests that the simpler model for quantum correction is as effective as the elaborate model.

C. Roughness of Gate Oxide Interface

In the absence of surface roughness, the sub-band levels are well defined and the 2D DOS has the well-known staircase form. A rough oxide-semiconductor interface causes random fluctuations of the boundary wall of the triangular-like potential well. As a consequence, the step-like DOS smears out to form tail states as shown in the right part of Fig. 5. A simplified model for 2D DOS tails originating from an arbitrary random field has been derived by Quang et al. [17]. Here, we apply their model to the case of a random potential due to surface roughness. Instead of rejecting the tunnel paths below the lowest sub-band energy, the energy of the tunnel path relative to the ideal sub-band level is determined and the DOS factor is calculated for this relative energy. The spectral tunnel rate from the "dynamic nonlocal path BTBT model" is then multiplied with the 2D DOS to obtain the integrated tunnel rate. This approach has been made available in S-Device by modifying the code for the PMI model of channel quantization. The same "vertical" TFET as shown in Fig. 6 was simulated using the above model. Its transfer characteristics without and with

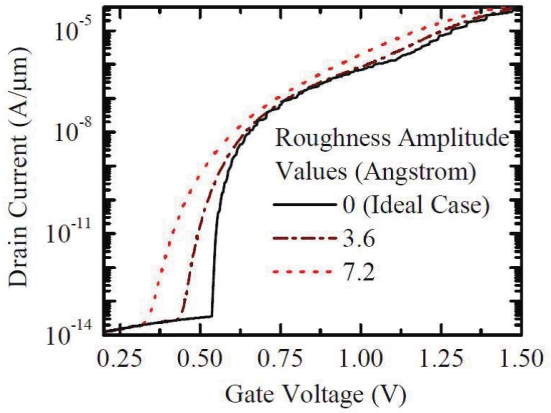

Fig. 8: Effect of the surface roughness amplitude on the transfer characteristics of the TFET in Fig. 6.

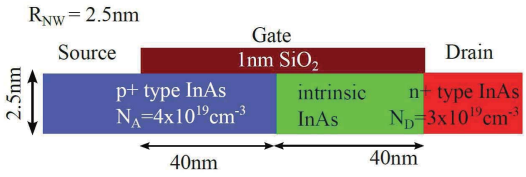

Fig. 9: Device structure to model the effect of DOS tails caused by random dopant fluctuations.

the inclusion of 2D DOS tails due to surface roughness are presented in Fig. 8.

D. DOS Tails from Random Dopant Fluctuations

Kanes model of DOS tails [18] includes the effect of random dopant fluctuations (RDF). Using his tail-state description, we derived a semi-analytical expression of the generation rate to model the effect of RDF-induced DOS tails in the frame of zero-phonon TAT [22]. Tail states are considered as trap states with sufficient localization. Transitions from VB tail states to CB Bloch states and from VB Bloch states to CB tail states are taken into account, whereas transitions between tail states are neglected due to their much smaller tunnel probability. The proximity of the tail states to the band edges allows easy thermal excitation into conducting states which completes the inter-band process. There are two differences to a BTBT transition: (i) the tunnel gap is effectively reduced, (ii) instead of the reduced effective mass the imaginary dispersion is dominated by the respective single-band mass, the mass of the localized state being a fit parameter. The second fit parameter is the characteristic energy of the DOS tail. The model was implemented in S-Device using the PMI for "non-local recombination". To analyze the impact of RDF-induced tail states on TFETs, we simulated a gate-all-around InAs nanowire TFET (see Fig. 9). The transfer characteristics of this device with and without RDF-induced DOS tails are presented in Fig. 10. One observes an earlier onset of tunneling and a degradation of the SS as consequence of the tail states.

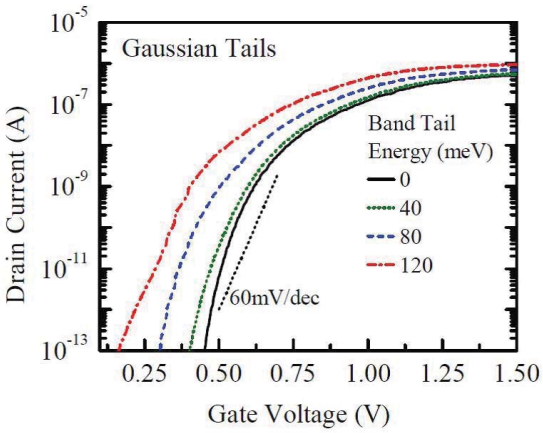

Fig. 10: Effect of RDF-induced DOS tails on the transfer characteristics of the TFET shown in Fig. 9.

V. Conclusion

Various mechanisms contribute to the degradation of the SS in GoS-TFETs. As main effects we identified (i) the surface SRH generation at the oxide interface in combination with zero-phonon tunneling from trap levels into the CB and (ii) TAT at the material interface with weak phonon assistance. Channel quantization delays the onset of line tunneling which may be hidden if point tunneling dominates, but it also reduces the on-current. Two TCAD solutions to include channel quantization were presented. Furthermore, we developed models for the impact of surface roughness and RDF-induced DOS tails on the SS. Although values of the characteristic energies by which the DOS is smeared out are not known, it seems that both effects could become visible if TAT is suppressed.

VI. Acknowledgment

Funding from the European Communitys Seventh Framework Programme under Grant Agreement No. 619509 (Project E^{2}SWITCH) is acknowledged.

References

[1] S. Banerjee et al., IEEE Electron Dev. Lett. **8**, 347 (1987).
[2] A. S. Verhulst et al., IEEE Electron Dev. Lett. **29**, 1398 (2008).
[3] M. T. Björk et al., Appl. Phys. Lett. **80**, 1058 (2002).
[4] M. Borg et al., Nano Letters **14** (4), 1914 (2014).
[5] M. T. Björk et al., Appl. Phys. Lett. **97**, 163501 (2010).
[6] K. Tomioka et al., Appl. Phys. Lett. **98**, 083114 (2011).
[7] K. E. Moselund et al., IEEE Electron Dev. Lett. **33**, 1453 (2012).
[8] H. Riel et al., IEDM Tech. Digest, pp. 391-394, 2012.
[9] D. Cutaia et al., IEEE Journal of the Electron Dev. Soc. **3** (3), 176 (2015).
[10] Sentaurus-Device User Guide, version 2015.06, Synopsys Inc., Mountain View, California (2015).
[11] H. Carrillo-Nuñez et al., Solid-State Electronics **113**, 61 (2015).
[12] K. E. Moselund et al., J. Appl. Phys. **113**, 184507 (2013).
[13] W. G. Vandenberghe et al., Appl. Phy. Lett. **98**, 143503 (2011).
[14] P. Mensch et al., IEEE Trans. Nanotechnol. **12** 3, 279 (2013).
[15] E. O. Kane, J. Phys. Chem. Solids **12**, 181 (1959).
[16] W. G. Vandenberghe et al., Proc. SISPAD, 271 (2011).
[17] D. N. Quang et al., phys. stat. sol. (b) **207**, 111 (1998).
[18] E. O. Kane, Phys. Rev. **131**, 79 (1963).
[19] S. Sant et al., Trans. Electr. Dev., DOI: 10.1109/TED.2015.2489844.
[20] A. Schenk et al., Proc. DRC, 201 (2012).
[21] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
[22] A. Schenk, J. Appl. Phys. **71** (7), 3339 (1992).