Energy Filtering Effect at Source Contact on Ultra-Scaled MOSFETs

Johan Saltin, Nguyen Cong Dao, Philip H. W. Leong, Senior Member, IEEE and Hiu Yung Wong, Senior Member, IEEE

Abstract— We postulate that in ultra-scaled Field Effect Transistors (FET), such as nanowires in sub-7nm technology, the source contact will act as an energy filter and increase the effective temperature of carriers arriving at the channel barrier. This is due to the absence of inelastic scattering in the short source-contact-to-channel region. As a result, the Sub-threshold Slope (SS) will increase substantially. In this paper, we verify this energy filtering effect through numerical calculations and Technology Computer-Aided-Design (TCAD) simulations calibrated to quantum solvers for electrostatics. It is found that SS degradation increases as the source metal workfunction increases. At 300K, in the nanowire simulated, SS increases from 94mV/dec to 109mV/dec for gate length, $L_G = 10$nm and from 72mV/dec to 88mV/dec for $L_G = 15$nm, representing an increase of effective carrier temperature from 300K to more than 340K. The simulation result is also verified by including the Schroedinger equation (SE) for tunneling in TCAD simulation. It is also found that such an effect is worse at higher device temperature and disappears at cryogenic temperature.

Index Terms—Sub-threshold Slope, Energy Filter, Nanowire, Schottky Contact, Ballistic Transport

I. INTRODUCTION

One of the main goals in transistor scaling is to increase the drain current ON/OFF ratio ($I_{ON}/I_{OFF}$) [1] so that the transistor is a close approximation to an ideal switch. There are three ways to achieve this goal. The first is to increase $I_{ON}$ through material engineering, e.g. using high-mobility or strained materials [2]. The second is to reduce $I_{OFF}$ by suppressing leakage, e.g. via Gate-Induced-Drain-Leakage (GIDL) [3]. The third is to reduce the Sub-threshold Slope (SS). This is important especially when the supply voltage is also reduced. The smaller the SS, the more the drain current is reduced as the gate voltage is decreased, and so lower leakage can be achieved for a given $I_{ON}$. Due to the operating mechanisms in conventional Field Effect Transistors (FET), the minimal SS one can achieve is $ln(10)kT/q$ [4], where $k$ is the Boltzmann Constant, $T$ is the temperature and $q$ is the basic electronic charge. At 300K, this is about 60mV/dec because FET operation is based on the thermionic emission of carriers over a gate-controlled barrier in the channel. SS generally becomes larger (worse) when the transistor is scaled due to the reduction in gate control. Novel transistors based on different mechanisms such as the Tunnel Field Effect Transistor (TFET) [5] and Negative Capacitance FET (NCFET) [6] have been proposed to reduce SS below 60mV/dec at 300K. However, such devices usually either have low $I_{ON}$ or are difficult to optimize.

Therefore, it is expected that traditional FETs will continue to dominate and be scaled to the 1nm technology node for most applications [1][7]. To avoid degradation of SS in ultra-scaled transistors, it is likely that gate-all-around (GAA) structures such as nanowires will be used [1].

It is also well known that Fermi level pinning is inevitable in metal-semiconductor contacts [8]. Therefore, regardless of the metal workfunction, every metal-semiconductor contact is of Schottky type in nature, behaving in an Ohmic manner because of very efficient carrier tunneling through the thin Schottky barrier when the contact is heavily doped [9].

In this paper, we postulate that when a device is sufficiently scaled, due to lack of inelastic scattering in the source region, the source contact can act as an energy filter and increase (worsen) the SS of future ultra-scaled devices. This is confirmed in numerical calculations and then in TCAD simulation of a nanowire (Fig. 1). Therefore, Fermi level de-pinning [10] or source region engineering will be crucial to maintaining low SS in future ultra-scaled devices.

This paper is based on a paper entitled “Degradation of Sub-Threshold Slope in Ultra-Scaled MOSFETs due to Energy Filtering at Source Contact” presented at the 2019 IEEE S3S Conference.

Johan Saltin and Hiu Yung Wong are with the Electrical Engineering Department, San Jose State University, 1 Washington Sq, San Jose, CA 95192, USA. (Corresponding author: Hiu Yung Wong. E-mail: hiuyung.wong@sjsu.edu).

Nguyen Cong Dao and Philip H. W. Leong are with the School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia.
II. ENERGY FILTERING AT SCHOTTKY CONTACT POSTULATE

Here, an n-type metal/semiconductor (M/S) Schottky contact is used as an example. As shown in Fig. 2, a Schottky contact with a heavily doped semiconductor (e.g. $10^{21}$ cm$^{-3}$) has a thin triangular barrier. Therefore, electrons can tunnel through it very easily in both directions, rendering it Ohmic. At the metal, electrons are thermalized and obey a Fermi-Dirac distribution. In contrast to the common definition of energy, here $\Delta E$ is defined as the energy difference between the top of the Schottky barrier ($E_{\text{top}}$) and the energy considered ($E$ ($\Delta E = E_{\text{top}} - E$)). Since only the electrons far away (compared to $kT$) from the Fermi-level are important in this discussion, we approximate the distribution as a Boltzmann distribution, $f \approx e^{\Delta E/kT}$, with the electron density decreasing exponentially as the energy ($E$) increases (i.e. when $\Delta E$ decreases). Since the barrier is thinner at the top, the transmission probability, $M(\Delta E) \sim e^{-3(2m)^{1/2}E^{3/2}/4\hbar^2}$, derived using WKB approximation, is higher for higher energy electrons. Here, $F = 0.5V/nm$ (which is the typical value obtained in the following TCAD simulations) is used and $m = 0.2m_e$ where $m_e$ is the rest electron mass. Therefore, right after tunneling from metal to semiconductor, the electron population will no longer follow a Boltzmann Distribution (with more energetic electrons in the total population) and has a higher effective temperature ($T_{\text{eff}}$). $T_{\text{eff}}$ can be considered as a parameter to map the new distribution to an equilibrium Boltzmann Distribution. As seen in Fig. 2, $T_{\text{eff}}$ is 340K because $f(T=300K) \times M(T=300K)$ and $f(T=340K)$ match pretty well overall. This is called the energy filtering effect of a Schottky contact similar to the idea in [12]. In a regular M/S contact, due to phonon scattering, electrons will quickly thermalize to reach thermal equilibrium with the lattice and restore the Boltzmann distribution before reaching the channel barrier, and the regular subthreshold slope based on $\ln(10)kT/q$ will still be obtained.

However, if the contact to channel distance is sufficiently short, the electrons will not undergo inelastic scattering with phonons and will reach the channel with higher $T_{\text{eff}}$, increasing SS. This is expected to be true even the source is heavily doped and experiences significant Coulomb scattering, which is an elastic scattering and will not alter the electron distribution in energy space [13].

Reference [14] shows that the mean free path of the electron in silicon nanowire can be as large as 40nm in the $<110>$ direction. According to IRDS [1], the Contacted Poly Pitch is expected to reach 42nm in 5nm technology node, with contact critical dimension (CD) = 14nm and gate length $L_C=18nm$. Therefore, source contact to channel distance is only (42nm-14nm-18nm)/2=5nm. It is thus expected that electrons tunneling through the source contact (which is Schottky in nature) will arrive at the channel without inelastic scattering and have higher $T_{\text{eff}}$ than the device operating temperature.

III. TCAD SIMULATIONS

To further verify our theory, TCAD Sentaurus was used to simulate an n-type silicon nanowire [15]. The drift-diffusion carrier transport model (DD) is used to perform the study. DD allows the inclusion of various scattering mechanisms in carrier mobility such as Coulomb scattering due to impurities, surface acoustic phonon and roughness scatterings, and optical phonon scattering (captured in high field saturation model). These models are included by activating Masetti Model, Lombardi Model and Extended Canali Model in the simulation [15]. The carrier transport in nanowire is assumed to be in the $<100>$ direction. The density gradient model for electrons is used to account for quantum confinement effects (except simulations in Fig. 6, 7 and 8 where convergence is difficult due simulation at cryogenic temperature or inclusion of Schroedinger equation) and is calibrated against a Non-Equilibrium Green’s Function (NEGF) transport model [16]. Self-heating is not included for simplicity and better convergence. The gate workfunction is set to 4.6eV. A Schottky boundary condition is used for the source contact with various metal workfunctions. The non-local tunneling model using WKB approximation is used to model the tunneling of electrons into the nanowire from the contacts with tunneling mass = 0.2$m_e$. Fig. 1 shows the structure simulated.

The non-local tunneling model is set up to have two options [17][15]. One is to have the electrons tunneled from the metal thermalized immediately after the Schottky barrier. This represents the case when there is no energy filter effect as all tunneled carriers are thermalized before reaching the channel. Another option is to send the tunneled electron to the top of the channel barrier after Schottky barrier tunneling, which is equivalent to ballistic transport from the Schottky Source to
the top of the channel barrier. This represents the case when there is no inelastic scattering and the electrons reach the channel barrier at a higher temperature (Fig. 2 and Fig. 3).

IV. SIMULATION RESULTS

Figure 3 shows that when the source metal workfunction is 5.1eV, the subthreshold slope (using the slope from I_D = 10^-10A to 10^-8A) is degraded from 94mV/dec to 109mV/dec for L_G=10nm as predicted. This is equivalent to T_eff = 347K using SS = n ln \frac{kT}{q}, where n is the non-ideality factor. By assuming the ballistic transport and non-ballistic transport cases have the same n and since the non-ballistic case has device temperature equals to the ambient temperature (300K in this case), T_eff is found by considering the ratio of SS of both cases. This is similar to the numerical calculation. It also shows the conduction band profile from source to drain when the transport in the source is treated ballistically. As the gate voltage increases, the channel barrier reduces, and more electrons can go over the barrier. However, due to the energy filtering effect at the source contact, the number of extra electrons that can overcome the barrier when the channel barrier is lowered, less than when the distribution was Boltzmann. As a result, it has a large SS.

Figure 4 shows the change of SS as a function of the source metal workfunction (WF) for L_G=10nm and 15nm. The SS degradation starts when WS is at 4.4eV and becomes worst when WS = 5.1eV. When WS is 4.3eV or less, in the off state, the channel barrier is higher than the Schottky barrier and the transmission coefficient is 100% (quantum reflection is ignored) and there is no energy filtering effect. As a result, there is no SS degradation when the WS is low.

V. TEMPERATURE EFFECT

The effect of energy filtering at source contact is expected to have strong temperature dependency because the effect is a result of the deviation of the carrier distribution in the energy space from equilibrium Boltzmann and Fermi-Dirac distributions. Transistors usually operate at elevated temperatures due to self-heating, especially in high-performance CPUs and GPUs. Therefore, it is important to understand how energy filtering effect changes at elevated temperatures. Nowadays, there is also an increased interest in the cryogenic applications of CMOS circuits as quantum computer peripheries. Cryogenic CMOS shows abnormal SS when the device temperature is less than 30K [18][19]. For example, at 4.2K, the SS is ~11mV/dec instead of the theoretical value of 0.8mV/dec. However, there is no consensus on the root cause of such abnormal behavior [20]. Therefore, it is also important to understand if such behavior can be a result of source contact energy filtering because at cryogenic temperature, the inelastic scattering rate reduces

![Fig. 3. The subthreshold region of the I_D-V_G of the nanowire with L_G=10nm with source metal WF=5.1eV (top). The corresponding conduction band profiles from source to drain are shown at the bottom. One of the tunneling paths is illustrated. In the non-ballistic case, electrons are thermalized at “A” once passing through the Schottky barrier. In ballistic case, electrons are not allowed to thermalized until they have passed the peak “B” of the channel barrier.](image)

![Fig. 4. SS of nanowires as a function of source metal Workfunction.](image)

![Fig. 5. Calculated change of T_eff as a function of device temperature. Markers are results from calculation and red line is a power law fitting curve. source metal workfunction = 5.1eV is assumed.](image)
substantially due to lack of phonon scatterings. It is possible such an energy filtering effect exists in larger transistors due to the much longer mean free path for ballistic transport.

By using the methodology illustrated in Fig. 2, T_{eff} in the subthreshold region is calculated for the device with source metal WF = 5.1eV from 4K to 400K. Fig. 5 shows that the change of T_{eff} increases with device temperature. This is because at higher temperatures, electron distribution in energy space is broadened and this increases the effect of energy filtering. On the other hand, at lower temperatures, the change of T_{eff} reduces and approaches zero at 0K. This is because the energies of the electrons participating in the tunneling in the subthreshold region concentrate in a very narrow band around the Fermi level and will not “feel” the different transmission coefficients at different energies as much. It can also be better understood by considering in the extreme case when T = 0K, only the electron at the Fermi level will participate in the conduction when the transistor switches from off-state to on-state, and thus, the shape of the source Schottky barrier does not have any effect in altering the electron distribution after tunneling.

To verify the calculation, TCAD simulation is also performed for the device in Fig. 1 with L_G = 10nm from 20K to 400K as showed in Fig. 6. Since convergence is known to be challenging at sub-100K TCAD simulation, V_D is set to 0.1V and quantum correction is turned off in the simulations in Fig. 6 and Fig. 7. It is found that TCAD simulation gives a similar trend as the calculation in Fig. 5. The major difference is, at 400K, the change of T_{eff} is not as much as in calculation. This is because thermionic emission over the source Schottky barrier becomes more important at higher temperatures while thermionic emission contribution to the transport is neglected in the calculation. This reduces the effect of the filtering effect due to tunneling. At device temperature less than 50K, the ballistic case has smaller T_{eff} than the non-ballistic case but only by about 3K. This is due to the noise in the simulation, such as the non-straight curve in the sub-threshold region of the 20K case (Fig. 7). Based on the calculation and simulation results, therefore, the energy filtering effect at source contact cannot cause the abnormal SS in cryogenic CMOS.

VI. DISCUSSION

One of the major assumptions in this study is that the electrons will retain the distribution of f(T=300K) × M(T=300K) in the source region due to the lack of inelastic phonon scattering. However, electron-electron scattering is not considered, which redistributes the energy and momentum among the electrons [21]. Due to the high electron concentration at the source, the relaxation time is about 0.1fs [22]. Therefore, the electrons are in equilibrium with each other but still at an elevated effective temperature as a whole when they reach the channel barrier. As a result, it is expected that in the ballistic case, the SS slope may be different from what has been calculated but still worse than the non-ballistic case.

As pointed out in [23], the metal-semiconductor interface is one of the most challenging problems in solid-state physics. It is expected rigorous self-consistent quantum transport simulation is needed for quantitatively correct results, in order to account for e.g. source starvation [24], sub-band formation and image force. However, since both ballistic and non-ballistic cases share the same setup, it is expected that our result is still qualitatively correct.
This study doesn’t take the backscattering of electrons from the channel barrier into account, which can reduce the ballisticity [25][26]. However, as shown in [25], backscattered carriers are only a small portion of the carriers even at 15nm and it is expected to be less at 10nm or below. Therefore, most of the carriers are still expected to transport ballistically.

In the extreme case of WF=5.1eV and doping of 10^{24} cm^{-3}, the depletion width is only 1.5nm and comparable to the de Broglie wavelength of the electrons. Error in WKB approximation may become large theoretically. It is desirable to solve the Schroedinger equation (SE) for M. However, convergence is difficult in TCAD when SE is included. By using a reduced model setup (no quantum correction and high-field saturation model is turned off) and a low drain bias (V_D = 0.01V) for better convergence, it is confirmed that WKB gives a similar result as in SE (Fig. 8). Therefore, WKB still gives qualitatively correct results despite the possible error in the approximation.

This results in this study applies not only to MOSFET but also to other transistors such as Schottky-Barrier (SB) Transistor using Si or 2D materials [27]. Since the source in SB transistor are undoped and thus lack of elastic Coulomb scatterings due to ionized dopants, it is expected that the energy filtering effect will be more prominent.

VII. CONCLUSION

We report that in ultra-scaled FET, SS will be larger than expected due to the energy filtering effect in the source contact. This is because every source contact is Schottky in nature and if there is a lack of inelastic scattering in the source region, carriers cannot thermalize with the lattice before they reach the channel. Numerical calculation and TCAD simulation show that T_eff increases from 300K to more than 340K. The result is also validated by solving the Schroedinger equation (SE) for M in TCAD simulation. Therefore, it is important to either avoid Schottky contact at the source (e.g. through Fermi-Level de-pinning engineering [10]) or special source region engineering is required. Moreover, it is also found that such an energy filtering effect becomes worse at higher temperatures and disappears at cryogenic temperatures. Therefore, such energy filtering effect is not the cause of the abnormal SS in cryogenic CMOS.

ACKNOWLEDGMENT

The authors thank Synopsys Inc. for the donation of TCAD licenses and Dr. Andreas Wettstein of Synopsys for his fruitful discussion and advice on the use of non-local tunneling model. H. Y. Wong thanks the support from San Jose State University College of Engineering’s new faculty start-up fund.

REFERENCES


