Impact of Nanowire Variability on Performance and Reliability of Gate-all-around III-V MOSFETs

*S. H. Shin, M. Masuduzzaman, J.J Gu, M. A. Wahab, N. Conrad, M. Si, P. D. Ye, and *M. A. Alam

*E-mail: {shin136, alam}@purdue.edu, Phone: (765) 494-5988, Fax: (765)-494-2706
Department of ECE, Purdue University, West Lafayette, IN 47907, USA

Abstract

Gate-all-around (GAA) transistors use multiple parallel nanowires to achieve the desired ON current. The fabrication and performance of GAA transistors have been reported, however, a fundamental consideration, namely, the scaling and variability of transistor performance as a function of the *number of parallel NWs* is yet to be discussed. In this paper, we (i) examine how the overall performance matrix (e.g., I_{ON}, I_{OFF}, V_{th}, SS, R_C) depends on the number of parallel NWs, (ii) theoretically interpret the results in terms of *variability* and *self-heating* among the NWs, (iii) compare the reliability of multiple NW devices (ΔV_{th}, ΔSS, both stress and recovery) with a planar device of similar technology. We find that the self-heating and NW-to-NW variability are reflected in novel properties of variability and reliability of GAA transistors that are neither anticipated nor observed in the corresponding planar technology.

Introduction

Three-dimensional electrostatic control of MOSFET channel is essential for technologies beyond the 20nm node. The Gate-all-around (GAA) MOSFET ensures surrounding-gate electrostatic control of each channel/nanowire (NW) for immunity from short channel effects. An intrinsic trade-off of such a device geometry is the reduced cross-section per NW, and thus, several NWs connected in parallel, along with high mobility materials such as InGaAs, are required to achieve the desired drive current, (I_{ON}). We may expect that (i) the I_{ON} will scale with the *number* of NWs (#NW), and (ii) the variability/reliability would average over the NWs. However, we know of no reports in the literature that have analyzed these important propositions in any detail.

In this paper, we carefully examine the performance and reliability characteristics of GAA transistors with variable number of NWs. We find that both aspects of the classical proposition are wrong: We demonstrate that the performance for multiple NWs devices *does not* scale linearly with #NWs. This is because self-heating becomes more important at higher density of NWs, leading to increased I_{ON} (Fig. 4, 5), yet reduced I_{ON}/I_{OFF} ratio (Fig. 6). Moreover, the V_{th} and/or SS variability of individual NWs lead to increased SS as the number of NWs increases. As for PBTI, we find that V_{th} degradation of multiple NWs is remarkably similar to that of a conventional planar device for a wide range of stress conditions; however, the SS degradation in multiple NWs device is found to occur in two steps (Fig. 12). A plausible explanation of this distinctive two-step SS degradation has been proposed.

Characterization/Simulation of Multiple NWs Transistor

The devices used in this study are InGaAs GAA n-MOSFET (Fig. 1), with varying # of NWs and different EOT. The fabrication process is described in [1, 2] and the device dimensions are listed in Table 1. The transfer characteristics with different # of NWs show 2-3 orders of magnitude in ON/OFF ratio (Fig. 2a). The gate leakage, normalized by surface area (Fig. 2b, linear plot), compares very well with different # of NWs, implying remarkable uniformity in oxide thickness. The self-heating effect of the devices was characterized by AC conductance method [3]. The technique is based on comparing the output conductance (g_{ds}) at high frequency to that from the DC output characteristics, in order to extract the self-heating induced temperature within the NWs. The result shows that the heat dissipation becomes inefficient at higher NW density (Fig. 3). Indeed, the temperature could be as high as 420K for 19 parallel NWs!

For simulation, the transistor transfer characteristic (consisting of multiple NWs) is calculated by summing over single NW transfer characteristics [4]. The individual NW characteristics are calculated using a BSIM-based compact model (Table 2). A random V_{th} distribution (± 40mV width, see Fig. 7) is used to account for the variability of the individual NWs (Fig. 7). The overall transistor $V_{th,}$ and SS are extracted from this composite transfer characteristics.

Self-Heating Effect on Device Performance

Fig. 4a shows that I_{ON}, normalized by total cross sectional area of all the NWs, is not a constant, but increases as a function of the # of NWs. The counter-intuitive increase in I_{ON} correlates to a superlinear decrease in contact resistance (R_C) as a function NWs, see Fig. 4b. This dramatic drop in R_C (faster than expected for classical parallel channels, (i.e., $R_C(n\ NWs) = R_C(1\ NW)/n$) is traced to (i) the increasing number of NWs, as one expects, and (ii) the increasing self-heating with the number of NWs (Fig. 3), an assertion also confirmed by detailed analytical simulation. The increased temperature from self-heating allows more activated injection of carriers at the contact [5, 6] and thus reduces R_C .

Indeed, the self-heating is also reflected in exponential increase of I_{OFF} with #NW (Fig. 5a) in two ways, (i) the decrease in R_C, as in I_{ON}, and (ii) higher thermally activated current over the barrier (Fig. 5b). This also explains the

reduced ON/OFF ratio with increasing #NW (Fig. 6). Finally, the increased temperature with #NW raises V_{th} by ~100mV (Fig. 7), consistent with the theoretical prediction.

Effects of NW-NW Variability

Another important consequence of multiple NWs comes from the variability among NWs. For example, the distribution of V_{th} of NWs within a device is shown in Fig. 8a. Since the transistor transfer characteristics is determined by the summation of the individual NWs, *the SS of the overall device, will always be higher than that of the individual NWs*, and can be characterized by the spread of V_{th}, as well as SS of individual NWs. Assuming a normal distribution of V_{th} for the NWs, the expected spread in transistor V_{th} increases with the # of NWs, which in-turn is reflected in the degradation of the overall SS of the transistor as a function of #NW (Fig. 8b). The theoretical calculation which accounts both the variability of V_{th} of individual NWs as well as the temperature increase with #NW verifies such increase of SS.

Reliability of Planar vs. Multiple NW devices

To compare the reliability of multiple NW devices with that of a planar device, we focus on the PBTI stress and recovery characteristics of such devices. The measurement is done with an automated measure-stress-measure (MSM) setup at various stress conditions. The power law time exponent of $\Delta V_{th} (\propto t^n)$ lies within 0.13-0.15 (Fig, 9a). These values are robust across a broad range of devices, with varying #NW, oxide thickness, channel width, length, and voltage (Fig. 9a-d) and are similar to that of a planar device [7]. The voltage acceleration is optimistically lower than that of Si/high-*k*, see Fig. 10 [8].

In general, both trapping and defect generation are responsible for the shift in V_{th}. To resolve these components, we perform the stress-recovery measurements for multiple cycles, as shown in Fig. 11. First, ΔV_{th} data (Fig. 11a) indicates the presence of relaxation in these devices. However, the N_{IT} relaxation is confirmed by the SS recovery (Fig. 11b). Such partial N_{IT} relaxation is also reported for planar devices of similar technology [7]. For a deeper understanding of the N_{IT} characteristics, we analyze the time exponents of ΔSS (Fig. 12a) of *single* NW GAA transistor and find that the exponent is similar to that of a planar device. Remarkably, however, *the SS degradation and recovery of multiple NW devices are found to occur in two-stages* (Fig. 12b). We offer a plausible explanation based on V_{th} variability in NWs, as follows:

The SS degradation of multiple NW devices has two components (shown schematically in Fig. 12c): (i) the first component (ΔSS_1) is the traditional, (positive) component due to N_{IT} generation. (ii) The second component (ΔSS_2) is negative (SS decreases with time!) and is a reflection of the heterogeneous V_{th} shift of the individual NWs. More specifically, in the stress phase, an individual NW which has lower V_{th} degrades faster compared to those with higher V_{th}. This is because $\Delta V_{th} \propto f(V_{G\_Stress} - V_{th}) \times t^n$, and the NW having lower V_{th} is affected by higher effective stress $(V_{G\_Stress} - V_{th})$ than the NW with higher V_{th}. Thus, there is a heterogeneous shift in V_{th} that tends to *reduce* the *spread* in V_{th} as a function of time. This, in turn, is reflected in reduced SS (or, negative ΔSS_2). The net SS degradation ($\Delta SS = \Delta SS_1 + \Delta SS_2$) thus has a lower slope than that of ΔSS_1. Beyond a critical time (t_c), however, the threshold voltages of different NWs merge, and all the NW degrades with identical effective stress ($\Delta SS_2 = 0$). Therefore, for $t > t_c$ the overall SS is only dictated by the N_{IT} generation (ΔSS_1), and is reflected in the higher exponent. The interpretation is confirmed by the fact that the two-step process is reversed during recovery, as expected.

Conclusion

We have analyzed, for the first time, the impact of the number of NWs on the performance and reliability (PBTI) of InGaAs multiple NW GAA devices. The variability within NWs and self-heating significantly affect the characteristics of multiple channel GAA transistors that have no counterparts in classical planar transistors. Such variability and the heat dissipation must be carefully optimized to fully realize the dramatic scaling potential promised by surrounding-gate transistors.

Acknowledgement

We acknowledge Birck Nanotechnology Center for the fabrication and characterization facilities. Prof. Ye thanks Xinwei Wang and Prof. Roy G. Gordon from Harvard University for the technical support in device fabrication.

References

[1] J. J. Gu, Y. Q. Liu, Y. Q. Wu, R. Colby, R. G. Gordon, and P. D. Ye, in IEEE International Electron Device Meeting (IEDM), 2011 pp. 769-772.

[2] J. J. Gu, X. W. Wang, H. Wu, J. Shao, A. T. Neal, M. J. Manfra, R. G. Gordon, P. D. Ye, in IEEE International Electron Device Meeting (IEDM), 2012 pp. 633-636.

[3] Tu, R.H, C. Wann, J.C. King, P.-K. Ko and Chenming Hu, Electron Device Letters, IEEE , vol.16, no.2, pp.67,69, Feb. 1995

[4] B.J. Sheu, D.L. Scharfetter, P.-K. Ko, Min-Chie Jeng, Solid-State Circuits, IEEE Journal of , vol.22, no.4, pp.558,566, Aug 1987

[5] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices,* 3rd ed., Wiley-Interscience, 2006.

[6] Kaustav Banerjee, Ajith Amerasekera, Girish Dixit, and Chenming Hu, in IEEE International Electron Device Meeting (IEDM), 1997 pp. 115-118.

[7] Guangfan Jiao, Chengjun Yao, Yi Xuan, Darning Huang, P.D. Ye, Ming-Fu Li, Electron Devices, IEEE Transactions on , vol.59, no.6, pp.1661,1667, June 2012

[8] Jin Ju Kim, Moonju Cho, L. Pantisano, Ukjin Jung, Young Gon Lee, T. Chiarella, M. Togo, N. Horiguchi, G. Groeseneken, Byoung-Hun Lee, Electron Device Letters, IEEE, vol.33, no.7, pp.937,939, July 2012

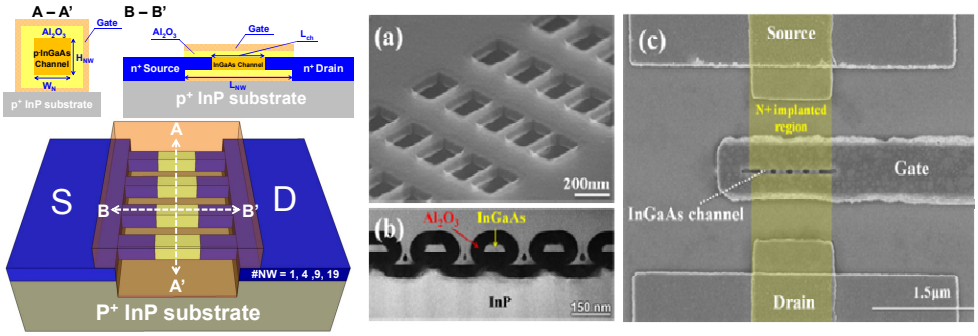

Fig. 1: (left) Schematic image of a NW GAA n-channel InGaAs MOSFETs having EOT = 1.7nm and 4.5nm, respectively. (a) SEM image of parallel NWs. (b) STEM image of the cross section of InGaAs NWs (A-A'). (c) SEM image of parallel InGaAs NWs (right). Taken from Ref. [1].

	Sample A IEDM 2011 [1]	Sample B IEDM 2012 [2]
Channel Material	$In_{0.53}Ga_{0.47}As$	$In_{0.65}Ga_{0.35}As$
L_{ch} (nm)	50-120	20-80
W_{NW} (nm)	30-50	20-35
H_{NW} (nm)	30	30
L_{NW} (nm)	200	200
Gate Oxide	10nm Al_2O_3	3.5nm Al_2O_3
EOT (nm)	4.5	1.7
# of NW	1, 4, 9, 19	4

Table. 1: Description of the two types of samples (different EOT) used in this study.

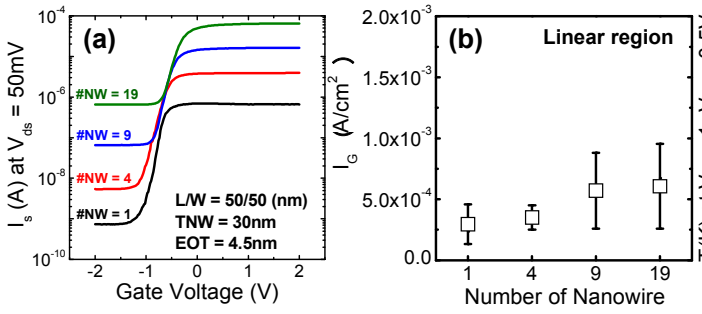

Fig. 2: (a) The transfer characteristics of the device with EOT =4.5nm depending on number of NWs. (b) Leakage current (I_G) depending on number of NWs. I_G is normalized by the total surface of NWs (i.e., I_G (A/cm$_2$) = $I_{G,n}$(A)/ (n × 2(W + H) × L_{ch})), where W and H are the width and height of the NWs, respectively). (The error bar reflects measurement of more than 20 samples).

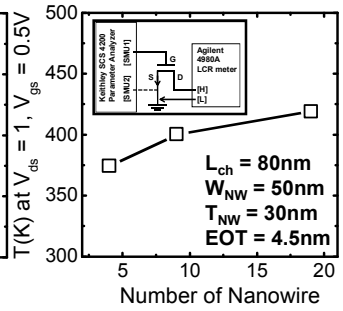

Fig. 3: Extracted temperature ΔT is $T_{lattice}$ – $T_{ambient}$. Self-heating effect (SHE) from NWs is correlated to the number of NWs in given area (NW density). The measurement setup for Self-heating effect (inset).

$$I_{DEV} = \sum_i I_{NWi}(V_{thi}, m_i, R_{Ci}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

$$I_{SUB} = \frac{A\mu_{EFF}C_{OX}(4W)}{L}\exp\left[\frac{q(V_{GS}-V_{th})}{mk_BT}\right] \quad \ldots\ldots\ldots (2)$$

$$I_{WEAK} = \frac{I_{SUB}[1.3I_{SUB}(V_{GS}=V_{th})]}{(I_{SUB}+1.3I_{SUB}(V_{GS}=V_{th}))} \quad \ldots\ldots\ldots\ldots (3)$$

$$I_{STRG} = \begin{cases} \frac{\mu_{EFF}C_{OX}(4W)}{L}\left[(V_{GS}-V_{th})V_{DS}-\frac{mV_{DS}^2}{2}\right] & \text{(Lin)} \\ \frac{\mu_{EFF}C_{OX}(4W)}{2mL}(V_{GS}-V_{th})^2 & \text{(Sat)} \end{cases} \quad \ldots (4)$$

$$I_{NW} = \frac{(I_{WEAK}+I_{STRG})V_{DS}}{(I_{WEAK}+I_{STRG})R_C+V_{DS}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Table. 2: The equations used in this work. (1) Total transfer characteristic (2) ~ (4) BSIM model. (5) Transfer characteristic for individual NW.

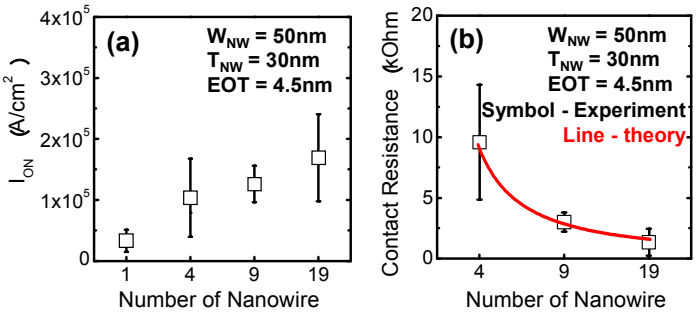

Fig. 4: (a) On current (I_{ON}) is normalized by the total cross section area of NWs (i.e., I_{ON} (A/cm^2) = $I_{ON,n}$(A)/(n × W × H)). I_{ON} increases with the # of NWs due to reduction of series resistance caused by self-heating effect.. (b) Contact resistance (R_c) extracted by conventional method and the Y-function method yield similar results.

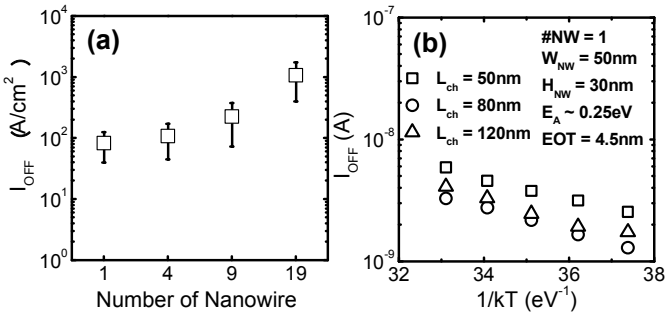

Fig. 5: (a) Off current (I_{OFF}) is normalized by the total cross section area of NWs (i.e., I_{OFF} (A/cm^2) = $I_{OFF,n}$(A)/(n × W × H)). I_{OFF} increases with # of NWs due to self-heating effect. (b) I_{OFF} vs. T defines activation energy (E_A ~ 0.25eV). The results demonstrate T-activated reduction in energy barrier at the contact and channel.

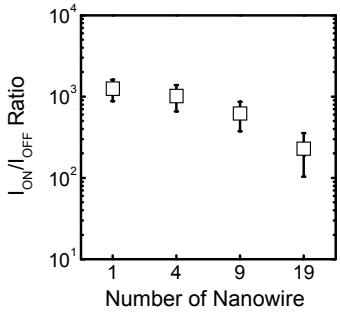

Fig. 6: The reduction in the ON/OFF ratio as a function of # of NW is correlated to the self-heating within the transistors.

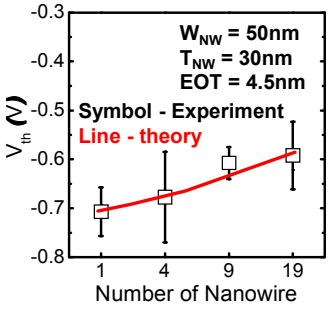

Fig. 7: Experimental data of V_{th} depending on the number of NWs are well-reproduced by simulation results.

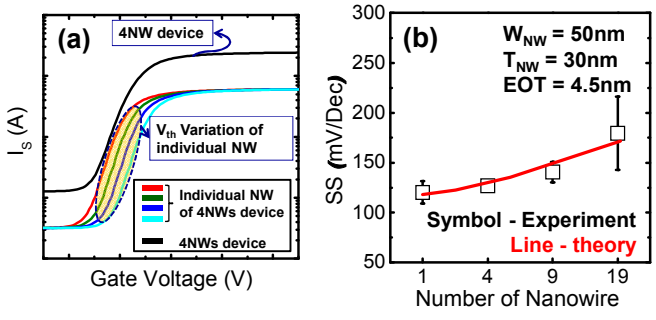

Fig. 8: (a) Schematic of the transfer characteristic of a 4NWs transistor (black line), obtained by summing up the 4 *individual* NW transfer characteristics. (b) Experimental results for SS as function of # of NWs is reproduced by BSIM-based simulation.

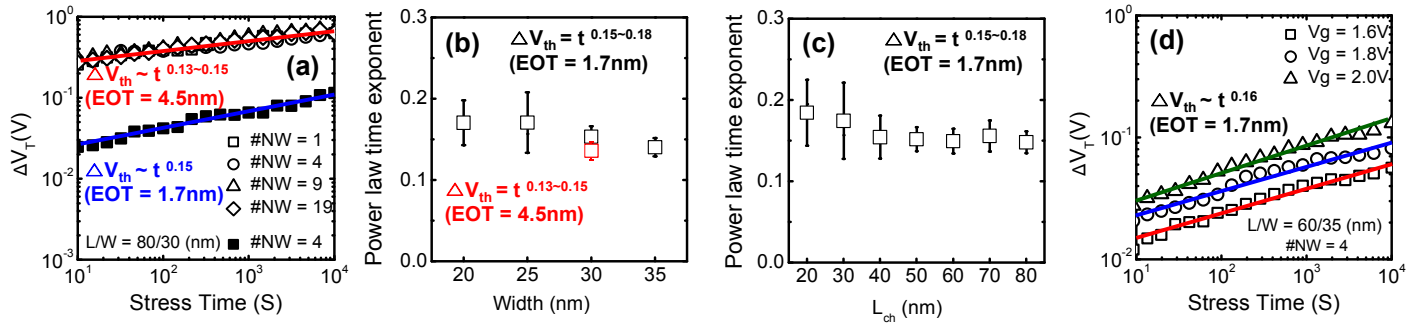

Fig. 9: (a) The time evolution of ΔV_{th} in PBTI (0-10^4s) under stress of 5V and 1.8V for the device with EOT = 4.5nm and EOT = 1.7nm, respectively. (b) Power-law time exponent of ΔV_{th} depending on channel width of NW extracted from ΔV_{th} (10^1-10^4s). (c) Power-law time exponent of ΔV_{th} depending on channel width extracted from ΔV_{th} (10^1-10^4s). (d) Time evolution of ΔV_{th} with different bias. Power-law time exponent of ΔV_{th} is robust about 0.13~0.15 for the device with EOT = 4.5nm and 0.15~0.18 for the device with EOT = 1.7nm, respectively.

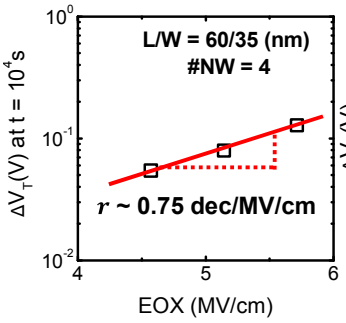

Fig. 10: Voltage acceleration factor (***r***) extracted from Fig.9 (d).

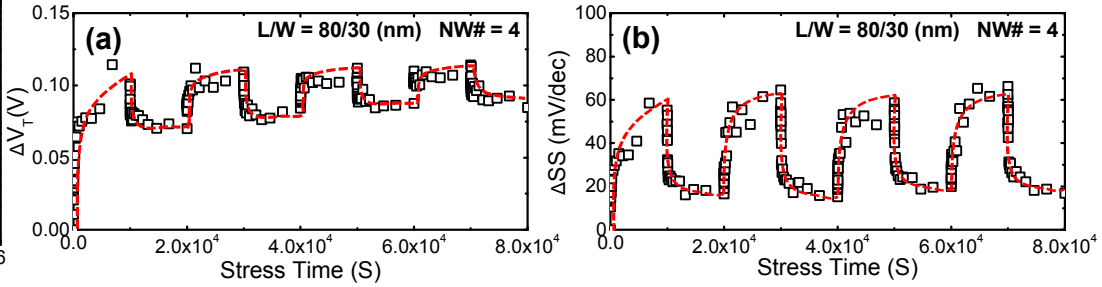

Fig.11: time evolution of (a) ΔV_{th} and (b) ΔSS under stress and recovery cycle offers insights into the dynamics of trapping and trap generation.

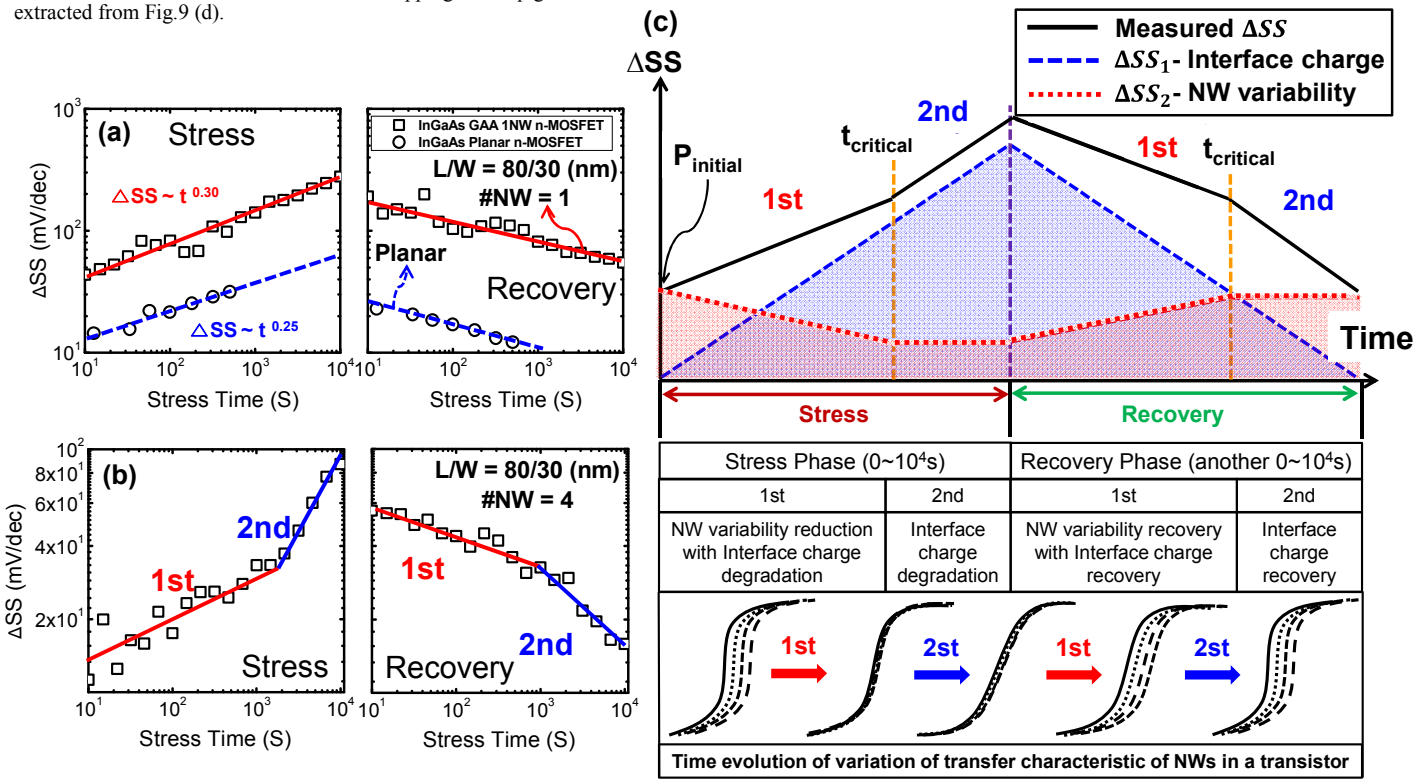

Fig.12: (a) Time evolution of ΔSS for a *single* NW device vs. a planar device. ΔSS degradation (0-10^4s) and recovery (10^4-2×10^4s) shows monotonous increasing and decreasing behaviors. (b) Time evolution of ΔSS for *multiple* NWs: Two step slopes are observed. (c) The two-step subthreshold slope (black line) can be interpreted as follows: In stress phase (left) before $t < t_{critical}$, the increase in N_{IT} generation (blue, dashed line) is compensated by the reduction in V_{th} variability (ΔSS_2, red dotted line). For $t > t_{critical}$, the subthreshold slope is defined exclusively by N_{IT} generation, with V_{th} of all NW evolving at the same rate. In recovery phase (right), the process is reversed, i.e., the self-compensation of N_{IT} (ΔSS_1, blue dashed line) and V_{th} recovery makes initial delay of SS slower, once the V_{th} recovery is complete, N_{IT} recovery completes the process.