# *Stacked Nanowires/Nanosheets GAA MOSFET From Technology to Design Enablement*

J.-Ch. Barbé[1], S. Barraud[1], O. Rozeau[1], S. Martinie[1], J. Lacord[1], P. Blaise[1], Z. Zeng[2], L. Bourdet[2], F. Triozon[1] and Y. Niquet[2]

[1] Univ. Grenoble Alpes, F-38000 Grenoble France
CEA, LETI, MINATEC, F-38054 Grenoble

[2] Univ. Grenoble Alpes, F-38000 Grenoble France
CEA, INAC/MEM, L_Sim, F-38054 Grenoble, France

E-mail: jean-charles.barbe@cea.fr

***Abstract*— GAA nanowires (NW) transistors are promising candidates for sub 10 nm technology nodes. They offer optimal electrostatic control, thereby enabling ultimate CMOS device scaling. Horizontally stacked they are a natural extension of today's mainstream technology. Considering enlarged NWs in Nanosheets (NS) allows to target the best compromise in power and performance for future applications. In this paper we will first briefly introduce the technology and then review what can bring advanced simulation focusing on both mobility and contact resistance. Then we will focus on devoted compact modeling fed by both TCAD capturing electrostatics of the Device and above mentioned advanced simulation for mobility. Finally we will demonstrate the capability of the model to capture actual hardware data as well as to benchmark the different architectures in competition down to 5 nm technology node.**

***Keywords: Nanowires, Nanosheets, Process Integration, NEGF, TCAD, Compact model, Leti-NSP, MHC, Benchmark.***

## I. INTRODUCTION

The vertically stacked NWs MOSFET architecture pushes further the scaling limits of the CMOS technology [1-4]. Due to its excellent immunity to the Short Channel Effects (SCE), undoped stacked-GAA MOSFET is one of the most promising candidate for sub-7nm CMOS technologies [5]. GAA quasi-cylindrical or square cross-section NanoWires (NW) can be used for an optimized gate control of the channel in order to minimize LSI power consumption. Meanwhile, GAA NanoSheet (NS) (e.g. *thin* and *wide* NW) MOSFETs can be proposed in order to increase the drive current for high performance applications. From a benchmark point of view addressing sub 10 nm technological nodes, these considered 3D architectures necessitate consolidated process assumptions based on specific technological module development and full assessment of the behavior of the devices; from electrostatics to transport properties. All these ingredients will then allow to build physics based compact models aiming to predict performances at circuit level and then to demonstrate which device is the most promising for specific applications or identifying which architecture allow the best compromise for a wide range of applications.

## II. NW/NS DEVICE FABRICATION IN BRIEF: PROCESS ASSUMPTION CONSOLIDATION

NW/NS process flow (Fig. 1) starts with the epitaxial growth of ($Si_{0.7}Ge_{0.3}/Si$) multilayers with typical layer thicknesses ranging from 7 to 12nm [6]. Then, individual and dense arrays of fins (60 nm high and 20 nm wide) are patterned to fabricate stacked-NWs/NSs FETs. The Sidewall Image Transfer technique (SIT) was used in order to meet the density targets of advanced nodes with fins pitch as low as 40 nm-pitch (Fig. 2, left). Dummy gates and spacers are then defined prior to the anisotropic etching of the (*SiGe/Si*) multilayers.

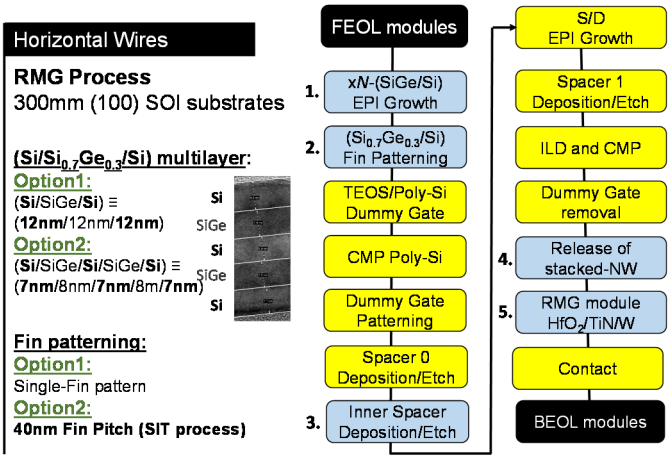


Fig. 1. Process flow of stacked NW/NS FETs including Inner spacers and SiGe:B raised-S/D with High density Fin patterning (FP=40nm) obtained by a SIT process. Steps numbered '1' to '5' are specific to stacked wires FETs as compared to FinFET devices.

Then, SiGe layers are partially etched selectively to Si ones with depth of the SiGe recess adjusted to match the thickness of future inner spacers. A nitride layer is deposited in the formed cavities and then etched back preserving Nitride spacers in between two adjacent wires.
$Si_{0.7}Ge_{0.3}$:B raised-S/D are then epitaxially grown (Fig. 2 right) and Si wires are fully released during the Replacement Metal Gate fabrication module. Gate stack is formed by a conformal $HfO_2$/TiN/W deposition while a specific attention is paid to the bottom wire (optimized Ω-shaped gate Si channel) to maintain a good electrostatic integrity. Having demonstrated the main specific technological process steps and then consolidated realistic process assumptions, benchmark will have to be fed by realistic electrical characteristics of the considered devices.

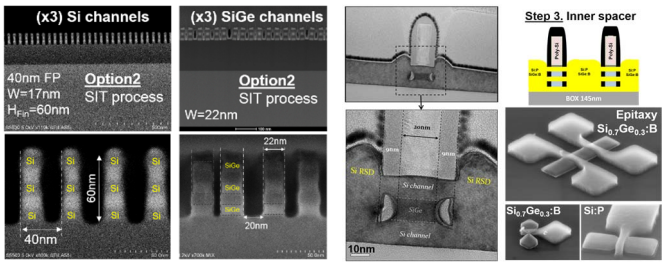

Fig.2. Cross-sectional TEM images of (Right) dense arrays of fins with 40 nm fin pitch and (Left) 2 stacked Si NWs channels with raised S/D ((Si0.7Ge0.3:B for p-FETs and Si:P for n-FETs) and inner spacers for optimized electrostatics.

## III. Transport Properties: Atomistic Simulation

Addressing some low-maturity technologies and targeting predictive benchmark, mobility models are one of the main ingredients needed to feed devoted compact model. Predictive simulation environment has then to be set-up.
Mobility models in NW must account for quantum confinement and carrier scattering mechanisms. In that way Non-Equilibrium Green's Function approach [7], is particularly adapted because it naturally captures quantum confinement and allows explicit accounting for the impact of scattering by phonons (PH), surface roughness (SR), and Remote Coulomb Scattering (RCS). Below are reported results obtained with Surface roughness disorder (and remote Coulomb charges in the gate dielectrics) explicitly included in the geometries. Gaussian autocorrelation function model is used for SR with correlation length ΛSR = 1.0 nm [8]. Carrier-phonon scattering is described by local, imaginary self-energies and the deformation potential theory. For electron–phonon scattering, intravalley acoustic phonon scattering is included (with deformation potential Dac = 14.6 eV) as well as the 3 f-type and the three g-type processes [9]. For hole-phonon scattering, a diagonal hole–phonon interaction is adopted with one single acoustic deformation potential (Dac = 16.5 eV) and one single optical deformation potential (DKopt = 15 eV/Å [10]).

The current through the device is computed in a self-consistent NEGF framework [11,12], on top of the effective mass approximation (EMA) or two bands k p model for electrons, and on top of the three bands k p model for holes. The NEGF equations are solved in a fully coupled mode-space approach (80 to 420 modes depending on the device cross section and on the band structure model), on a finite differences grid with step 2 Å.
In this framework, both electrons and holes long channel mobilities can then be computed considering different crystallographic orientations, a wide range of height H and width W (7 to 50nm) and chemical disorder when considering SiGe NWs. It naturally captures both quantum confinement and scattering mechanisms. Planar devices (DG) are also considered for H or W large enough compared to W or H. Mobilities are extracted from the slope of the quasi-Fermi level in the channel [13].
A simple long channel mobility model, compatible with compact modeling constraints, can then be derived while considering that the charge in both the inner and side channels is ruled by simple electrostatics and is thus proportional to the inner/side perimeters.
GAA MOSFETs' Mobilities can then be fairly interpolated as a linear combination of the computed square shape NW's mobility ($\mu_{SQ}$) and the DG computed mobility ($\mu_{DG}$). Proposed mobility models follow:

$$\mu_{W>H} \approx \frac{W-H}{W+H}\mu_{DG} + \frac{2H}{W+H}\mu_{SQ}$$

and

$$\mu_{H>W} \approx \frac{H-W}{W+H}\mu_{DG} + \frac{2H}{W+H}\mu_{SQ}$$

As an illustration, we compare in figure Fig. 3 the calculated mobilities with experimental data on rectangular Trigate devices with height H = 11 nm fabricated at CEA/LETI [6]. Considering Tri-gate devices, the usefull mobility model for compact modeling take the following form:

$$\mu^{TG}_{W>H} \approx \frac{W-H}{W+2H}\mu_{FDSOI} + \frac{3H}{W+2H}\mu_{SQ}$$

and

$$\mu^{TG}_{W>H} \approx \frac{H-W}{W+2H}\mu_{FDSOI} + \frac{3H}{W+2H}\mu_{SQ}$$

where $\mu_{FDSOI}$ stands for the mobility of the tri-gate device with width large compared to height.
Having captured the shape of the long channel mobility models based on calibrated predictive NEGF simulations, in a form which comply with constraints of compact modeling and associated SPICE simulation tools, compact model itself can be considered.

## IV. Compact Modeling: Leti-NSP

Here we will introduce a complete surface potential-based core model for GAA stacked-NW/NS MOSFETs (SPICE

like model) essential for the design of integrated circuits [14]. The modeling of this MOSFET architecture is a great challenge for the simulation of integrated circuits due to the variation of the Surface potential (SP) along the NW/NS perimeter.

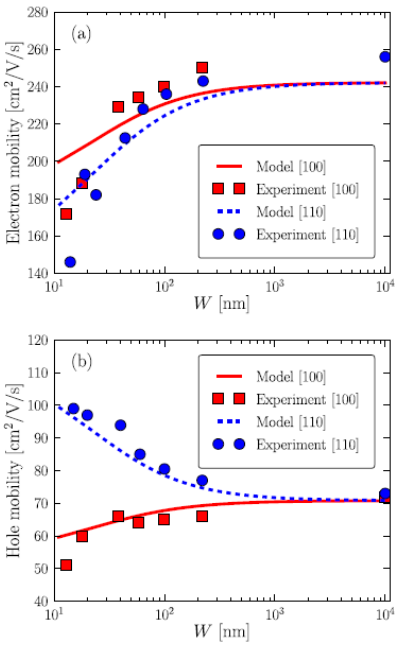


Fig. 3: Electron (a) hole (b) long channel mobility models as a function of device's width in rectangular Trigate devices (H = 11 nm), at carrier density n = $10^{13}$ $cm^{-2}$. The symbols are the experimental data, while the solid and dashed lines are the interpolated long channel mobility models

Then, an effective surface potential based on a partitioning of the structure cross-section was proposed to account for a large variety of cross-section geometries (Fig. 4) ranging from Double-Gate, Nanosheet, Square and cylindrical Nanowires. FinFET architecture is also naturally accounted for in the model. Then Leti-NSP can be used as a unique model allowing fair comparison of all the above mentioned architectures.

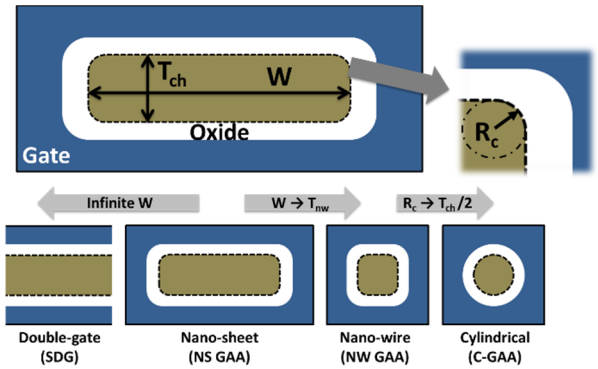


Fig. 4. GAA NS MOSFET architectures accounted for in Leti-NSP compact model

The key rely on proposing an unified way to compute the surface potential for all the architecture to account for. It was demonstrated that surface potential is the solution of the following differential equation [14]:

$$(x_g - x)^2 + B \cdot (x_g - x) = \Delta_f \cdot exp(x - x_n) \quad (1)$$

with $B = \frac{4 \cdot C'_{si}}{C'_{ox}}$, $C'_{si} = \frac{\varepsilon_{si}}{R}$, $C'_{ox} = \frac{\varepsilon_{ox}}{R \cdot \ln\left(1+\frac{T_{ox}}{R}\right)}$ and $\Delta_f = \frac{2 \cdot q \cdot n_i \cdot \varepsilon_{si}}{\phi_T C'^{\,2}_{ox}}$ and the notations in Table 1.

The strong inversion is controlled by $C'_{ox}$ only, while both capacitances $C'_{si}$ and $C'_{ox}$ are included in the weak inversion calculation. In addition, the total inversion charge used for the charge and the current calculation is given by:

$$Q_i = W_{eff} \cdot C'_{ox} \cdot \phi_T \cdot q_i. \quad (2)$$

$W_{eff}$ denotes the NS/NW perimeter and can be easily calculated considering the geometrical parameters. Thus, using a rigorous description of $W_{eff}$, $C'_{ox}$ and $C'_{si}$, all GAA MOSFET architectures can be modeled.

| | |
|---|---|
| $\phi_T$ | Thermal voltage |
| $T_{ox}$ | Gate oxide equivalent thickness |
| $W$ | Nano-sheet width |
| $L$ | Gate length |
| $R_c$ | Radius of nano-sheet corners |
| $T_{ch}$ | Nano-sheet silicon thickness |
| $C'_{ox}$ | Gate oxide capacitance per area unit |
| $C'_{si}$ | Body capacitance per area unit |
| $x_g$ | Gate to source voltage normalized to $\phi_T$ |
| $x_{nds}$ | Drain to source voltage normalized to $\phi_T$ |
| $x_n$ | Quasi-Fermi level normalized to $\phi_T$ |
| $x$ | Surface potential normalized to $\phi_T$ |
| $q_i$ | $(= x_g - x)$ Gate charge density normalized to $C_{ox}\phi_T$ |

Table 1. Notations used in Equation 1

Equation 1 is solved through providing an initial guess followed by two successive corrections based on Taylor's developments of second order. This analytical calculation of surface potential was validated on gate capacitance $C_{gg}$ by TCAD numerical simulations. An excellent agreement is achieved especially in moderate inversion (Fig 5) for both asymptotic cases (Double-Gate and cylindrical GAA MOSFETS).

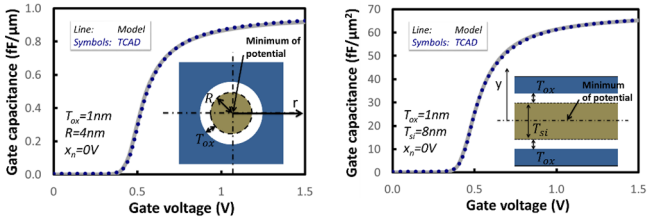


Fig.5. Gate capacitance vs gate voltage of cylindrical GAA (left) and Double-Gate (right) MOSFETS: comparison between TCAD and analytical solution of Eq. (4).

In Leti-NSP model, the quantum confinement (QC) was introduced thanks to a correction on the oxide capacitance accounting for the effect of carrier effective mass on the charge centroid position and with a classical correction of the flat-band voltage in order to include structural quantum confinement. In order to validate quantum confinement effects in our model, the software TB_SIM [15] has been

used to solve the Poisson-Schrödinger equations in GAA NS MOSFETs. $C_{gg}$ are demonstrated to be in perfect agreement over the full range of gate voltage and geometries.
Analytical calculation of drain current is performed by considering calculated SP at the drain and source sides ($x_d$ and $x_s$). The drain current is defined as

$$I_{ds} = -\frac{W_{eff}}{L} \cdot C'_{ox} \cdot \phi_T \cdot \int_{x_s}^{x_d} \left( q_i \cdot \frac{dx_n}{dx} \right) dx \qquad (3)$$

The model's core was validated using TCAD simulations for long channel transistor with a constant mobility [14]. In the same way to ref. [16], additional physical features like electrical field dependence on mobility, SCE, DIBL, channel length modulation effect, access resistances, velocity saturation, GIDL current, and gate tunneling current were implemented in the model's code.
Finally parameter extraction was performed based on experimental data demonstrating the ability of Leti-NSP to fairly reproduce these data (Fig. 6).

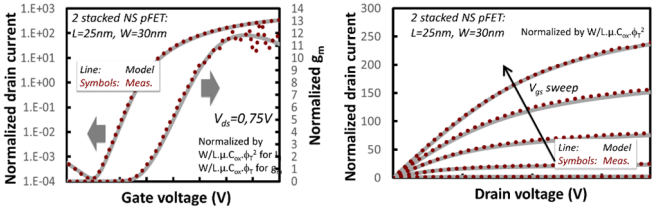


Fig.6: Model Hardware Correlation - Normalized drain current and its derivative vs gate voltage in saturation (Left) and Normalized drain current vs drain voltage (Right).

## V. Benchmarking down to 5 nm and conclusions

Comprehensive comparison of sub-10nm device architectures first needs a clear definition of the scaling rules. Actual pertinent design parameters for advanced nodes is no longer Gate Length but has to consider minimum distance between two consecutive transistor gates, namely 'contacted poly pitch' (CPP), and minimum pitch of the first metal level (M1 pitch) (Fig. 7, Left). Another critical parameter to be considered is the equivalent oxide thickness (EOT) with its associated supply voltage (Vdd) (Fig. 7 Right). These different parameters are extrapolated based on available data in letarature.

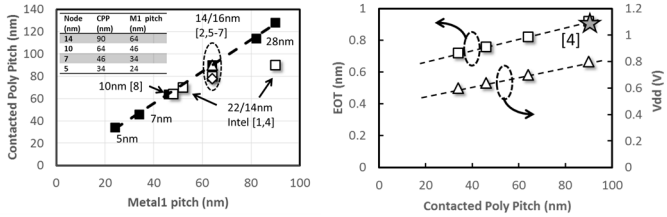


Fig. 7: Fig. 1. Contacted poly pitch versus metal1 pitch (Left) and Equivalent Oxide Thickness vs supply voltage (Right)

The technology performances can then be evaluated following the SPICE model calibration using experimental data, transport properties projections and TCAD simulations for electrostatic behavior. Benchmark can first be run considering simple design such as ring oscillators (RO) including pre-layout parameters such as parasitic effects of interconnections depending of the technological nodes.

As a conclusion of the predictive architectures evaluation down to 5nm technological node, RO based on FDSOI, FinFET and stacked-Gate All Around FET are considered. RO delays are reported on Figure Fig.8.
For 10nm node, FDSOI and FinFET are equivalent in speed at constant static power. However, FDSOI is more frugal in dynamic power thanks to lower load capacitances per stage (19% lower than FinFET).
For 7nm node FinFET needs stressors to target performances. From process point of view, the 5nm with a CPP as low as 34nm seems to be an insurmountable difficulty. At this node stacked-GAA with conservative 7nm design rules but increased density thanks to 3D integration, like monolithic 3D integration [15], seems to be the only viable pathway to reach targeted performances.

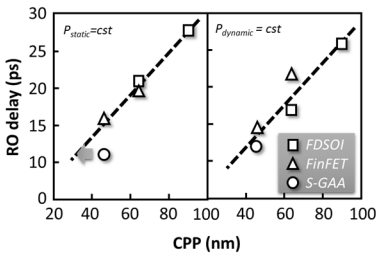


Fig. 9. RO delay versus nodes for all architectures. Symbols are RO simulations, dashed line is the trend of +35% on speed per node.


## Acknowledgment

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 688101 "SUPERAID7".


## References


[1] T. Ernst et *al*, IEDM, 10.1109/IEDM.2006.346955, 2006 ;
[2] L.K. Bera et *al*, IEDM, 10.1109/IEDM.2006.346841, 2006.
[3] C. Dupré et *al*, IEDM, 10.1109/IEDM.2008.4796805, 2008.
[4] E. Bernard et *al*, VLSI Technology, pp. 16-17, 2008.
[5] K. Tachi et *al*. IEDM.2010. 34.4.4.
[6] S. Barraud et *al*, IEDM.2016.17.6.1-4
[7] Z. Zeng et *al*, IEEE TED.2017.64.6.2485-2491
[8] S. M. Goodnick,et *al*, Phys. Rev. B, vol. 32, pp. 8171–8186, Dec. 1985
[9] C. Jacoboni et *al*, Rev. Mod. Phys., vol. 55, no. 3, pp. 645–705, Jul. 1983
[10] V. H. Nguyen et *al*, IEEE TTED.2014.61.9.3096–3102
[11] M. P. Anantram et *al*; Proc. IEEE, vol. 96, no. 9, pp. 1511–1550, Sep. 2008
[12] Y.-M. Niquet et *al*, J. Appl. Phys., vol. 115, no. 5, p. 054512, Feb. 2014
[13] L. Bourdet et *al*, J. Appl. Phys., vol. 119, no. 8, p. 084503, Feb. 2016
[14] O. Rozeau et *al*, IEDM.2016. 7.5.1-7.5.4
[15] TB_SIM simulator: Site web: http://inac.cea.fr/L_Sim/TB_Sim
[16] T. Poiroux et *al*, IEEE IEDM.2013.12.4.1-12.4.4.