

SOLID-STATE ELECTRONICS

An International Journal

Special Issue:
Extended papers selected from
EUROSOI-ULIS 2016

Guest Editors
Viktor Sverdlov
Siegfried Selberherr

EDITORS:

E. Calleja
Madrid

S. Cristoloveanu
Grenoble, France

Y. Kuk
Seoul, South Korea

A. Zaslavsky
Providence, RI

Available online at www.sciencedirect.com

ScienceDirect



ELSEVIER

Amsterdam • Boston • London • New York • Oxford • Paris • Philadelphia • San Diego • St. Louis

Solid-State Electronics

An International Journal

Volume 128 (2017)

Founding Editor
W. CRAWFORD DUNLAP

Editors

E. CALLEJA
Madrid

S. CRISTOLOVEANU
Grenoble, France

Y. KUK
Seoul, South Korea

A. ZASLAVSKY
Providence, RI

Special Issue:
Extended papers selected from EUROSOI-ULIS 2016

Guest Editors
Viktor Sverdlov
Siegfried Selberherr



Amsterdam • Boston • London • New York • Oxford • Paris • Philadelphia • San Diego • St. Louis

SOLID-STATE ELECTRONICS

FOUNDING EDITOR

DR W. CRAWFORD DUNLAP

EDITORS

E. CALLEJA

*Dept. of Electronic Engineering (ISOM), ETSI Telecommunication,
Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain*

S. CRISTOLOVEANU

*Grenoble INP, Ref: LPCS-SSE, 46 av. Felix Viallet,
F-38031 Grenoble Cedex 1, France*

Y. KUK

*Dept. of Physics & Astronomy, Seoul National University (SNU),
Gwanack-ku Silim-dong, Seoul, South Korea*

A. ZASLAVSKY

*Solid State Electronics, Brown University Engineering, 182 Hope Street,
Providence, RI 02912, USA*

EDITORIAL ADVISORY BOARD

P. AVOURIS, *Yorktown Heights, NY, USA*

G. BACCARANI, *Bologna, Italy*

L. COLOMBO, *Dallas, TX, USA*

J. G. FOSSUM, *Gainesville, FL, USA*

G. GHIBAUDO, *Grenoble, France*

G. GILDENBLAT, *Tempe, AZ, USA*

S. HALL, *Liverpool, UK*

T. HASHIZUME, *Saitama, Japan*

S. HWANG, *Yongin-si, Gyeonggi-do, South Korea*

C. JAGADISH, *Canberra, ACT, Australia*

K. KAKUSHIMA, *Midori-ku, Yokohama, Japan*

S. KELLER, *Santa Barbara, CA, USA*

J.-H. LEE, *Gwanag-Gu, Seoul, South Korea*

C. MCANDREW, *Tempe, AZ, USA*

C. MCCONVILLE, *Coventry, UK*

S. RINGEL, *Columbus, OH, USA*

J. SCHMITZ, *Enschede, Netherlands*

A. SEABAUGH, *Notre Dame, IN, USA*

M. S. SHUR, *Troy, NY, USA*

A. WAAG, *Braunschweig, Germany*

Author inquiries: You can track your submitted article at <http://www.elsevier.com/track-submission>. You can track your accepted article at <http://www.elsevier.com/trackarticle>. You are also welcome to contact Customer Support via <http://support.elsevier.com>.

Orders, claims, and journal inquiries: please contact the Elsevier Customer Service Department nearest you:

St. Louis: Elsevier Customer Service Department, 3251 Riverport Lane, Maryland Heights, MO 63043, USA; phone: (877) 8397126 [toll free within the USA]; (+1) (314) 4478878 [outside the USA]; fax: (+1) (314) 4478077; e-mail: JournalCustomerService-usa@elsevier.com

Oxford: Elsevier Customer Service Department, The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK; phone: (+44) (1865) 843434; fax: (+44) (1865) 843970; e-mail: JournalsCustomerServiceEMEA@elsevier.com

Tokyo: Elsevier Customer Service Department, 4F Higashi-Azabu, 1-Chome Bldg, 1-9-15 Higashi-Azabu, Minato-ku, Tokyo 106-0044, Japan; phone: (+81) (3) 5561 5037; fax: (+81) (3) 5561 5047; e-mail: JournalsCustomerServiceJapan@elsevier.com

The Philippines: Elsevier Customer Service Department, 2nd Floor, Building H, UP-Ayalaland Technohub, Commonwealth Avenue, Diliman, Quezon City, Philippines 1101; phone: (+65) 6349 0222; fax: (+63) 2 352 1394; email: JournalsCustomerServiceAPAC@elsevier.com

Publication information: *Solid-State Electronics* (ISSN 0038-1101). For 2017, volumes 115–126 (12 issues) are scheduled for publication. Subscription prices are available upon request from the Publisher or from the Elsevier Customer Service Department nearest you or from this journal's website (<http://www.elsevier.com/locate/sse>). Further information is available on this journal and other Elsevier products through Elsevier's website (<http://www.elsevier.com>). Subscriptions are accepted on a prepaid basis only and are entered on a calendar year basis. Issues are sent by standard mail (surface within Europe, air delivery outside Europe). Priority rates are available upon request. Claims for missing issues should be made within six months of the date of dispatch.



Contents lists available at ScienceDirect

Solid-State Electronics

journal homepage: www.elsevier.com/locate/sse

Editorial

Special Issue of Solid-State Electronics, dedicated to EUROSOI-ULIS 2016



The current special issue of Solid-State Electronics includes 29 extended papers presented at the 2016 Second Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS 2016) held in Wien, Austria, on January 25–27, 2016. The papers entering to the special issue have been selected by the EUROSOI-ULIS 2016 Technical Program Committee based on the excellence of abstracts submitted and presentations delivered at the conference. In order to comply with the high standards of Solid-State Electronics the manuscripts went through the standard reviewing procedure.

The papers cover the vast variety of topics including fabrication, characterization and modeling of the cutting-edge fully depleted SOI, FinFET and nanowire-based devices, tunnel FETs, and novel emerging devices for memory and logic applications.

The invited paper by H. Stork and G. Hossey from ON Semiconductor introduces an exciting field of commercial SOI technology and provides an extensive review of its use for power management in automotive and industrial applications.

The problem of an accurate extraction of the effective oxide thickness and the work function in extremely scaled 14 nm fully depleted SOI devices is addressed in the paper by B. Mohamad et al.

Simulation study of silicon ultra-scaled gate all around nanowire FETs including the line-edge roughness by 3D Finite Element Monte Carlo Simulations is performed by M. Elmessary et al.

Another aspect of variability due to electron scattering on randomly positioned impurities and the problem of self-averaging of impurity-limited resistance in quasi-one dimensional nanowires is addressed in the paper by N. Sano.

Drain current variability in 28 nm n-MOSFETs in both linear and saturation regimes is studied in details by T. Karatsori et al.

Continuing the line of extremely scaled devices, the computational benchmarking of III–V semiconductors tunnel FETs against the 10 nm CMOS FinFET technology is performed by S. Strangio et al. at an example of basic arithmetic circuits.

An extensive comparison of the analog performance of fabricated line tunnel FETs with respect to other tunnel FET- and MOSFET-based architectures is reported in the paper by P. Agopian et al.

Confinement orientation effects on source to drain tunneling are addressed by C. Medina-Bailon et al.

Process modules for GeSn nanoelectronics with high Sn-concentration reported by C. Schulte-Braucks et al. provide a base for further optimization of GeSn FETs and novel tunnel FET devices.

An experimental evaluation of the analog performance of n- and p-type SOI nanowire FETs in a broad temperature range from room temperature down to 100 K is performed by B. Paz et al.

The influence of temperature on drain-induced barrier lowering in ultra-thin box and body fully depleted SOI MOSFETs is analyzed by A. Pereira et al. based on extensive experimental data, numerical simulations, and analytical modeling.

Continuing the line of exploiting materials with improved transport characteristics for the next generation transistors, the influence of layout effects including mechanical stress in 14 nm UTBB FDSOI SiGe channel p-MOSFETs is investigated and characterized both experimentally and by modeling in the paper by R. Berthelon et al.

Fabrication process and conduction mechanisms in back-gated InGaAs-on-insulator lateral N⁺NN⁺ junctionless MOSFETs are reported by H.J. Park et al.

DC and first RF characterization of InGaAs replacement metal gate n-FETs on SiGe-on-insulator FinFETs fabricated by 3D monolithic integration is presented by V. Deshpande et al.

A new MOSFET threshold voltage characterization method suggested by D. Tomaszewski et al. allows elimination of the channel current effects by using junction capacitance measurements.

D. Boudier et al. describe how low frequency noise measurements can provide an essential information on the quality of the gate oxide and the silicon film of tri-gate SOI FinFETs and thus can be used as a non-destructive characterization tool at sub-10 nm technology nodes.

A systematic method for electrical characterization of random telegraph noise in MOSFETs is presented by C. Marquez et al. aiming to facilitate massive on-wafer characterization of noise MOSFET characteristics.

In the work by B. Kazemi Esfeh et al. it is shown that a trap-rich layer in trap-rich high resistivity SOI wafers with thinner BOX does not alter RF characteristics of the MOSFET transistors while it allows for improved thermal properties thus making the technology viable for SoC applications.

The paper by C. Jungemann et al. continues along the line of frequency-dependent phenomena in MOSFETs and provides an efficient numerical simulation tool capturing plasma oscillation effects in silicon MOSFETs used for THz detection.

Efficient approaches for accurate high-performance TCAD simulations are reported by the CDL Laboratory for high-performance TCAD at TU Wien: V. Simonka et al. describe the anisotropic inter-polarization method of SiC oxidation growth rates for three-dimensional simulation, while P. Manstetten et al. present a computationally efficient framework to evaluate the neutral flux in high aspect ratio structures during three-dimensional plasma etching simulations.

T. Baldauf et al. address by simulations the effect of mechanical stress on tunneling and the transfer characteristics in two independently gated Schottky junctions of Schottky barrier-based reconfigurable nanowire transistors.

Continuing the reconfigurable FETs topic, C. Navarro et al. highlight the main advantages, limitations, and ways to improve the performance by employing TCAD for careful optimization and benchmarking reconfigurable field effect transistors.

Concerning new applications and devices employing advanced FD-SOI technology, a novel scalable three-dimensional single-photon avalanche diode pixel with high fill factor is proposed and extensively studied by simulations in the work of M. Vignetti et al.

A novel electrostatic discharge protection based on reconfigurable ultra-thin film gated diode merged NMOS is proposed and fabricated with 28 nm UTBB FD-SOI high-k metal gate technology in the work by S. Athanasiou et al.

Another new sharp-switching band-modulation back-gated device fabricated in advanced FD-SOI technology is presented by H. El Dirani et al.

Finally, recent developments in promising devices and concepts beyond SOI MOSFETs for future non-volatile memory and logic applications are given in the paper by A. Grossi et al., where electrical characterization and modeling of one-transistor/one-resistor resistive random access memory arrays with amorphous and polycrystalline HfO_2 as an active medium is presented.

A substantial reduction of the switching current in spin-transfer torque due to inverse magnetostriction induced switching barrier

lowering is predicted by Y. Takamura et al. A magnetic random access memory cell based on the magnetic tunnel junction surrounded by a piezoelectric gate operates at low voltages, which promises a large reduction of the write energy without degradation of the thermal stability.

We would like to express our gratitude and thank all the authors for accepting the invitation and preparing the excellent quality manuscripts within a tight time frame. We would also like to acknowledge all the reviewers for their time spent to provide highly professional comments to the authors, which indisputably helped boosting the overall paper quality to meet the highest publication standards adopted by Solid-State Electronics. Our special thank is dedicated to the Solid-State Electronics Editor Sorin Cristoloveanu for his constant encouragements, support, and help.

Guest Editor
Viktor Sverdlov

Guest Editor
Siegfried Selberherr
Institute for Microelectronics, TU Wien, Vienna, Austria
E-mail addresses: sverdlov@iue.tuwien.ac.at (V. Sverdlov),
selberherr@iue.tuwien.ac.at (S. Selberherr)

Available online 18 October 2016



SOI technology for power management in automotive and industrial applications



Johannes M.C. Stork IEEE Fellow, George P. Hosey*

ON Semiconductor, Phoenix, AZ, USA

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Semiconductor on Insulator
SOI
Power management
Automotive electronics
Industrial electronics

ABSTRACT

Semiconductor on Insulator (SOI) technology offers an assortment of opportunities for chip manufacturers in the Power Management market. Recent advances in the automotive and industrial markets, along with emerging features, the increasing use of sensors, and the ever-expanding “Internet of Things” (IoT) are providing for continued growth in these markets while also driving more complex solutions. The potential benefits of SOI include the ability to place both high-voltage and low-voltage devices on a single chip, saving space and cost, simplifying designs and models, and improving performance, thereby cutting development costs and improving time to market. SOI also offers novel new approaches to long-standing technologies.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Today's modern cars contain an enormous amount of electrical wiring and electronic circuits used for a variety of applications, and the trend is for even greater integration and complexity. Likewise, the industrial segment continues to grow and expand with the “Internet of Things”. With each new advance and each new opportunity come new challenges. In power management, these challenges include the ability to fabricate high-voltage power electronics and the low-voltage MOS devices required to control them without the additional space and cost of multi-chip and co-packaged solutions. Semiconductor-on-Insulator technology offers the advantage of integrating both high-voltage devices and low-voltage circuitry on the same chip [1]. The monolithic integration of vertical power devices together with control circuits increases the functionality and minimizes the size of the chips [2]. SOI also offers improved performance as a result of lower parasitic leakage and capacitance, leading to improved model accuracy and fewer development iterations, resulting in reduced development costs and quicker time to market, as well as improved reliability. In addition, SOI provides new ways to improve older technologies and opens paths to novel process integrations.

2. Automotive and industrial applications

The automotive segment has enjoyed one of the fastest growth rates of any large segment in the worldwide chip market, averaging 8% annual growth between 2002 and 2012. The average automobile currently has around \$350 in semiconductor content with even higher dollar content found in hybrid and luxury vehicles [3]. The vast majority of this content is in microcontroller units, analog, and power. (Fig. 1 shows a breakdown of the semiconductor content in various automobile families by family and by semiconductor type.)

Key trends in automotive semiconductors include:

- Fuel Economy & Emissions Reduction
- Active Safety & Autonomous Driving (autonomous cruise control, collision warning/avoidance)
- Vehicle Electrification (Hybrids, plug-ins, electric cars)
- Connectivity (Bluetooth, in-car WiFi)
- Light Emitting Diode (LED) Lighting

ON Semiconductor offers a vast array of solutions in these areas, including

- Power Management, Igniters, Application-Specific Integrated Circuits (ASICs), Application-Specific Standard Products (ASSPs)
- Image Sensors, Communications, Sensor Interface

* Corresponding author.

E-mail address: George.Hosey@onsemi.com (G.P. Hosey).

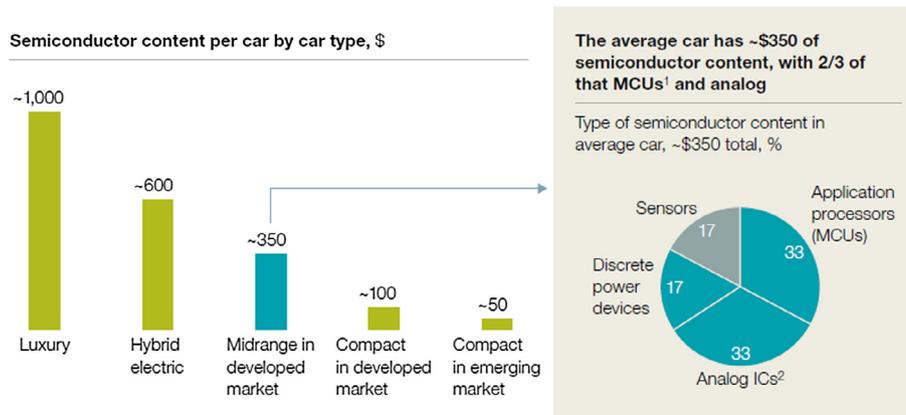


Fig. 1. Automobile semiconductor content [3].

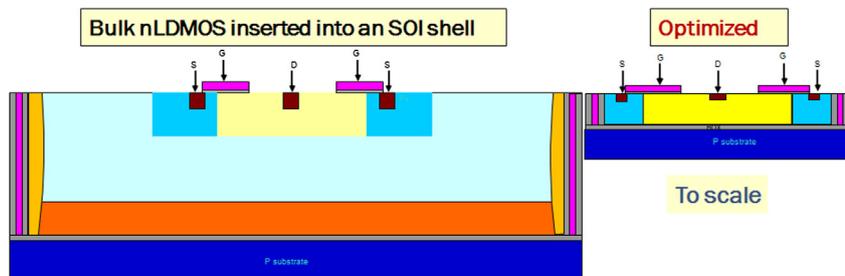


Fig. 2. Size comparison of a bulk NLD MOS device inserted into an SOI shell to the same device once fully optimized for an SOI platform.

- Insulated Gate Bipolar Transistors (IGBTs), High Efficiency FETs, Power Integrated Modules (PIMs), Gate Drivers, Intelligent Power Modules (IPMs)
- Wireless Charging Circuits, In-Vehicle Networking
- LED Drivers, Motor Control, MOSFETs
- Smart Passive Sensors, Protection

The industrial segment also offers incredible growth opportunities in the coming years courtesy of trends in

- The “Internet of Things” – Home security, smart homes, smart office buildings, building security and surveillance
- High-efficiency power supplies and motors
- The seemingly ubiquitous sensors we encounter every day.

for which ON offers solutions that include

- Power modules (IPMs, PIMs)
- Smart Passive Sensors, Image sensors, ISPs, DSPs
- Wired/wireless communications, motor drivers, and LED drivers
- As well as a host of comprehensive power management solutions.

3. Automotive BCD process integration

Bipolar-CMOS-DMOS (BCD) is a key technology in power integrated circuits, and thus, has become a key technology in automotive power applications. The marriage of BCD with SOI wafer technology allows the same chip to carry both low-voltage control circuitry and high-voltage DMOS power devices. Devices with ratings exceeding 35 V are common in 12 V automotive systems, while 120 V rated devices are seen in the emerging 42 V battery systems as designers and manufacturers look for robust “load

dump” capability. Even higher voltages, 500 V or more, are required for some applications, such as ignition IGBTs [1]. As automotive manufacturers continue to seek the simplest, lowest-cost solutions, SOI technologies, with the ability to house control and driver circuits in a single chip will become more and more a technology of choice.

SOI offers additional benefits as well. Wafer costs for SOI substrates are considerably higher than those for traditional bulk silicon wafers, however, when the layout and integration are fully optimized, SOI offers chip size (silicon area) savings, as shown in Fig. 2, and mask count savings as well.

As shown in Fig. 3, effective savings of 10–30%, depending on the SOI wafer fabrication techniques employed, are achieved when SOI-based processing is used in place of HV-bulk-Si technology.

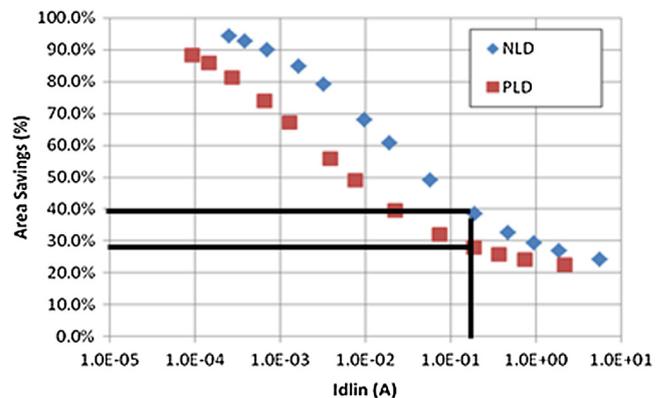


Fig. 3. Area savings for optimized NLD MOS and PLD MOS devices in SOI platforms compared to bulk platforms, by device size (current capability).

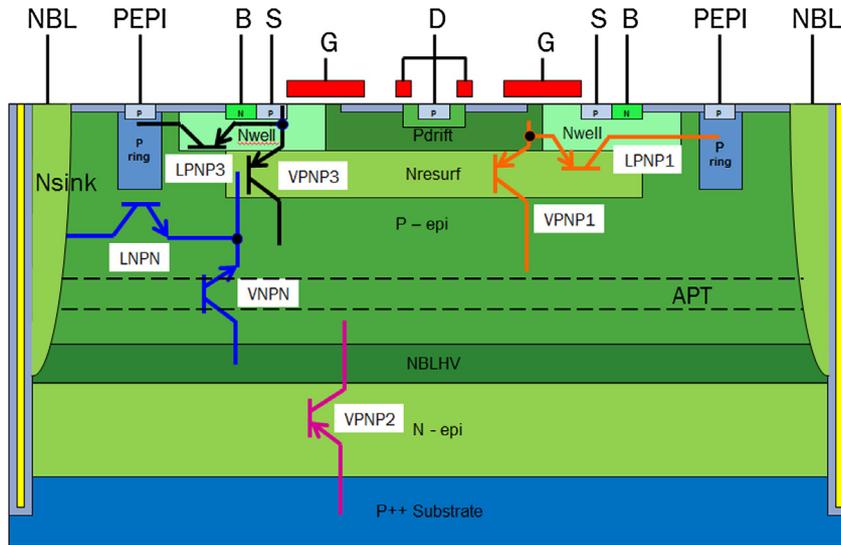


Fig. 4. Parasitic junctions in a typical bulk DMOS device.

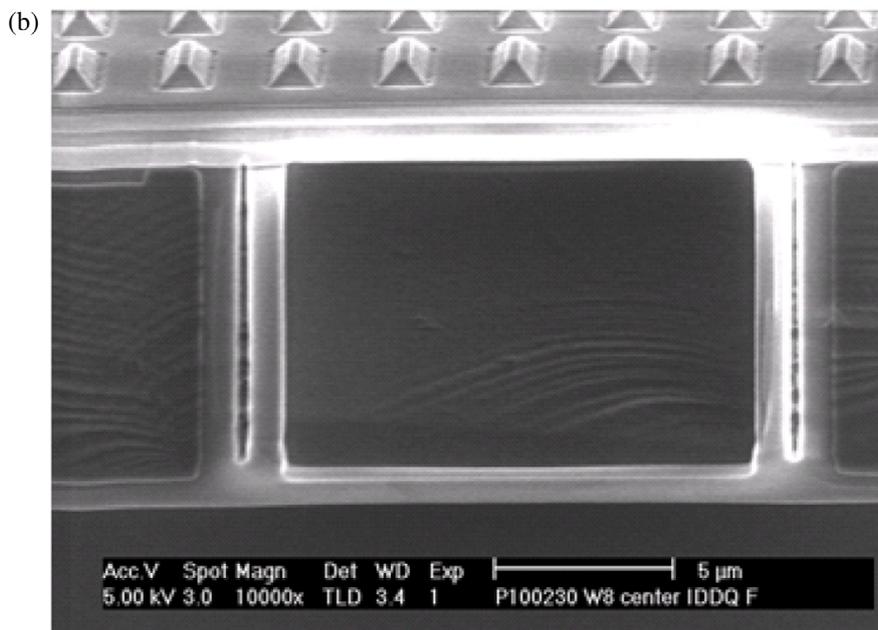
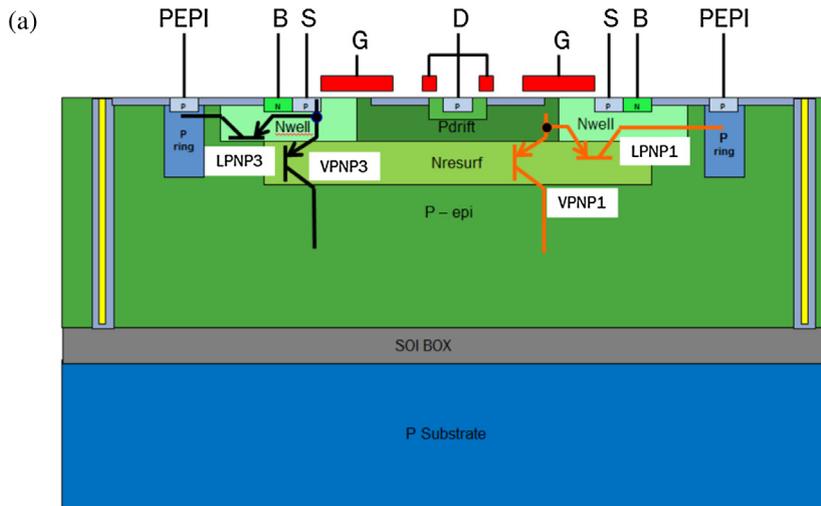


Fig. 5. (a and b) Parasitic junctions in a typical SOI DMOS device and cross-section.

SOI technology also brings improved IC performance through reductions in parasitic leakage currents and parasitic capacitances. Every set of PNP or NPN junctions creates a parasitic transistor with some inherent gain and an associated, resulting, leakage current. Typical parasitics for a bulk device are shown in Fig. 4.

The same is, of course, also true for devices in SOI technology platforms, however, the sheer number of such parasitic junctions is greatly reduced by the SOI integration. Fig. 5 shows the parasitic junctions for the same device when fabricated in an SOI platform.

It is also worth noting that, for each of these parasitic junctions, not only is a parasitic transistor created, but a parasitic junction capacitance as well.

While any one of these parasitic junctions might not greatly impact device or circuit performance, taken together, they can greatly affect the results, so the more parasitic junctions are present, the greater the risk.

Parasitics lead to greater model complexity as modeling teams work to identify and characterize each and every parasitic path. With this increased complexity comes greater inherent inaccuracy of the models. Inaccuracies in the models, in turn, result in addi-

tional design/development cycles which prolong development, increase development costs, and delay release to market.

Conversely, SOI technology offers simpler, more accurate models, leading to fewer design cycles, shorter development times, reduced development cost, and quicker time to market (and profitability).

SOI is not, of course, without its integration challenges as well. One of the primary challenges in using SOI is the accumulation and dissipation of heat within the chip and especially the power devices. As semiconductor devices shrink, heat dissipation becomes ever more challenging. While the task is already difficult in bulk silicon devices, it takes on even greater difficulty in a silicon-on-insulator (SOI) device. In bulk silicon, the heat can travel through the silicon to the package lead frame. In SOI devices, however, the buried oxide and oxide trenches surrounding the device act as heat insulation, trapping the generated heat within the SOI tub. The thermal resistance of SOI devices has been seen to be as much as an order of magnitude higher than that for bulk silicon devices [4].

Fig. 6 shows a comparison of SOI vs bulk silicon temperature response. It is seen that there is a significant difference in self-heating between SOI and bulk silicon, and that this difference is greatest for short power transients because the initial temperature rise in the SOI devices is more rapid than that for the bulk devices [5].

Fig. 7 shows a modeled temperature response for BOX thicknesses ranging from 0.5 μm to 1.5 μm . As would be expected, the self-heating for the devices increases with the thickness of the BOX layer as the thicker BOX increases the thermal resistance of the device.

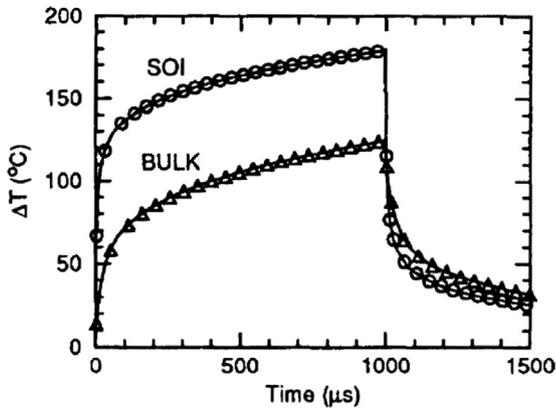


Fig. 6. Temperature rise and decay in SOI and bulk-Si LDMOS transistors following an application of 90 W/mm², 1 ms power pulse [5].

4. Motor control using IGBT's

SOI technology offers tremendous opportunities for device performance in Insulated Gate Bipolar Transistors (IGBTs) as well.

IGBTs have enjoyed popularity for automotive ignition applications because they combine the fast-switching capability of a MOSFET with the low conduction losses (high efficiency) of a bipolar junction transistor (BJT).

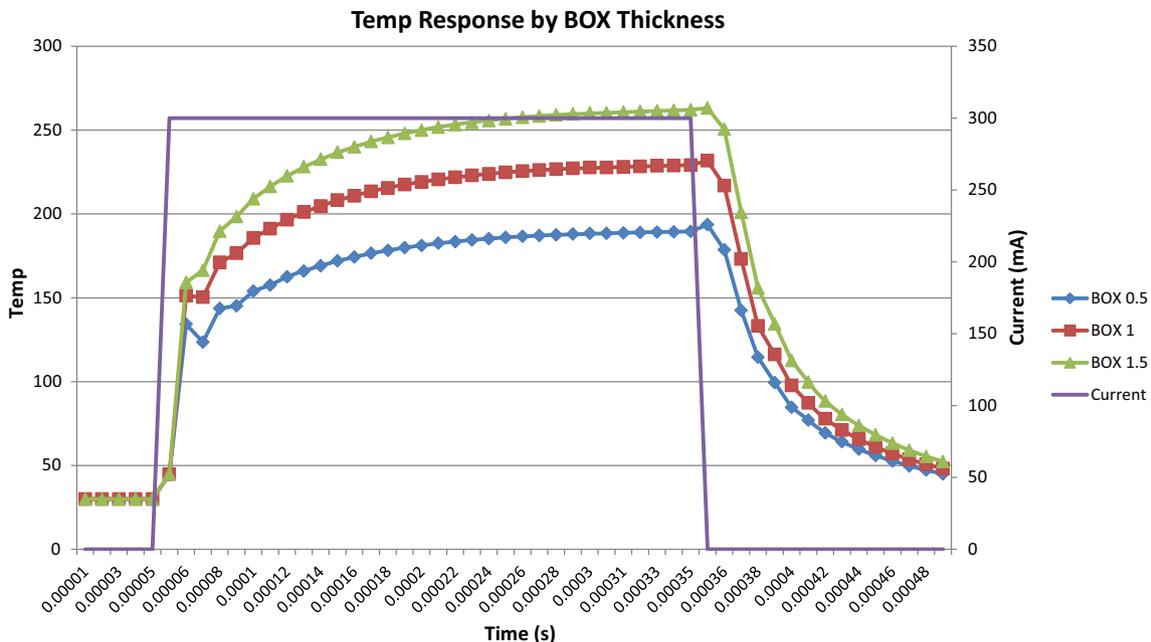


Fig. 7. Modeling results for temperature response of SOI object with various BOX thicknesses [6].

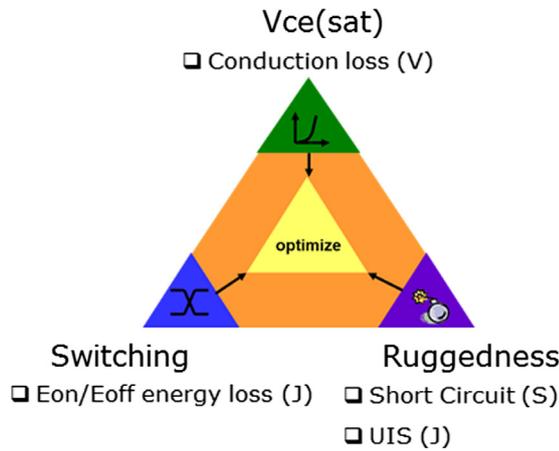


Fig. 8. The IGBT triangle.

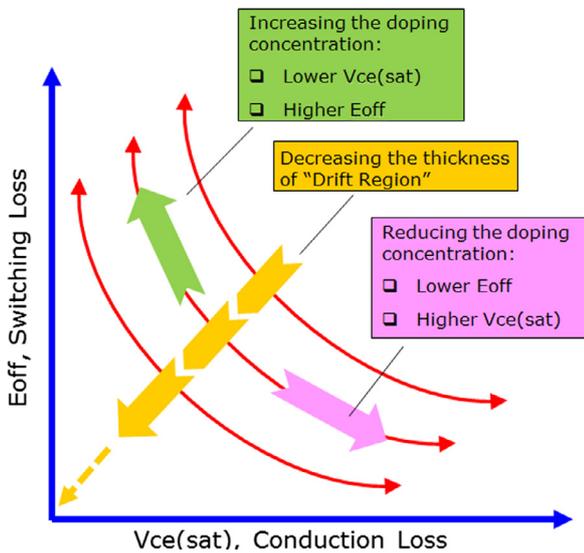


Fig. 9. Switching loss/conduction loss trade-off.

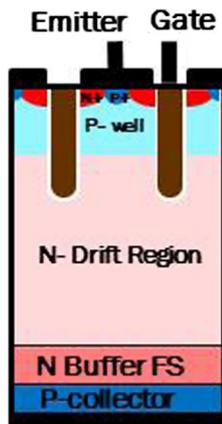


Fig. 10. Cross-section of N-channel trench IGBT with drift region labeled.

Until recently, automotive systems other than the ignition module had no need for the high voltage performance of IGBTs, but new developments in both conventional and hybrid autos are changing that. High-Intensity Discharge (HID) headlamps and super-accurate direct fuel-injection systems have become common in

today's cars and commercial vehicles, such as buses and trucks. Their associated electrical drives and controls operate at voltages above 100 V. In addition, hybrid/electric vehicles (HEVs), which require efficient power control at voltages close to 1000 V, will drive demands for new types of automotive IGBTs [7].

IGBT design and performance are a study in trade-offs. Fig. 8 shows what is called the "IGBT Triangle". The points of the triangle are the three key aspects of IGBT performance – Switching Losses (efficiency), Conduction Loss (Vce), and Ruggedness (energy absorption ability as displayed in short circuit and UIS testing).

Efforts to improve one aspect of performance (one point of the triangle) diminish one or both of the other aspects. Fig. 9 illustrates an example of such a trade-off. As shown, increasing the dopant concentration results in reduced conduction loss, but it also increases switching loss. Conversely, decreasing the dopant concentration improves switching loss while bringing about an unwanted increase in Vce.

Decreasing the thickness of the "drift" region, however, offers the opportunity to reduce both the switching loss and conduction loss. The drift region (visible in Fig. 10) is an epitaxy layer through which minority carriers must travel when the device is in operation. The drift region is typically quite thick and very lightly-doped relative to other regions of the device, so it is not surprising that this region of silicon contributes significantly to the "on" resistance of the device, and thus the conduction loss (also known as Vceon using the bipolar nomenclature, or Vdson using the MOS nomenclature).

Process options for fabricating these thinner drift regions include the use of glass carrier wafers and the use of SOI silicon carrier wafers. The silicon wafer carrier method is illustrated in Fig. 11.

While both the glass carrier method and the SOI wafer carrier method will work to achieve the desired thinning of the drift region, the glass carrier process encounters two serious challenges

- The glue attaching the glass carrier to the silicon limits the backside processing temperature to less than 300 °C.
- The glass and the silicon have different thermal expansion coefficients, resulting in wafer bow and/or warp issues, or other manufacturing issues such as equipment handling.

The SOI wafer carrier method overcomes both of these challenges as the BOX does not limit processing temperature like glue does, and both the device and carrier wafers are silicon, so the coefficients of thermal expansion are perfectly matched.

5. Image sensors

Image sensors are a rapidly-growing segment of the automotive electronics market. As today's (and tomorrow's) cars become more and more sophisticated and incorporate collision avoidance and autonomous driving, the need for image sensors will continue to expand. Today's cars have the potential for 20 or more sensors per vehicle for back-up cameras, blind spot detection, collision warning, and numerous other applications as displayed in Fig. 12.

Like any emerging market, automotive image sensors present both challenges and opportunities. The rapid adoption of LED lighting in both vehicles and traffic signs has made the problem of LED flicker in video viewing and machine vision a high priority for automotive OEMs.

ON Semiconductor has positioned itself as a leader in LED Flicker Mitigation (LFM) technology with the release of AR0231AT, a CMOS image sensor that eliminates high frequency LED flicker from traffic signs and vehicle LED lighting and allows Traffic Sign Reading algorithms to operate in all light conditions. AR0231AT

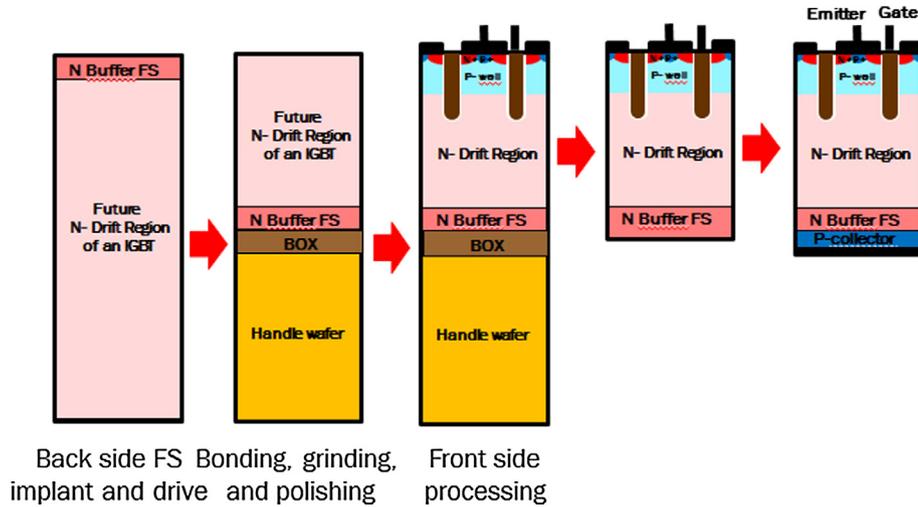


Fig. 11. SOI silicon wafer carrier IGBT process flow.

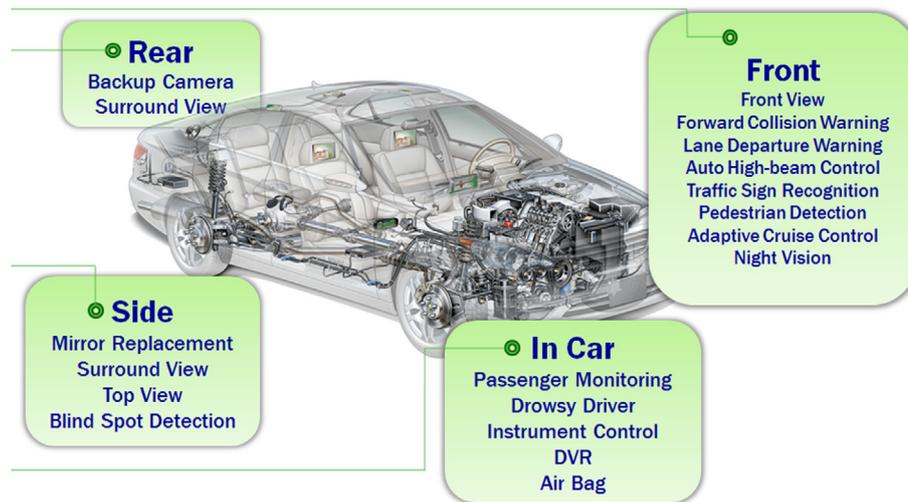


Fig. 12. Automobile imaging sensor opportunities.

uses the latest 3.0 μm Back Side Illuminated (BSI) pixel with ON Semiconductor’s DR-Pix™ technology.

BSI technology also offers higher quantum efficiency (QE) and reduced cross-talk.

Stacked wafer technology is another image sensor technology opportunity. In a stacked wafer process, one wafer contains the pixels while the other contains the mixed signal devices. The “sensor” wafer is then bonded to the “ASIC” wafer as shown in Fig. 13. This approach allows for individual optimization of each wafer for both cost and performance. Fig. 14 shows the resulting cross-section of Pixel and ASIC wafers connected using through-silicon vias (TSV).

6. Conclusions

The market for power management ICs and Discretes continues to grow at a steady pace thanks to new automotive technologies, the proliferation of ever-more sensor applications, and the “Internet of Things.” SOI technologies offer numerous cost, performance, and integration improvement opportunities. Using SOI integration, chip manufacturers are able to shrink parts, offer both low and high voltage devices in a single chip, improve IGBT performance without suffering the trade-offs seen in bulk technologies, and create novel sensor technologies. SOI technologies also offer the

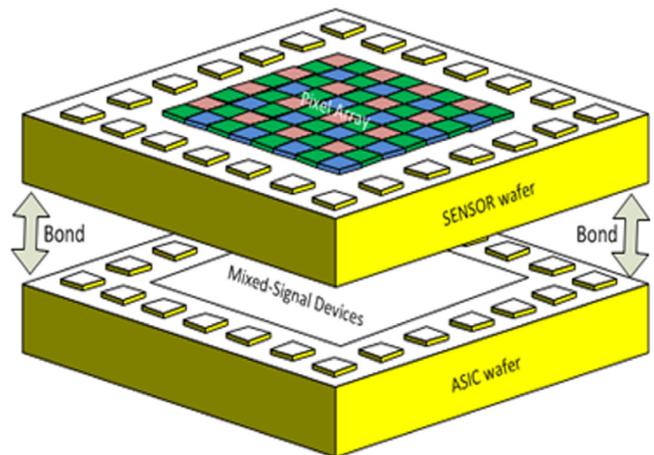


Fig. 13. Stacked wafer configuration.

chance to reduce parasitics, leading to improved performance and more accurate models, which in turn, lead to reduced development costs and faster times to market.

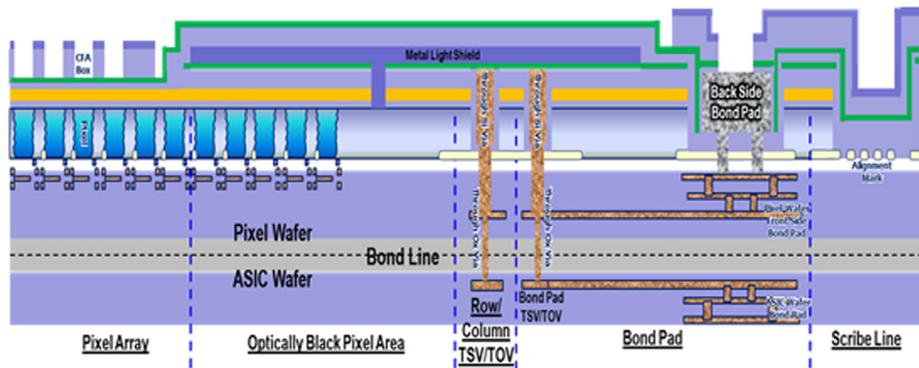


Fig. 14. Stacked wafer cross-section.

Acknowledgements

The authors extend their thanks to Technische Universitat Wien (TUWien), the hosting institute for EUROS0I-ULIS 2016, as well as the Conference's Chairmen, Dr. Viktor Sverdlov and Dr. Siegfried Selberherr.

The authors also thank our many colleagues at ON Semiconductor whose projects and products are represented in the information presented here.

References

- [1] Olsson J. High voltage SOI devices for automotive applications. *Science and Technology of Semiconductor-on-Insulator Structures and Devices Operating in a Harsh Environment*. Netherlands: Springer; 2006. p. 155–66.
- [2] Heinle U, Pinardi K, Olsson J. Vertical high voltage devices on thick SOI with back-end trench formation. In: *Solid State Device Research Conf.*. p. 295–8.
- [3] S. Bernstein, *Auto Semiconductors Report*, September 2012.
- [4] Pop E et al. Energy dissipation and transport in nanoscale devices. *Nano Res.* 2010;3(3):147–69.
- [5] Arnold E, Pein H, Herko SP. Comparison of self-heating effects in bulk-silicon and SOI high-voltage devices. In: *Int. Electron Devices Meeting*, San Francisco, CA. p. 813–6.
- [6] G.P. Hosey, Modeling heat dissipation in an SOI device – finite element analysis at semiconductor die level, in: *Joint Int. EUROS0I Workshop and Int. Conf. on Ultimate Integration on Silicon*, Vienna, 2016, pp. 206–209.
- [7] H. Hauenstein, High-Voltage IGBTs Move into Automotive Apps, *Electronic Design*, October 2010.



Reliable gate stack and substrate parameter extraction based on C-V measurements for 14 nm node FDSOI technology



B. Mohamad^{a,b,*}, C. Leroux^a, D. Rideau^c, M. Haond^c, G. Reibold^a, G. Ghibaudo^b

^aCEA-Leti, MINATEC Campus, 38054 Grenoble Cedex 9, France

^bUniv. Grenoble Alpes, IMEP-LAHC, MINATEC/INPG, CS 50257, 38016 Grenoble, France

^cSTMicroelectronics, BP 16, 38920, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Effective work function
Equivalent oxide thickness
Channel thickness
Buried oxide thickness
Fully depleted silicon on insulator

ABSTRACT

Effective work function and equivalent oxide thickness are fundamental parameters for technology optimization. In this work, a comprehensive study is done on a large set of FDSOI devices. The extraction of the gate stack parameters is carried out by fitting experimental CV characteristics to quantum simulation, based on self-consistent solution of one dimensional Poisson and Schrodinger equations. A reliable methodology for gate stack parameters is proposed and validated. This study identifies the process modules that impact directly the effective work function from those that only affect the device threshold voltage, due to the device architecture. Moreover, the relative impacts of various process modules on channel thickness and gate oxide thickness are evidenced.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The effective Work function (WF_{eff}) and the Equivalent Oxide thickness (EOT) are key parameters for FDSOI device characterization. WF_{eff} and EOT are the values of Work Function and Thickness that an ideal Metal/SiO₂ gate stack should have to account for the measurements. They directly influence Capacitance Equivalent thickness (CET), Threshold voltage (V_T) and therefore on-current. Both are strongly related to the process: dielectric thickness, metal gate work function and dielectric interface dipole [1]. Former methodologies [2,3] for automatic and statistical extraction of WF_{eff} and EOT were fitted for bulk but are no longer sufficient for FDSOI technology. Capacitance–Voltage extraction of the flat band voltage (V_{FB}) in the accumulation region has traditionally been the method for the WF_{eff} extraction for the bulk technology. The absence of CV signal in the accumulation and flat band regions compared to standard bulk capacitance pushes us to look for alternative solution for WF_{eff} and EOT extraction. Indeed the inversion region (V_T) still available in FDSOI can be exploited to extract the WF_{eff} and EOT. The extraction difficulties are due to the complexity of the FDSOI structure (Fig. 1) and the strong influence on V_T of several technological parameters such as: channel thickness t_{si} , buried

oxide (BOX) thickness t_{box} and well doping level of substrate N_{Well} at BOX backside. We will present a comprehensive study on numerous devices combining the process modules on metal gate, dielectric, channel material (Si & SiGe) and well implantation type (P & N doped). Reliable parameter extraction is presented by comparison between quantum simulations and experimental CV characteristics allowing the identification of the process modules that really influence EOT and WF_{eff} .

2. Process technology and experimental results

Fig. 1 shows two standard FDSOI MOS devices (P (a) & N (b) MOS) that feature a gate stack characterized by a metal bilayer (Poly and TiN) on the top of oxide bilayer (High-k dielectric HfON & SiON). To get such devices, the process starts from an ultra-thin silicon body over a buried oxide. The strained Si_{0.75}Ge_{0.25} channel (cSiGe) is selectively processed in PMOS areas by epitaxy growth process followed by a Ge condensation, before shallow trench isolation (STI) regions are patterned and the back side wells are implanted. The channel thickness is around 6–8 nm of Si or Si_{0.75}–Ge_{0.25} over a 20 nm BOX with a P or N-Well with a 10¹⁸ cm⁻³ doping level at its backside. Two different interlayer (SiON) dielectrics, can be deposited corresponding to two different final oxide thicknesses: EOT = 1 nm for GO1 and EOT = 3 nm for GO2. The bilayer high-k HfON/SiON is deposited by Atomic Layer Deposition, Metal Organic Vapor deposition and decoupled plasma nitridation. To

* Corresponding author at: CEA-Leti, MINATEC Campus, 38054 Grenoble Cedex 9, France.

E-mail address: blend.mohamad@cea.fr (B. Mohamad).

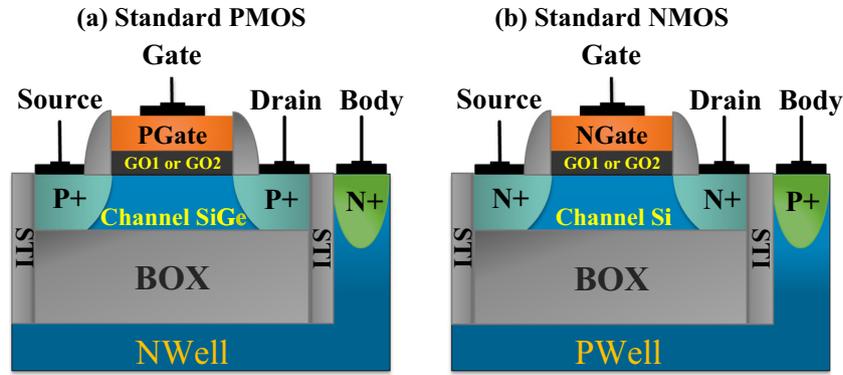


Fig. 1. Standard (a) PMOS and (b) NMOS FDSOI devices: both feature a gate stack TiN + Poly on a bilayer oxide HfON + SiON but (a) with SiGe channel and N-Well substrate and (b) with Si channel and P-Well substrate. The interlayer (SiON) dielectrics, is deposited in order to have two different final oxide thicknesses: EOT = 1 nm for GO1 and EOT = 3 nm for GO2.

Table 1

Set of different FDSOI devices featured by combining different process modules on metal gate, dielectric, channel (Si & SiGe) and well (P & N doped). The 4th and 5th columns report on the ΔV_T shift evidenced in Figs. 2 and 3.

Device	Device definition (GO1 & GO2)			GO1	GO2
	Channel	Metal	Well	ΔV_T [mV]	ΔV_T [mV]
NMOS	Si	N	P	Ref.	Ref.
	Si	N	N	45	115
	Si	P	P	90	45
PMOS	SiGe	P	N	Ref.	Ref.
	SiGe	N	P	60	90
	SiGe	N	N	100	45
	Si	P	N	320	250
	Si	N	N	×	295

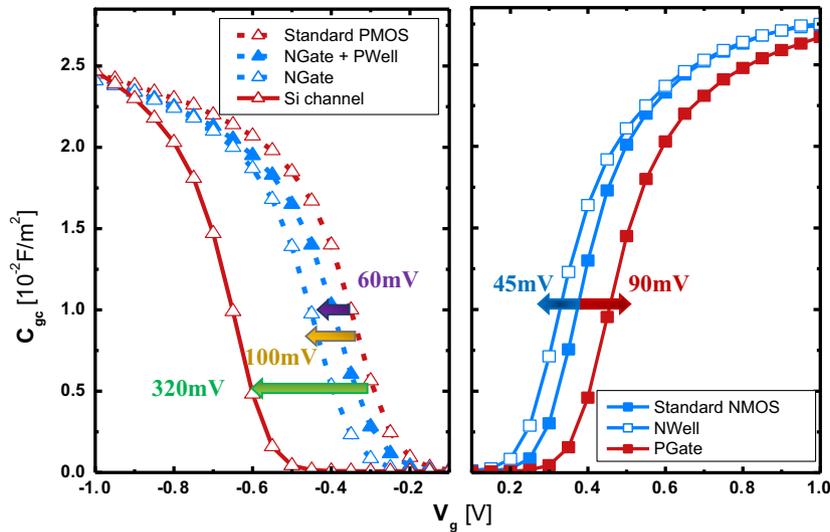


Fig. 2. Gate-to-channel capacitance at back bias $V_B = 0$ V for PMOS and NMOS FDSOI GO1 with V_T shift to reference architecture.

obtain N and P gate types, before the final metal gate, a sacrificial gate is deposited, specifically to each N or P gate type. Sacrificial gate includes specific additives which are diffused by thermal annealing in order to adjust dipole at oxide bilayer inner interface [1,4]. The final metal gate bilayer (TiN + Poly) is the same for all the devices (N and P-MOS) and it is deposited right after the sacrificial gate etching. After final gate deposition, the Source-Drain implantations (specific to each FDSOI MOS type) are carried out, followed

by final annealing. From these standard FDSOI MOS, a wide set of devices has been obtained by combining the process modules on metal gate, dielectric, channel and well type (Table 1).

In order to carry out the parameter extraction of W_{eff} and EOT, two types of measurements have been performed on this set of devices: the gate-to-channel capacitance (C_{gc}) and the back-gate-to-channel capacitance (C_{bc}) [5]. For C_{gc} measurements, the small AC signal is applied on the Gate and the induced capacitive current

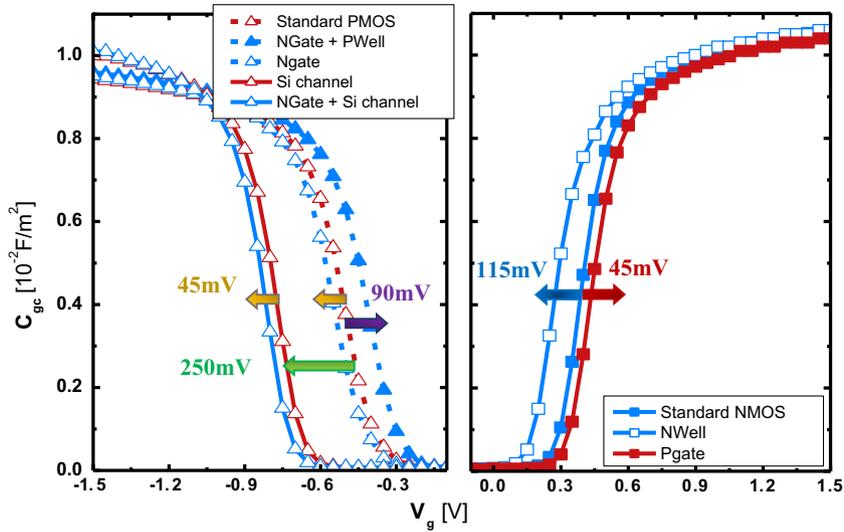


Fig. 3. Gate-to-channel capacitance at back bias $V_B = 0$ V for PMOS and NMOS FDSOI GO2 with V_T shift to the reference architecture.

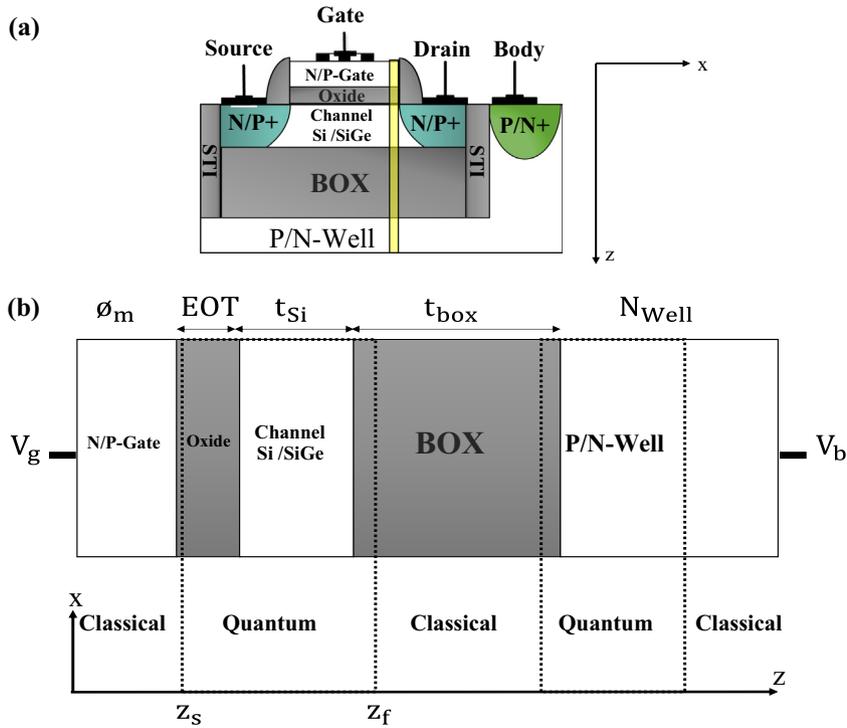


Fig. 4. 1-D (a) Cross section of typical FDSOI transistor designed and measured in this study. (b) 1-D structure simulated to account of the transistor electrostatics (yellow line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is measured on Source and Drain. For C_{bc} measurements, the small AC signal is applied on the back side contact and the induced capacitive current is still measured on Source and Drain. In both the experimental configurations Source and Drain are grounded. These measurements have been performed for all the devices and their relative gate-to-channel capacitances, corrected by subtracting their parasitic capacitances, are reported in Fig. 2 for GO1 devices and Fig. 3 for GO2 devices. As summarized in Table 1, we have the opportunity to test various devices, N-MOS and P-MOS, on which we can get various channel materials, Gate type and Well type for ground plane. A reference can be identified for each MOS structure. N-MOS has a N-Gate and P ground plane with a Si channel whereas P-MOS has P-Gate and N ground plane with a SiGe

channel. Arrows in Figs. 2 and 3 identifies the V_T shift from reference when we modify some process module of the gate stack. V_T shifts are reported in Table 1 and we will try in following part to account for such ΔV_T variations.

3. Simulation and extraction

The extraction of gate stack parameters (EOT and WF_{eff}) for these devices (Table 1) is obtained by fitting Quantum Simulation to experimental capacitance-voltage characteristics (CV). Indeed, for a specific gate stack (WF_{eff} , EOT, t_{Si} , t_{box} , N_{well}), we can simulate the channel charge of the MOS device and compare it to experi-

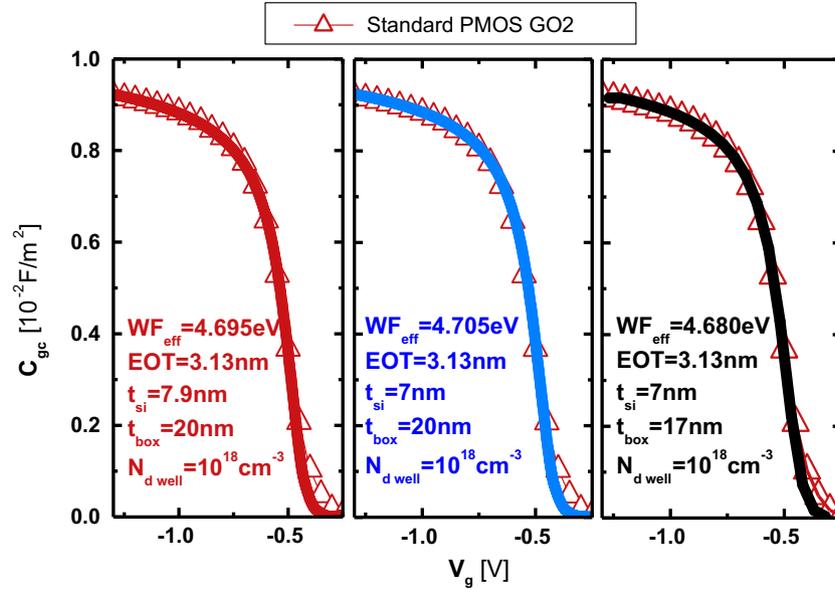


Fig. 5. Standard PMOS GO2 gate-to-channel capacitance compared with simulation (UTOXPP) with different channel and BOX thicknesses (each fit leads to a certain WF_{eff} and EOT extraction).

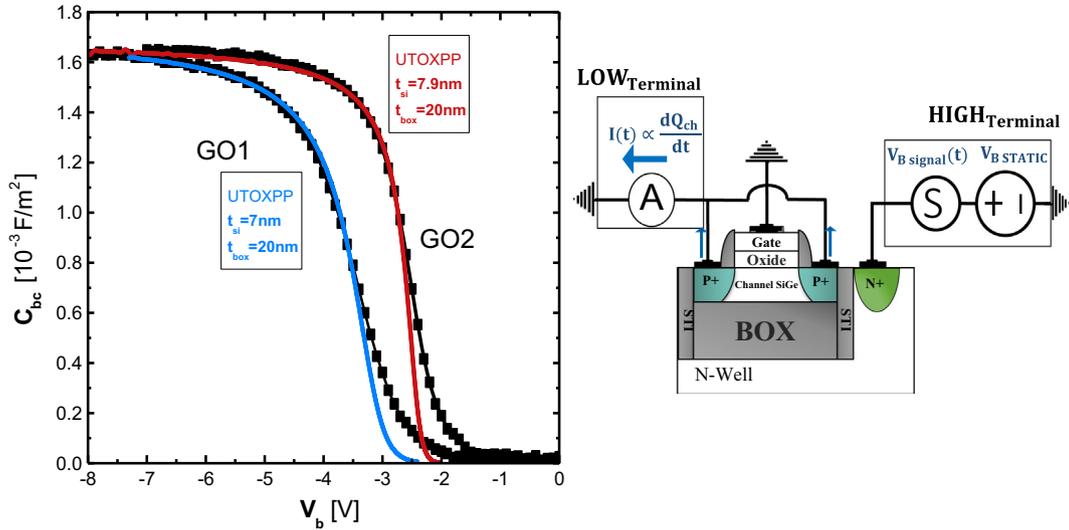


Fig. 6. Experimental standard PMOS GO1 & GO2 back-gate-to-channel capacitance compared to simulation (UTOXPP) leading to t_{box} extraction.

mental measurements. We can consider that WF_{eff} and EOT are correctly identified if we account for the various dependences of the experimental capacitances with Gate and Body biases. These simulations are based on the self-consistent solutions of the Poisson and Schrodinger equations (PS), carried out by NUMERICAD Simulator (UTOXPP [6]).

In order to accurately calculate the charge distribution in the channel and substrate of the FDSOI stack (Fig. 4), parabolic effective-mass approximation (EMA) of the valence and conduction band are taken into account. As shown in Fig. 4, the PS simulations are carried out in z -direction, orthogonal to the channel-oxide interface (Fig. 4a yellow line). With the purpose of including electron and hole wave function penetration, the PS formalism is not just solved in the channel and substrate (Well) but also in complementary regions as the front oxide (EOT) and back oxide (BOX) (Fig. 4b). To calibrate the EMA simulation parameters, a more complex and accurate 6-K·P band self-consistent simulation

has been performed, leading to corrected in-plane effective masses: i.e. for the heavy hole valence band ($m_{DOS-HH_{xy}}$) of $2.5m_0$ for Si and $1.5m_0$ for SiGe channel [3], with m_0 as the electron mass.

The calculated density of charge $\rho(z, V_g, V_b)$ from the PS equation can be used to compute the channel charge as

$$Q_{ch}(V_g, V_b) = \int_{z_s}^{z_f} \rho(z, V_g, V_b) dz \quad (1)$$

From such channel charge, we can calculate the gate-to-channel capacitance, defined by

$$C_{gc} = \frac{dQ_{ch}(V_g, V_b)}{dV_g} \quad (2)$$

and the back-gate-to-channel capacitance by

$$C_{bc} = \frac{dQ_{ch}(V_g, V_b)}{dV_b} \quad (3)$$

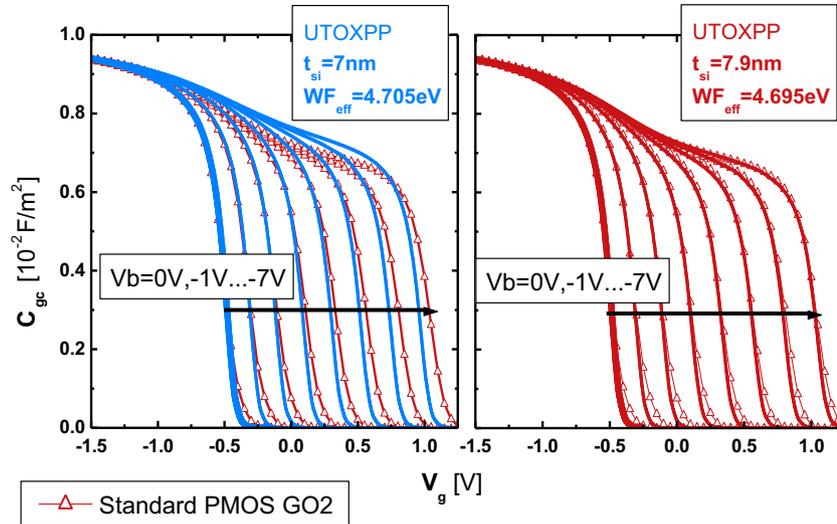


Fig. 7. Standard PMOS GO2 gate-to-channel capacitance vs simulation (UTOXPP) with two different channel thicknesses (7.9 nm t_{si} is mandatory for fitting all CVs).

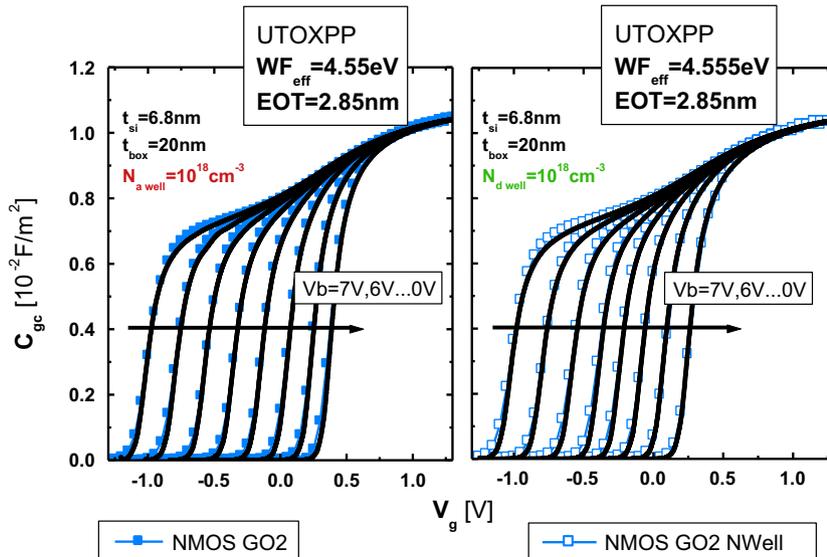


Fig. 8. Fitting between simulations and experiments for two MOS devices having two different well types at back side of buried oxide.

These simulated capacitances will be compared with the experimental CV curves in order to extract the FDSOI stack parameters. To evaluate the quality of parameter extraction, various hypotheses on physical parameters (t_{box} , t_{si}) have been compared for a same experimental device (Standard PMOS GO2). A good fit of $C_{gc}(V_g)$ can be obtained for the three different set of parameters (Fig. 5) leading to three different values for EOT and WF_{eff} . These values are given with a precision of 0.01 nm and 3–4 meV. Whereas, EOT is found to be independent of t_{box} and t_{si} values (with precision below 0.01 nm), a 25 meV WF_{eff} uncertainty is found. Such a result confirms the importance of a previous extraction on t_{box} and t_{si} in order to accurately extract WF_{eff} . To extract t_{box} , we have recently proposed a new CV measurement, the C_{bc} earlier mentioned, with back side wafer contact as Gate [5]. Fig. 6 shows back-gate-to-channel capacitance experimental configuration and measurements results.

The corresponding CV curves for GO1 and GO2 (Fig. 6) shows characteristics similar to the standard gate-to-channel capacitance but at a significantly lower level. Adjustment with quantum simu-

lation makes possible a reliable extraction of the box thickness, independently of t_{si} and EOT [5]. A 20 nm box thickness is necessary in order to fit C_{bc} capacitances for both the standard PMOS (GO1 and GO2). Concerning t_{si} , it can be obtained by comparison between simulations and experiments on a large set of CV characteristics for which back biases vary from 0 V to large values leading to back interface inversion (Fig. 7). We can notice that there is only one channel thickness that takes into account the real capacitance deformation for strong back interface inversion. t_{box} and t_{si} are expected to shift V_T , WF_{eff} could also vary with channel material (Si versus SiGe) but it is obviously independent of the Source and Drain or Well type. Indeed, during the device fabrication, the implantation of the Source-Drain as well as the back substrate well are carried out without any influence on the MOS stack. A reliable parameter extraction must verify the independence of WF_{eff} and EOT with variations on types of Source and Drain or Well implantation. This feature has been confirmed on three different MOS devices, corresponding to two different well types (Fig. 8) and two different MOS types (Fig. 9). Moreover Fig. 9 validates the good

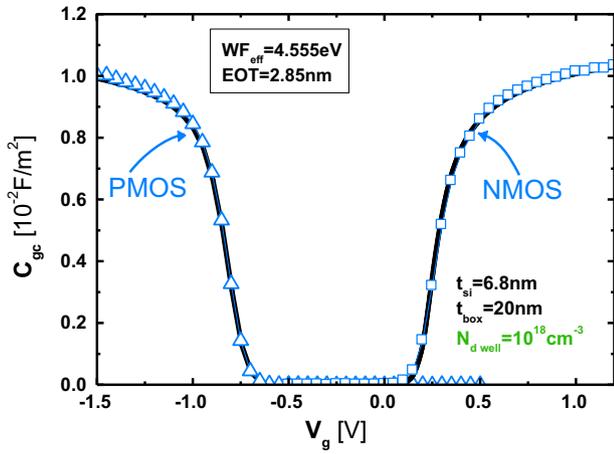


Fig. 9. Fitting between simulation and experimental data for two different MOS devices having a same gate stack and two different source and drain implantations (GO2 NMOS with N-Well & GO2 PMOS with N-Well and N-Gate metal and Si channel).

choice of m_{DOS-HH} of $2.5m_0$ for Si channel [3]. Indeed, only a good value of mass, related to electrons (conduction band) and holes (valence band), can reproduce both capacitances with a same WF_{eff} and EOT.

4. Results

In Fig. 10, are shown the WF_{eff} and EOT extracted for the fifteen different tested MOS devices (Table 1). The agreement reported above on WF_{eff} & EOT for three different MOS structures (Figs. 8 and 9) is identified in Fig. 10 with a blue arrow. Among the fifteen different devices, other configurations correspond to same gate stack and channel material but different Source and Drain or back side well types. They all lead to a good agreement which are identified with green arrows in Fig. 10.

We report in Fig. 11 the dependence of t_{si} with EOT. It appears to depend only on channel and oxide type. Indeed, it is equal to 6.1 nm for Si and 7 nm for SiGe for all GO1 devices and 6.8 nm for Si and 7.9 nm for SiGe for all GO2 devices. SiGe channel is found to be 1 nm thicker than Si channel and, in both cases, GO2 process leads to less channel consumption.

t_{box} is reported in Fig. 12 and it is found at the same 20 nm value for all the different MOS devices, a value corresponding to the target of the SOI substrate. EOT is reported in Figs. 10–12. It depends primarily on oxide type (GO1 and GO2) but also on channel material. As already evidenced in previous studies [7], the gate oxide formation leads to a slight interlayer regrowth in case of SiGe substrate. We notice here an interlayer regrowth around 0.12 nm for GO1 and 0.3 nm for GO2.

WF_{eff} appears to depend on gate type, channel material and oxide (GO1 and GO2). Considering first GO1, we notice an average shift of 100 mV from N to P metal gate. It is the expected effect of Al dipole created at HfON/SiON interface by the sacrificial gate process [1]. With SiGe channel, WF_{eff} report an additional shift of 145 mV. In Fig. 2 we can notice that V_T shift with SiGe channel reaches 320 mV. It is partially explained by the shift of valence band for SiGe semiconductor, but an additional WF_{eff} shift of 145 mV is required to account for the whole V_T shift. It has been attributed to an interface dipole at SiGe/SiON interface [7]. For GO2, both dipoles at HfON/SiON and SiGe/SiON device interfaces decreases of 45 and 95 mV, respectively.

5. Conclusion

Robust EOT and WF_{eff} extractions on FDSOI devices have been proposed through a methodology based on comparison between experimental CV characteristics and Poisson-Schrodinger simulations. To be reliable, it requires a first identification of buried oxide thickness (t_{box}) and channel thickness (t_{si}). It has been validated on a large set of MOS devices. Such analysis evidences the relative impact of process modules on t_{si} , EOT and WF_{eff} . Channel thickness and equivalent oxide thickness appear to depend on the channel material and the gate oxide interlayer process. Whereas effective

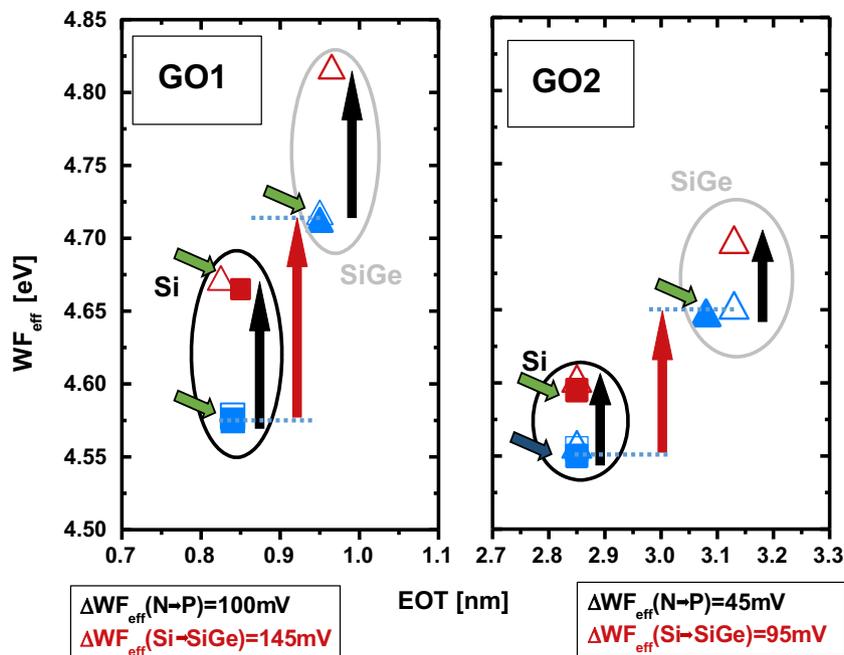


Fig. 10. WF_{eff} versus EOT for different PMOS and NMOS FDSOI structures. Impact of Ge and sacrificial metal gate and coherence of WF_{eff} and EOT extractions are shown.

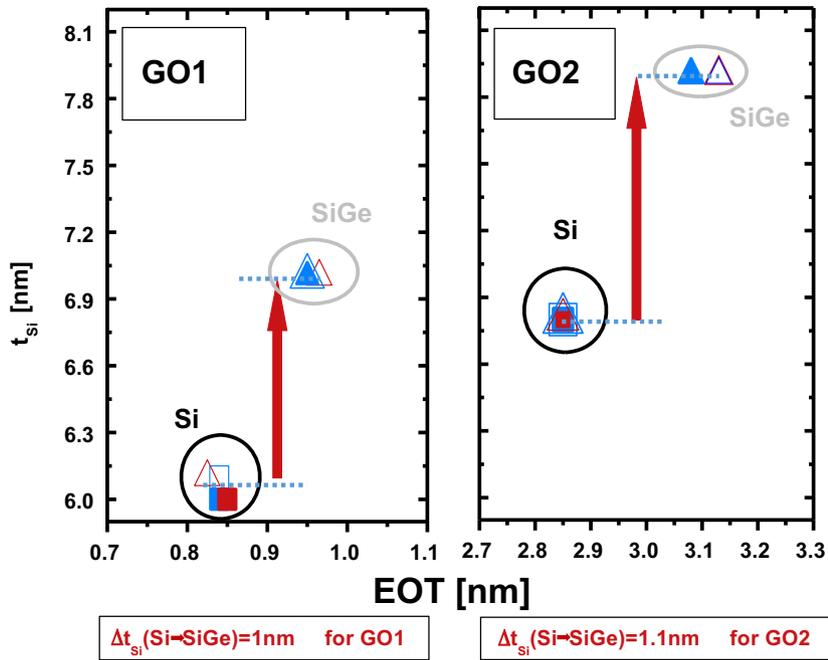


Fig. 11. t_{si} versus EOT for different PMOS and NMOS FDSOI structures. Impact of Ge and coherence of t_{si} and EOT extractions are shown.

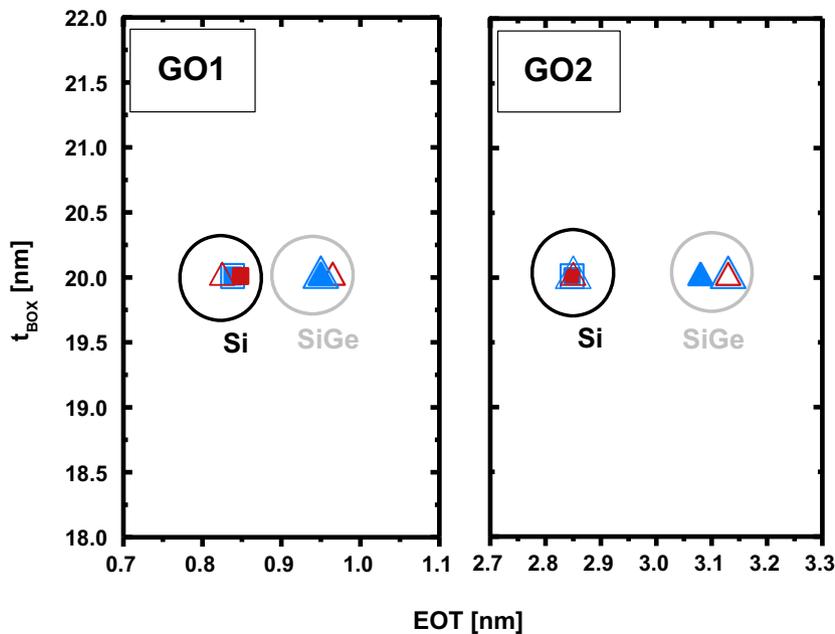


Fig. 12. t_{box} versus EOT for different PMOS and NMOS FDSOI structures. Impact of Ge and coherence of t_{box} and EOT extractions are shown.

work function is a combined effect of interlayer, channel interlayer and sacrificial gate type.

Acknowledgements

The authors would like to thank MINOS Laboratory of French ANR and WAYTOGO FAST ECSEL Project for their support for this work.

References

- [1] Suarez-Segovia C, Leroux C, Caubet P, Domengie F, Reimbold G, Romano G, et al. Effective work function modulation by sacrificial gate aluminum diffusion on HfON-based 14 nm NMOS devices. *Microelectr Eng* 2015;147:113–6.
- [2] Leroux C, Baudot S, Charbonnier M, Van Der Geest A, Caubet P, Toffoli A, et al. Investigating doping effects on high-k metal gate stack for effective work function engineering. *Microelectr Eng* 2007;84:2408–11.
- [3] Soussou A, Leroux C, Rideau D, Toffoli A, Romano G, Tavernier C, et al. Parameters extraction in SiGe/Si pMOSFETs using split CV technique. In: *Proc of ULIS conf, Cork (Ireland)*. p. 41–4.
- [4] Suarez-Segovia Carlos, Leroux C, Domengie F, Dabertrand K, Joseph V, Romano G, et al. Effective work function engineering by sacrificial lanthanum diffusion on HfON-based 14 nm NFET devices. In: *45th European Solid State Device Research Conference (ESSDERC)*, 14–18 September 2015. p. 246–9.
- [5] Mohamad B, Ghibaudo G, Leroux C, Josse E, Reimbold G. Full front and back split C-V characterization of CMOS devices from 14 nm node FDSOI technology. In: *Proc of S3S conf, Rohnert Park (USA)*. Section 9a.4.
- [6] Garetto D, Rideau D, Tavernier C, Leblebiciand Y, Schmid A, Jaouen H. Advanced physics for simulation of ultrascaled devices with UTOXPP Solver. In: *Proc of NTSI-nanotech 2011, Boston (USA)*. p. 607–10.
- [7] Soussou A, Leroux C, Rideau D, Toffoli A, Romano G, Saxod O, et al. Understanding Ge impact on VT and VFB in Si1–xGex/Si pMOSFETs. *Microelectr Eng* 2013;109:282–4.



Scaling/LER study of Si GAA nanowire FET using 3D finite element Monte Carlo simulations



Muhammad A. Elmessary^{a,c,*}, Daniel Nagy^a, Manuel Aldegunde^d, Natalia Seoane^e, Guillermo Indalecio^e, Jari Lindberg^f, Wulf Dettmer^b, Djordje Perić^b, Antonio J. García-Loureiro^e, Karol Kalna^a

^a ESDC, College of Engineering, Swansea University, Swansea SA1 8EN, Wales, United Kingdom

^b ZCCE, College of Engineering, Swansea University, Swansea SA1 8EN, Wales, United Kingdom

^c Dept of Mathematics & Engineering Physics, Faculty of Engineering, Mansoura University, Mansoura 35516, Egypt

^d WCPM, School of Engineering, University of Warwick, Coventry CV4 7AL, England, United Kingdom

^e CITIUS, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Galicia, Spain

^f Varian Medical Systems Finland, Helsinki, Finland

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Schrödinger quantum corrections
Monte Carlo simulations
GAA nanowire FET
LER
Variability

ABSTRACT

3D Finite Element (FE) Monte Carlo (MC) simulation toolbox incorporating 2D Schrödinger equation quantum corrections is employed to simulate I_D - V_G characteristics of a 22 nm gate length gate-all-around (GAA) Si nanowire (NW) FET demonstrating an excellent agreement against experimental data at both low and high drain biases. We then scale the Si GAA NW according to the ITRS specifications to a gate length of 10 nm predicting that the NW FET will deliver the required on-current of above 1 mA/ μ m and a superior electrostatic integrity with a nearly ideal sub-threshold slope of 68 mV/dec and a DIBL of 39 mV/V. In addition, we use a calibrated 3D FE quantum corrected drift-diffusion (DD) toolbox to investigate the effects of NW line-edge roughness (LER) induced variability on the sub-threshold characteristics (threshold voltage (V_T), OFF-current (I_{OFF}), sub-threshold slope (SS) and drain-induced-barrier-lowering (DIBL)) for the 22 nm and 10 nm gate length GAA NW FETs at low and high drain biases. We simulate variability with two LER correlation lengths (CL = 20 nm and 10 nm) and three root mean square values (RMS = 0.6, 0.7 and 0.85 nm).

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Gate-All-Around (GAA) nanowire (NW) FETs are considered to be excellent candidates for future CMOS integration for sub-10 nm digital technology to continue transistor downscaling [1]. The GAA NW FETs have superior electrostatics and immunity to short channel effects while still delivering a large on-current [2–4]. However, to have a full realistic assessment of these potential candidates, we must take into account the exact device geometry and also determine how different sources of device variability can affect device characteristics and reliability. Variability of transistor characteristics is induced by material properties and by fabrication processes and can affect their performance in circuits. One of such sources is line-edge roughness (LER) which has a major impact on variability in NW/FinFETs [3,5–7].

In this work, we report on performance, scaling and variability of nanoscale GAA Si NW FETs. We use an in-house 3D Finite Element (FE) Monte Carlo (MC) simulation toolbox which includes newly integrated calibration-free 2D FE anisotropic Schrödinger equation based quantum corrections (SEQC) [8] along the device channel. More details on the 3D FE MC toolbox are in Refs. [8–11]. Here, we start by comparing results from our 3D FE SEQC MC toolbox against experimental data of a 22 nm gate length GAA Si NW FET [2] with a $\langle 110 \rangle$ channel orientation. We then simulate the $\langle 100 \rangle$ channel orientation for the same NW for comparison. Next, we scale the NW to a gate length of 10 nm according to the International Technology Roadmap for Semiconductors (ITRS) specifications [12] and simulate the $\langle 100 \rangle$ and $\langle 110 \rangle$ channel orientations with the 3D FE SEQC MC. Finally, we use our 3D quantum corrected FE drift-diffusion (DD) simulation toolbox to study the LER-induced variability on the sub-threshold characteristics (threshold voltage (V_T), OFF-current (I_{OFF}), sub-threshold slope (SS) and drain-induced-barrier-lowering (DIBL)) at both low and high drain biases for the 22 nm and the 10 nm gate length GAA NW FETs. We simulate the variability with LER correlation lengths

* Corresponding author at: ESDC, College of Engineering, Swansea University, Swansea SA1 8EN, Wales, United Kingdom.

E-mail address: M.A.A.Elmessary.716902@swansea.ac.uk (M.A. Elmessary).

(CL) of 20 nm and 10 nm and three root mean square values (RMS) of 0.6, 0.7 and 0.85 nm chosen to represent the RMS observed in experiments [2,3].

2. 3D Monte Carlo simulations

The 3D FE method incorporated into 3D SEQC MC toolbox is capable of accurately describing the complex 3D geometry of the nanoscale devices. The accurate description of the simulation domain in nanoscale semiconductor devices is essential in determining quantum transport at highly non-equilibrium conditions. In our case, we chose a semi-classical transport technique, a 3D ensemble MC [8–11], with calibration-free quantum confinement corrections, the SEQC. Our in-house 3D FE MC simulation toolbox employs fully anisotropic 2D FE Schrödinger equation based quantum corrections (QC) [8] which depends on valley orientation and considers longitudinal and transverse electron effective masses. The MC transport engine considers an analytical anisotropic non-parabolic bandstructure model with the same longitudinal and transverse masses and the following scattering processes: the acoustic phonon scattering, non-polar optical phonon scattering (*g,f*-processes) [13], ionised impurity scattering using the third-body exclusion model by Ridley [14] with a static screening model self-consistently calculating Fermi energy and electron temperature [15], and the interface roughness scattering using Ando's model [16]. This combination has been shown to be a very good compromise between the precision and the speed for accurate physical simulations of carrier transport in nanodevices which are strongly quantum confined systems at highly non-equilibrium transport conditions [9–11].

We start by comparing results from our 3D SEQC MC toolbox against experimental data of a 22 nm gate length GAA Si NW [2] with a $\langle 110 \rangle$ channel orientation. The NW has elliptical cross-section (Fig. 1) with a shorter diameter of 11.3 nm and a longer diameter of 14.22 nm; with an effective diameter (elliptical circumference/ π) $D_{NW} = 12.8$ nm and EOT = 1.5 nm which can be accurately described by the FE method.

The 3D FE quantum corrected (QC) drift-diffusion (DD) simulations using density gradient (DG) [17] were used to reverse engineer a doping profile in the sub-threshold region at $V_D = 0.05$ V

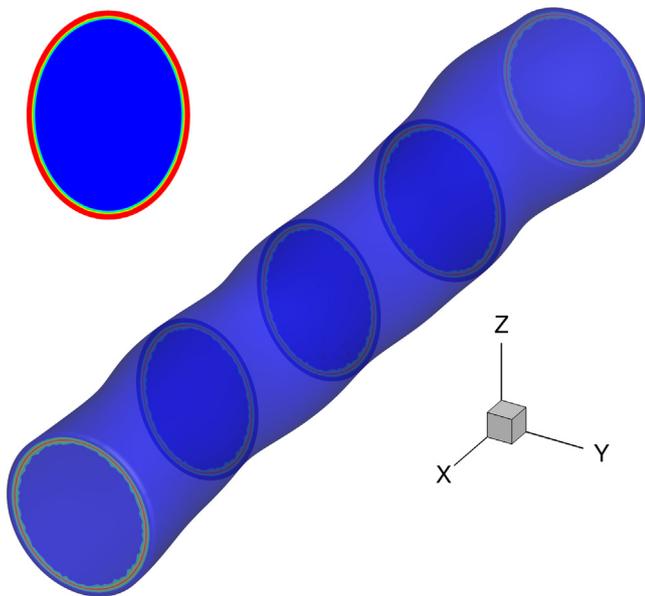


Fig. 1. Schematic of the 22 nm gate length n-channel Si GAA nanowire, showing LER and examples of 2D slices used for Schrödinger solver.

and $V_D = 1.0$ V by changing the Gaussian-like doping profile (a doping maximum and a spread X which determine the abruptness of the doping profile) as shown in Fig. 2. The effective masses in the DG approach were used as calibration parameters. Fig. 3 shows examples of this reverse engineering process for the sub-threshold region which achieved excellent agreement with a maximum doping of $5 \times 10^{19} \text{ cm}^{-3}$, a work function of 4.512 eV, and a S/D size of 30.8 nm. We then use the 3D FE SEQC MC toolbox to simulate the I_D - V_G characteristics of the 22 nm gate GAA Si NW at low and high drain biases achieving an excellent agreement with the experimental data [2] as can be seen in Fig. 4. The current is normalised by nanowire perimeter (elliptical circumference = 40.21 nm). Note here that the resulted drain current from 3D SEQC MC simulations gives the agreement with experiment without any need for additional lumping of external resistance from the experiment. This is because the S/D resistance is accurately reproduced in physically based 3D ensemble MC technique thanks to the size of the S/D access regions included into simulation domain.

Fig. 5 compares the average electron velocity at $V_G = 0.8$ V and $V_D = 1.0$ V for the 22 nm gate length GAA NW along the $\langle 100 \rangle$ and the $\langle 110 \rangle$ channel orientations, along with the average veloc-

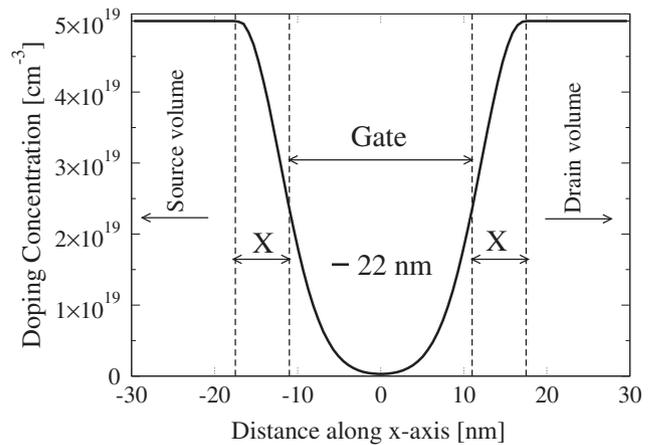


Fig. 2. Cross-section of Gaussian-like doping profile along the transport x -direction in the 22 nm gate length GAA NW FET.

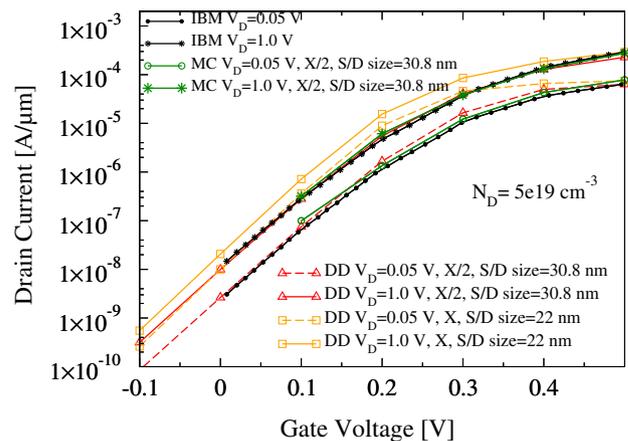


Fig. 3. Devising doping profile for the 22 nm GAA NW FET at $V_D = 0.05$ V and $V_D = 1.0$ V via DD simulations by changing the size of the S/D region and the doping spread X (open red triangles and orange squares). Final MC simulations (green open circles and stars) are compared to experimental data (black full circles and stars). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

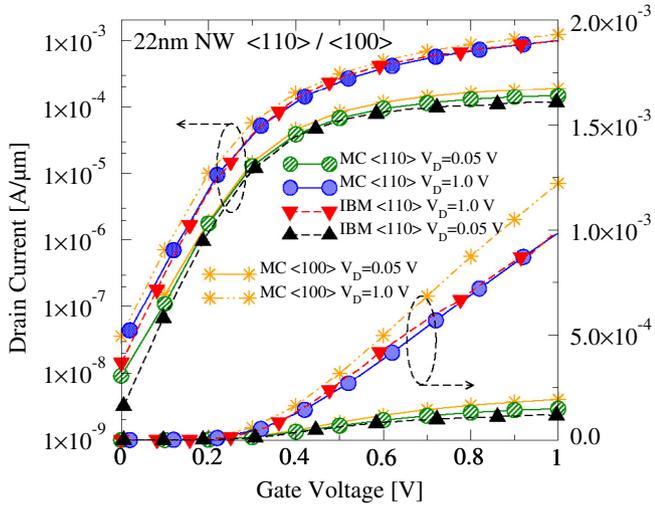


Fig. 4. I_D - V_G characteristics for the 22 nm GAA nanowire from the 3D FE MC with anisotropic Schrödinger quantum corrections, in the $\langle 110 \rangle$ (circles) and $\langle 100 \rangle$ (stars) channel orientations, compared against experimental data in the $\langle 110 \rangle$ (triangles) orientation [2].

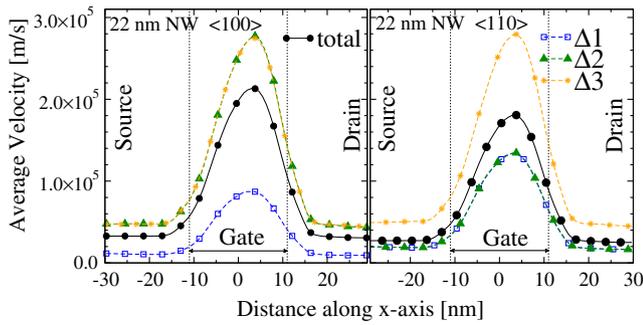


Fig. 5. Average electron velocity at $V_D = 1.0$ V and $V_G = 0.8$ V along the 22 nm gate length GAA NW (3D MC) along the $\langle 100 \rangle$ and $\langle 110 \rangle$ channel orientations. The zero is set in the middle of the channel.

ity in the three silicon valleys $\Delta 1$, $\Delta 2$ and $\Delta 3$. We see that the electron velocity exhibits a typical behaviour along the channel. The total average velocity in the $\langle 100 \rangle$ orientation is higher than the one in the $\langle 110 \rangle$ orientation due to a higher electron mobility in the $\langle 100 \rangle$ crystallographic orientation in Si. The source injects electrons at relatively large injection velocity of about 3.0×10^4 m/s where they are quickly accelerated along the gate reaching their maximum velocity of $2.0/1.75 \times 10^5$ m/s, respectively, on the drain side of the gate. Then electrons decelerate into a heavily doped drain due to enhanced optical phonon emission assisted by ionised impurity scattering [18].

In the $\langle 100 \rangle$ orientation channel, the $\Delta 1$ velocity is the smallest because it has the heaviest mass in the transport direction. The $\Delta 2$ and $\Delta 3$ velocities are equal in the $\langle 100 \rangle$ channel device because they have the same effective transport masses. On the other hand, in the $\langle 110 \rangle$ channel, the $\Delta 3$ velocity is the largest because it has the lightest effective transport mass. The $\Delta 1$ and $\Delta 2$ velocities are equal because they have equal effective transport masses.

We then scale the Si GAA NW according to the ITRS specification [12] to a gate length of 10 nm and an EOT of 0.8 nm. Fig. 6 compares the eigenmodes ($|\psi(y, z)|^2$) corresponding to the lowest energy eigenvalue, in the $\langle 110 \rangle$ channel orientation, of the three Δ valleys in the middle of the channel for the 22/10 nm gate length GAA NW, respectively, at $V_D = 1.0/0.7$ V and $V_G = 0.8$ V (note that $\Delta 1$ and $\Delta 2$ have the same effective mass tensor in the $\langle 110 \rangle$ channel

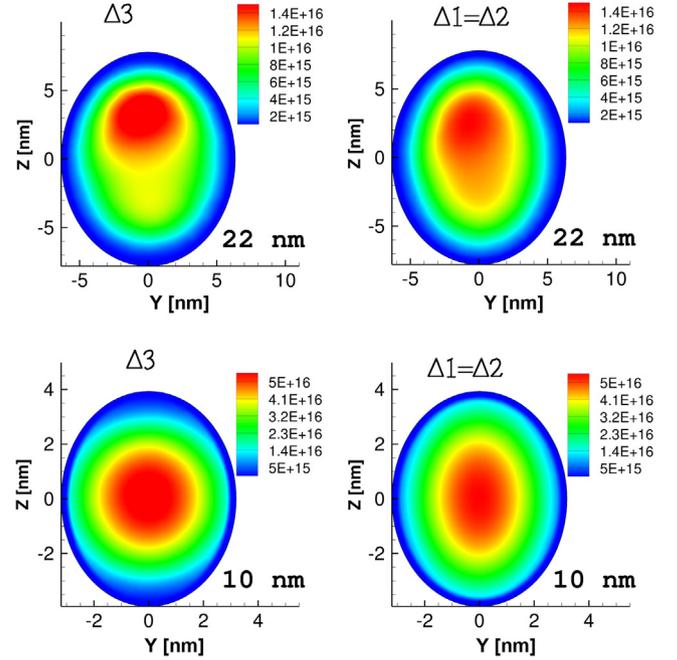


Fig. 6. Eigenmodes ($|\psi(y, z)|^2$) corresponding to the lowest energy eigenvalue of the three Δ valleys, in the middle of the $\langle 110 \rangle$ channel for the 22 nm (top, at $V_D = 1.0$ V) and 10 nm (bottom, at $V_D = 0.7$ V) Si GAA NW at $V_G = 0.8$ V.

orientation, so they will have the same wavefunction). The asymmetry seen in the eigenmode of the $\Delta 3$ valley in the 22 nm gate NW FET (top-left) is the result of drain induced change into potential in the channel at a large applied drain bias of 1.0 V. This effect will not occur in the 10 nm gate device because this transistor, with a much stronger quantum confinement, has a much better control of the transport so that the drain induced change into potential is negligible.

Fig. 7 shows I_D - V_G characteristics (the current is normalised by nanowire perimeter (elliptical circumference = 20.29 nm)) for the scaled 10 nm gate length NW FET at $V_D = 0.05$ V and $V_D = 0.7$ V along the $\langle 100 \rangle$ and the $\langle 110 \rangle$ channel orientations obtained from the 3D FE SEQC MC. Table 1 compares device operating characteristics with gate lengths of 22 nm and 10 nm predicting that the scaling to the 10 nm gate will ensure superior electrostatic integ-

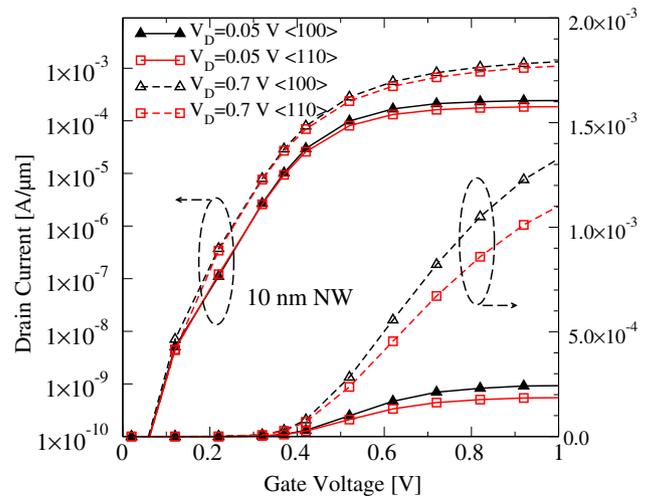


Fig. 7. I_D - V_G characteristics for the scaled 10 nm gate length GAA NW FET predicted by the 3D FE SEQC MC along the $\langle 100 \rangle$ and $\langle 110 \rangle$ channel orientations.

Table 1

V_T and sub-threshold slope (SS) at $V_D = 0.05$ V (LOW) and 1.0/0.7 V (HIGH) from the DD, DIBL from the DD and from the MC, and drive currents (I_{MC}) at $V_G = 1.0$ V comparing 22 and 10 nm GAA FETs.

Method	Gate length [nm]	22	10
MC	V_T [V]	0.3	0.35
DD	SS_{LOW} [mV/dec]	74	67
DD	SS_{HIGH} [mV/dec]	76	68
MC	$DIBL_{\langle 100 \rangle}$ [mV/V]	81	66
MC	$DIBL_{\langle 110 \rangle}$ [mV/V]	64	39
MC	$I_{\langle 100 \rangle}$ [$\mu A/\mu m$]	1222	1320
MC	$I_{\langle 110 \rangle}$ [$\mu A/\mu m$]	1000	1100

rity of a nearly ideal sub-threshold slope of 68 mV/dec and a DIBL of 39 mV/V and satisfactory on-current (the on-current increase is $\sim 8/10\%$ for the $\langle 100 \rangle/\langle 110 \rangle$ channel orientation with respect to the 22 nm NW).

The threshold voltage (V_T) is 0.3/0.35 V for the 22/10 nm gate length NW FET. The scaled GAA NW has a better sub-threshold slope (SS) at both low and high drain biases and a better drain-induced-barrier-lowering (DIBL) along the $\langle 100 \rangle$ and the $\langle 110 \rangle$ channel orientations. Due to the fact that electrons have a higher mobility in the $\langle 100 \rangle$ crystallographic orientation, both devices deliver a higher current for the $\langle 100 \rangle$ channel orientation than for the $\langle 110 \rangle$ channel at both low and high drain biases. For the 22 nm and the 10 nm GAA NW FETs, the drain current in the $\langle 100 \rangle$ channel device is larger than in that with the $\langle 110 \rangle$ channel by $\sim 30/20\%$ at a low/high drain bias.

3. Line-edge roughness (LER)

Downscaling transistors to the nano regime increases the undesirable performance mismatch in identically designed transistors [19]. The line-edge roughness (LER) is considered as one of the major sources of device variability [20] which may lead to serious device parameter fluctuations and limit the performance in the VLSI circuit applications. Therefore, studying the LER variability, especially in GAA NWs which have all the channel interfaces affected by the LER, is essential for predicting device behaviour in digital circuits. Here, we study the effect of uncorrelated LER (where we apply different LER profile at each side of the device, thus changing the width of the device across its length) using Fourier synthesis with Gaussian autocorrelation [21] implemented as described in Refs. [20,22]. The LER is characterised by a correlation length ($CL = \Lambda$), and a root mean square value ($RMS = \Delta$). The simulation method is based on the inverse discrete transformation and the application of a Gaussian filter over a list of random phases. The correlation length will be accounted for by the width of the Gaussian filter, and the amplitude will set the root mean square (RMS) value. To model the Fourier spectra, we use the following autocorrelation function:

$$S_G(\mathbf{k}) = \sqrt{\pi} \Lambda^2 \Lambda e^{(-k^2 \Lambda^2 / 4)},$$

The simulations of variability for the 22 nm and the 10 nm gate length NW FETs are carried out using the 3D density gradient (DG) quantum corrected FE DD with a LER correlation length (CL)

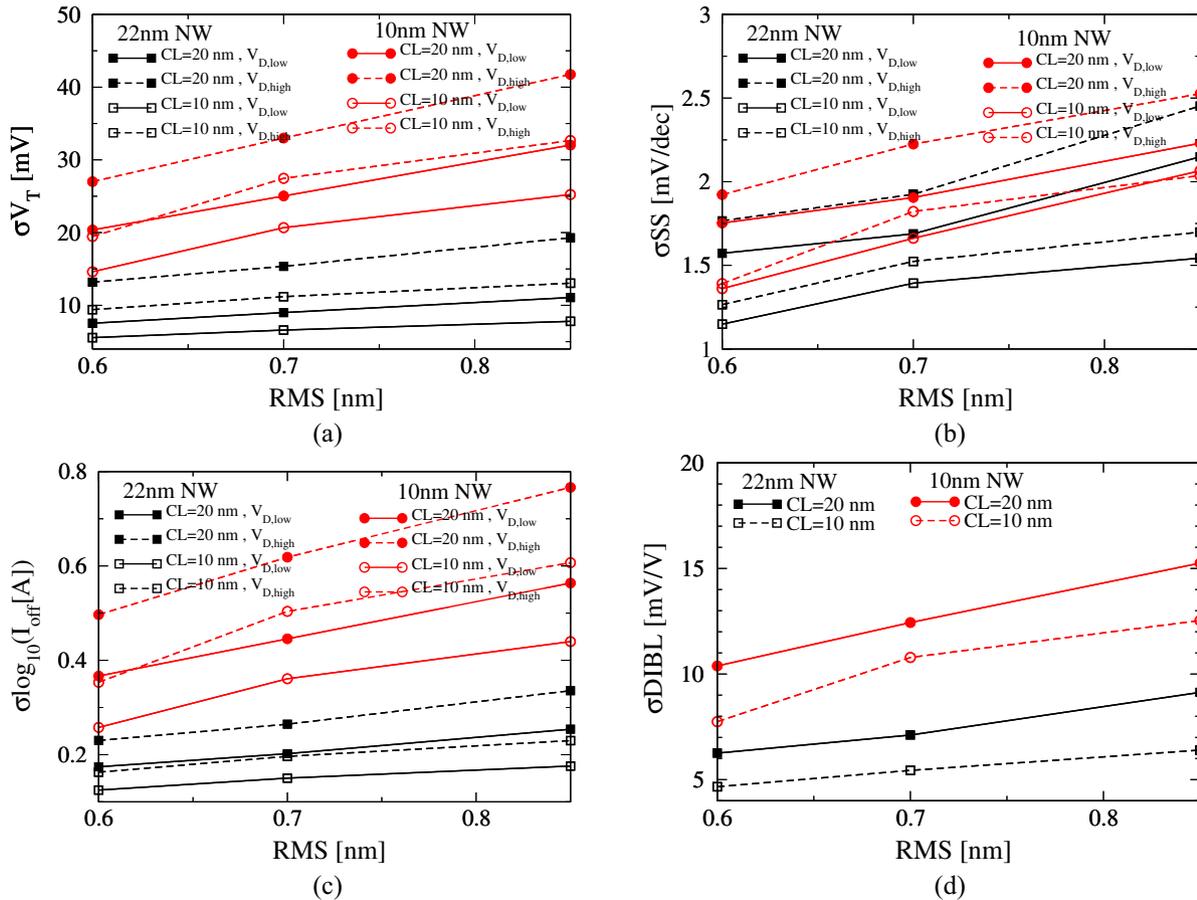


Fig. 8. Comparison of the V_T , SS, $\log_{10}(IOFF)$ and DIBL variability versus the RMS height due to LER for the studied 22 nm and 10 nm gate length GAA NW FETs as a function of the drain bias, and the correlation length (CL). $V_{D,low} = 0.05$ V and $V_{D,high} = 1.0$ V for the 22 nm gate length NW, and $V_{D,low} = 0.05$ V and $V_{D,high} = 0.7$ V for the 10 nm gate length NW.

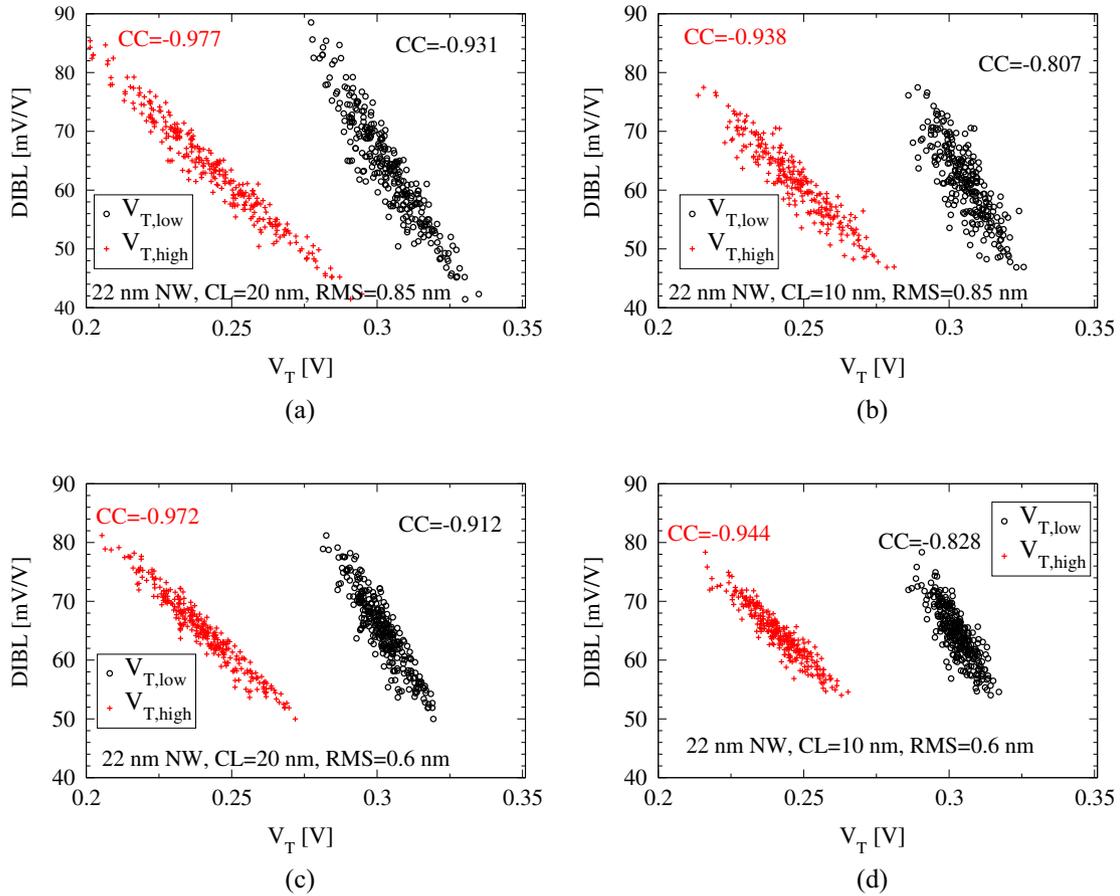


Fig. 9. Scatter plots showing the DIBL variation as a function of the V_T , at both low and high drain biases, due LER variations ($CL = 20/10$ nm and $RMS = 0.6/0.85$ nm) for the 22 nm gate GAA NW FET. $V_{D,low} = 0.05$ V and $V_{D,high} = 1.0$ V. Correlation coefficients (CC) are also calculated.

of 20 nm and 10 nm and three root mean square values ($RMS = 0.6, 0.7$ and 0.85 nm) chosen to represent the RMS values observed in experiments [2,3].

After we have calibrated DD-DG simulations to the results from 3D FE SEQC MC simulations, we analyse the LER-induced variability affecting the sub-threshold region of the device comparing four figures of merit: threshold voltage (V_T), OFF-current (I_{OFF}), sub-threshold slope (SS), drain-induced-barrier-lowering (DIBL). Ensembles of 300 devices have been used to investigate the LER-induced variability in the sub-threshold regions. We extract the threshold voltage using the fixed current approach and the OFF-current is extracted at $V_G = 0$ V. Fig. 8 shows a comparison of the V_T , SS, $\log_{10}(I_{OFF})$ and DIBL variability due to the LER for the studied 22 nm and 10 nm gate length GAA NWs as a function of the drain bias, the correlation length, and the RMS height. In the presence of LER, the observed variations for the four figures of merit are smaller in the 22 nm gate length GAA NW at low and high drain biases than the ones observed in the 10 nm gate length GAA NW. As expected, the standard deviations for the four figures of merit are increasing with the increase of the RMS value. Note here that the standard deviations for the four figures of merit are strongly affected by the drain bias and the correlation length values in both NWs. The standard deviations for the four figures of merit are increasing with the increase of the correlation length value, and also with the increase on the drain bias.

Fig. 9 shows the DIBL variability as a function of V_T at low and high drain biases due to LER with $CL = (20$ and 10 nm) and $RMS = (0.6$ and 0.85 nm) for the 22 nm gate length NW. In all cases, the DIBL shows strong negative correlations with $V_{T,low}$ (correlation

coefficient (CC) ranges from -0.807 to -0.931) and shows even stronger correlations with $V_{T,high}$ (CC ranges from -0.938 to -0.977). The larger the CC value, the less sensitive the variability is to a change in the drain bias. The Q-Q plot (not shown here) indicates near-to-Gaussian behaviour especially at a high drain bias.

Fig. 10 shows the DIBL variability as a function of V_T at low and high drain biases due to LER with $CL = (20$ and 10 nm) and $RMS = (0.6$ and 0.85 nm) for the 10 nm gate length NW FETs. In all cases, the DIBL shows larger strong negative correlations with $V_{T,high}$ (CC ranges from -0.969 to -0.988) than with $V_{T,low}$ (CC ranges from -0.945 to -0.979). We can see that, for the 10 nm gate length NW, the CC values are larger than those for the 22 nm gate length NW. The Q-Q plot indicate more Gaussian behaviour as expected in a larger device but the DIBL for some specific 10 nm gate length devices can overtake the DIBL in the 22 nm one.

Fig. 11 shows the scatter plots of the threshold voltages at a high drain bias ($V_{T,high}$) against the threshold voltages at a low drain bias ($V_{T,low}$) for the 22 nm gate length Si GAA NW with $CL = (20$ and 10 nm) and $RMS = (0.6$ and 0.85 nm). The threshold voltage at low and high drain biases are strongly correlated so we have used the same ranges for both the horizontal and vertical axis to show clearly the different behaviours. We can see that the threshold voltages with $CL = 20$ nm have a larger CC value that those for the $CL = 10$ nm which means the device variability is less sensitive to the change at the drain bias. In addition, the threshold voltages at a low drain bias ($V_{T,low} = 0.05$ V) is more spread in the case of $CL = 20$ nm. Fig. 12 shows the OFF-current ($I_{OFF,high}$) versus the threshold voltages ($V_{T,high}$) at $V_D = 1.0$ V for the 22 nm GAA NW with $CL = (20$ and 10 nm) and $RMS = (0.6$ and 0.85 nm). The log

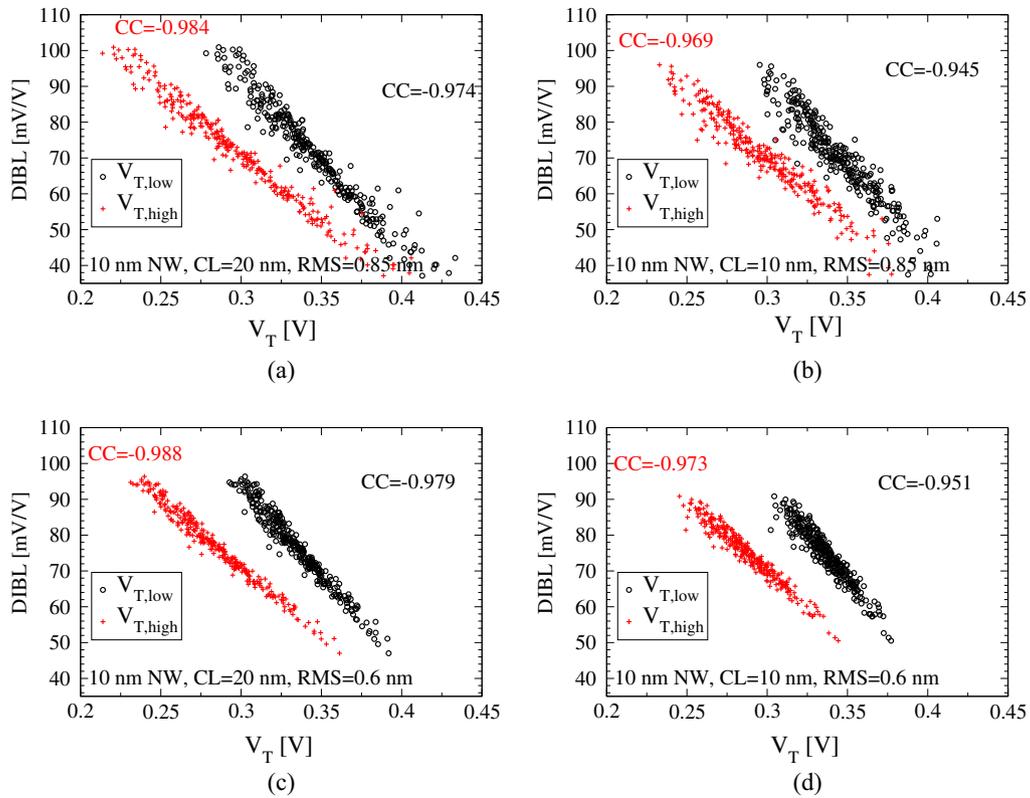


Fig. 10. Scatter plots showing the DIBL variation as a function of the V_T , at both low and high drain biases, due to LER variations (CL = 20/10 nm and RMS = 0.6/0.85 nm) for the 10 nm GAA NW. $V_{D,low} = 0.05$ V and $V_{D,high} = 0.7$ V. Correlation coefficients (CC) are indicated as well.

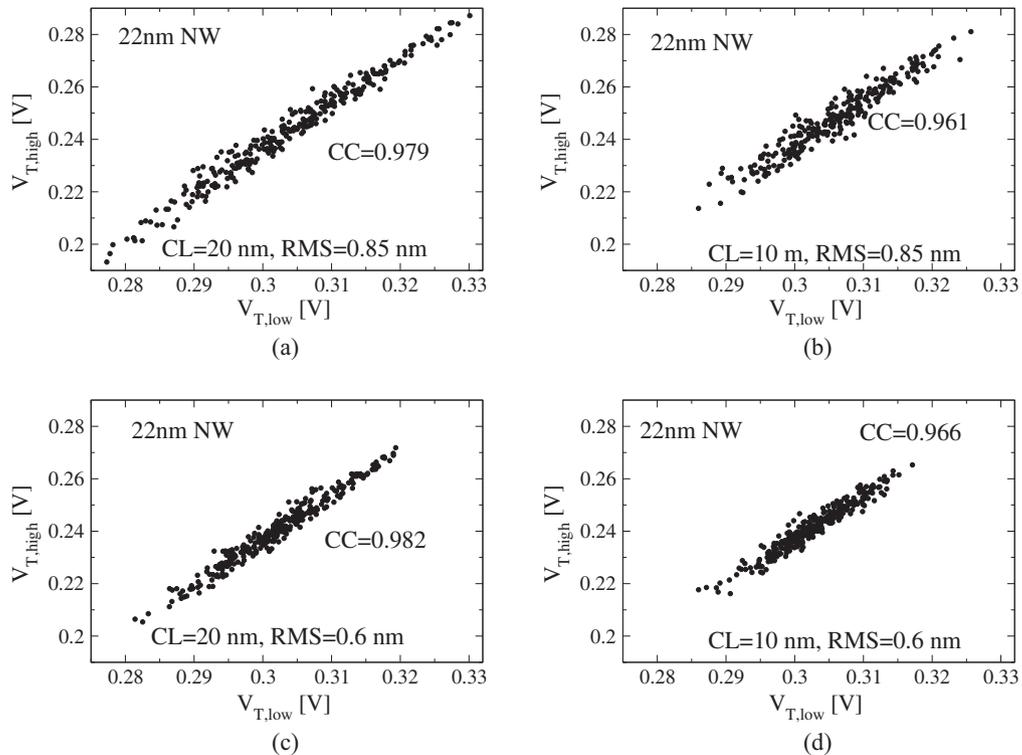


Fig. 11. Scatter plot showing the distribution of the threshold voltages at high drain bias ($V_{T,high}$) against the threshold voltages at low drain bias ($V_{T,low}$) for the 22 nm GAA NW with LER (CL = 20, 10 nm and RMS = 0.6, 0.85 nm) with respective correlation coefficients (CC).

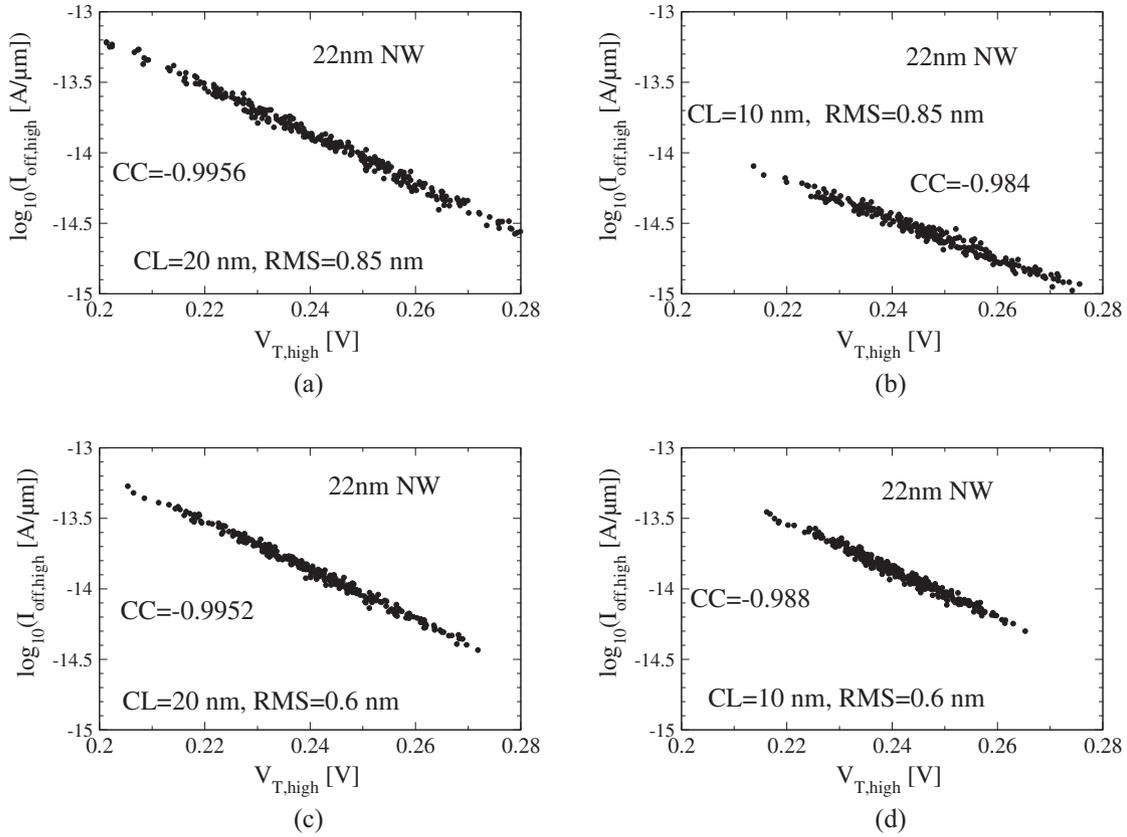


Fig. 12. $\log_{10}(I_{OFF,high})$ vs. $V_{T,high}$ at $V_D = 1.0$ V for the 22 nm GAA NW FET with the LER (CL = 20, 10 nm and RMS = 0.6, 0.85 nm).

of the OFF-current exhibits the typical linear dependence on the decreasing $V_{T,high}$ suggesting near-to-Gaussian behaviour as expected. Again, we can see that the variability with CL = 20 nm have a larger CC value than those for the CL = 10 nm, a characteristic of a lower variability in the SS.

4. Conclusion

We have employed our 3D SEQC FE MC simulation toolbox to obtain the I_D - V_G characteristics of a 22 nm gate length GAA Si NW FET. The simulation toolbox accurately describes the nanoscale geometry of multi-scale transistors using the FE method, and employs a completely parameter-free model of carrier transport which uses fully anisotropic transport model together with fully anisotropic quantum corrections which dependent on the valley orientation (longitudinal and transverse electron effective masses orientation along the device channel). The I_D - V_G characteristics at low and high drain biases obtained from the 3D MC toolbox demonstrated exceptional agreement with the experimental data [2] without any additional post-processing of lumping access resistance. We have then scaled the GAA Si NW FET to the 10 nm gate length and predicted that the scaled device will deliver an on-current of 1320/1100 $\mu\text{A}/\mu\text{m}$ for the $\langle 100 \rangle / \langle 110 \rangle$ channel with superior electrostatic integrity of a nearly ideal sub-threshold slope of 68 mV/dec and a DIBL of 39 mV/V. Finally, we have studied the effects of LER-induced variability on the sub-threshold characteristics (V_T , I_{OFF} , SS and DIBL) for both the 22 nm and the 10 nm gate length GAA NWs. Our simulations indicate that the 22 nm gate length NW is less sensitive to variability than the 10 nm one. We have found that the LER induced variability of the threshold voltage in the 22 nm gate length GAA NW FETs exhibits σV_T of about 9.5–19.2 mV and $\sigma \log_{10}(I_{OFF})$ of about 0.16–0.33 A at high

drain bias. The LER induced variability for the 10 nm GAA NW FETs is much larger. The variability of the threshold voltage, σV_T , is about 19.5–42 mV and, the variability of the OFF-current, $\sigma \log_{10}(I_{OFF})$, is about 0.35–0.76 A at a high drain bias. The 22 nm gate length GAA NW shows smaller variations for the four figures of merit at low and high drain biases than the ones observed in the 10 nm gate length GAA NW (see Fig. 8) as expected but the increase in the device variability is relatively small when comparing correlation coefficients (CC) [6,7]. This demonstrates that the GAA NW FETs are strong candidates for future generation of digital transistors delivering large on-current required in a circuit design accompanied by a well controlled device variability.

References

- [1] Ferain I et al. Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* 2011;479:310–6.
- [2] Bangsaruntip S et al. Density scaling with gate-all-around silicon nanowire MOSFETs for the 10 nm node and beyond. *IEDM Tech Dig* 2013:526–9.
- [3] Bangsaruntip S et al. High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling. *IEDM Tech Dig* 2009:297–300.
- [4] Suk SD et al. Investigation of nanowire size dependency on TSNWFET. *IEDM Tech Dig* 2007:1129–31.
- [5] Linton T et al. Determination of the line edge roughness specification for 34 nm devices. *IEDM Tech Dig* 2002:303–6.
- [6] Seoane N, Indalecio G, Aldegunde M, Nagy D, Elmessary MA, García-Loureiro AJ, et al. Comparison of fin-edge roughness and metal grain work function variability in InGaAs and Si FinFETs. *IEEE Trans Electron Dev* 2016;63(3):1209–16.
- [7] Kim SD, Wada H, Woo Jason CS. TCAD-based statistical analysis and modeling of gate line-edge roughness effect on nanoscale MOS transistor performance and scaling. *IEEE Trans Semicond Manuf* 2004;17(2):192–200.
- [8] Elmessary MA, Nagy D, Aldegunde M, Lindberg J, Dettmer WG, Perić D, et al. Anisotropic quantum corrections for 3-D finite-element Monte Carlo simulations of nanoscale multigate transistors. *IEEE Trans Electron Dev* 2016;63(3):933–9.

- [9] Lindberg J, Aldegunde M, Nagy D, Dettmer WG, Kalna K, García-Loureiro AJ, et al. Quantum corrections based on the 2D Schrödinger equation for 3D finite element Monte Carlo simulations of Nanoscaled FinFETs. *IEEE Trans Electron Dev* 2014;61(2):423–9.
- [10] Aldegunde M, García-Loureiro AJ, Kalna K. 3D finite element Monte Carlo simulations of multigate nanoscale transistors. *IEEE Trans Electron Dev* 2013;60(5):1561–7.
- [11] Nagy D, Elmessary MA, Aldegunde M, Valin R, Martinez A, Lindberg J, et al. 3D finite element Monte Carlo simulations of scaled Si SOI FinFET with different cross-sections. *IEEE Trans Nanotechnol* 2015;14(1):93–100.
- [12] ITRS(2012). *International technology roadmap for semiconductors* [Online]. Available: <<http://www.itrs.net/Links/2012ITRS/Home2012.htm>>.
- [13] Jacoboni C, Lugli P. *The Monte Carlo method for semiconductor device simulation*. Wien-New York: Springer-Verlag; 1989.
- [14] Ridley BK. Reconciliation of the Conwell-Weisskopf and Brooks-Herring formulae for charged-impurity scattering in semiconductors: third-body interference. *J Phys C: Solid State Phys* 1977;10(10):1589–93.
- [15] Islam A, Kalna K. Monte Carlo simulations of mobility in doped GaAs using self-consistent Fermi-Dirac statistics. *Semicond Sci Technol* 2011;26(5):055007. 9pp.
- [16] Ferry DK. *Semiconductor transport*. Taylor & Francis; 2000.
- [17] Garcia-Loureiro AJ et al. Implementation of the density gradient quantum corrections for 3D simulations of multigate nanoscaled transistors. *IEEE Trans Comput-Aided Des Integr Circ Syst* 2011;30(6):841–51.
- [18] Islam A, Benbakhti B, Kalna K. Monte Carlo study of ultimate channel scaling in Si and $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ bulk MOSFETs. *IEEE Trans Nanotechnol* 2011;10(6):1424–32.
- [19] Ban Yongchan, Sundareswaran Savithri, Pan David Z. Electrical impact of line-edge roughness on sub-45 nm node standard cells. *J Microlithogr Microfabr Microsyst (JM3)* 2010;9(4):041206.
- [20] Seoane N et al. Random dopant, line-edge roughness, and gate workfunction variability in a nano InGaAs FinFET. *IEEE Trans Electron Dev* 2014;61(2):466–72.
- [21] Asenov A, Kaya S, Brown AR. Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness. *IEEE Trans Electron Dev* 2003;50(5):1254–9.
- [22] Indalecio G, Aldegunde M, Seoane N, Kalna K, García-Loureiro AJ. Statistical study of the influence of LER and MGG in SOI MOSFET. *Semicond Sci Technol* 2014:045005.



Variability and self-average of impurity-limited resistance in quasi-one dimensional nanowires



Nobuyuki Sano

Institute of Applied Physics, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Nanowire
Impurity scattering
Variability
Phase interference
Self-average

ABSTRACT

The impurity-limited resistance in quasi-one dimensional (quasi-1D) nanowires is studied under the framework of the Lippmann–Schwinger scattering theory. The resistance of cylindrical nanowires is calculated theoretically under various spatial configurations of localized impurities with a simplified short-range scattering potential. Then, the relationship between the phase interference and the variability in the impurity-limited resistances is clarified. We show that there are two different and independent mechanisms leading to the variability in impurity-limited resistances; incoherent and phase-coherent randomization processes. The latter is closely related to the so-called “self-average” and its physical origin under nanowire structures is clarified. We point out that the ensemble average also comes into play in the cases of long channel nanowires, which leads to the self-average resistance of multiple impurities.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Si nanowires have been receiving great attention in the past few decades because of their possible application of future electronic and photonic devices [1,2]. However, because of their small structures in size, the device performance often fluctuates over great ranges, depending on the configuration of localized impurities in the substrate [3,4]. So far, theoretical studies on such variability observed in mobility or resistance of nanowires are limited with large-scale numerical simulations [5–10]. It has been demonstrated that the transport properties indeed fluctuate in short channel nanowires. Clearly, the phase interference would be of crucial importance in understanding the physics behind the variability.

It should be noted that the phase interference also plays a dominant role in self-averaging the transport properties such as resistance in *long channel* devices, in which many impurities are distributed uniformly in the substrate. Many different configurations of impurities in the substrate allows us to use the space-average impurity scattering rates in calculating the mobility etc, although the precise impurity configuration is different for each device. This is often referred to as “self-average” and the phase interference is deeply involved in its averaging mechanism [11]. Despite its importance, however, almost no attention has been paid so far on the interference effects among multiple impurities on transport properties under nanowire structures.

In the present paper, we study the interference effects associated with localized impurities on impurity-limited resistances in the quasi-1D nanowires. This is carried out based on the theory employed recently by the present author [12] and we clarify how and why the phase interference leads to the variability in the impurity-limited resistance. Then, the physical origin of “self-average” emerged under the fully coherent circumstances is clarified.

2. Theoretical foundations

2.1. Impurity-limited resistance

We consider a cylindrical nanowire with the radius of $r_s = 2$ nm and the impurity density in the substrate is assumed to be uniform ($n_{imp} = 2 \times 10^{19} \text{ cm}^{-3}$), i.e., localized impurities are distributed uniformly in the channel region. The channel length L of the wire changes in accordance with the number of impurities doped in the channel region, namely, $L = 4$ nm for one impurity, 8 nm for two impurities, etc. In addition, the extreme quantum limit, in which only the lowest subband is involved in electron transport, is assumed in which the phase interference is most effective.

Theoretical expressions of the impurity-limited resistance due to localized impurities in the nanowires are derived from the Landauer formula under the *linear response regime*. The conductance G of the doped channel region is calculated from the transmission coefficient $T_A(E)$ of the in-coming electrons in the lowest subband A with total energy E from the reservoirs (source and drain):

E-mail address: sano.nobuyuki.gw@u.tsukuba.ac.jp

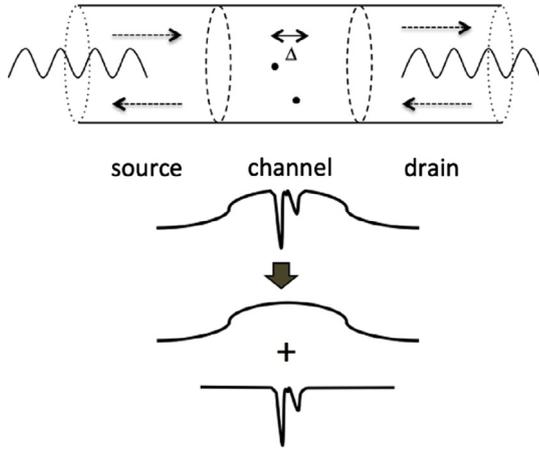


Fig. 1. Schematic drawings of the nanowire with two impurities doped in the channel and the corresponding potential profile under the applied gate voltage. The doped impurities are assumed to be donors so that the interaction is attractive. Under the framework of the scattering theory, the whole potential modulation is treated as a single scattering potential. The whole scattering potential consists of the long- and short-range potential modulation and the impurity-limited resistance results from the short-range screened scattering potential.

$$G = \frac{e^2}{\pi\hbar} \int_{-\infty}^{\infty} dE T_A(E) \left(-\frac{\partial f_{FD}(E)}{\partial E} \right), \quad (1)$$

where $f_{FD}(E)$ is the Fermi–Dirac distribution in the reservoirs.

It should be noted that the total resistance R_{tot} given by the inverse of G is comprised of two contributions; the contact resistance and the channel resistance [12]. The former results from the difference in the number of modes between the lead and the reservoirs, whereas the latter is due to the scattering potential by ionized impurities, phonons, surface roughness, and the long-range potential modulation. This long-range potential modulation is, furthermore, separated into the long-range part of the scattering Coulomb potential of ionized impurities/carriers and the electrostatic potential induced by the applied gate voltage,¹ as schematically drawn in Fig. 1. It should be noted that, in general, this long-range potential modulation is nearly irrelevant to the channel resistance and simply attributed to the carrier density modulation in the channel [12]. As the channel length shrinks, direct tunneling from the source to the channel and/or the drain regions comes into play and, then, the finite resistance due to the long-range potential modulation shows up.

In the present study, the separation between the long- and short-range parts of the scattering potential is always possible since the short-range part is defined as the singular part of the scattering potential and approximated by the delta-function potential. The rest of the potential should be regarded as the long-range part and attribute to the potential modulation through the Poisson equation, rather than the scattering potential. However, such separation for the *realistic* Coulomb potential might look nontrivial under nanostructures as the spatial distance between the impurities could be so small. This is, however, equivalent to asking whether or not the higher-order corrections associated with the interaction strength of impurity scattering is significant. In this case, the multiple impurities should be treated as a coherent single scattering center and the separation of the long- and short-range parts is made possible via Fourier transform of the potential.

In order to extract the impurity-limited resistance from the total resistance, we intentionally eliminate phonon scattering and consider only impurity scattering throughout this study.

¹ The work-function difference between the substrate and the gate also contributes to this long-range potential modulation.

Furthermore, the resistance caused by the long-range potential modulation mentioned above is ignored by assuming that the channel potential is flat along the wire axis. This could be achieved by properly imposing the gate voltage. This also implies that we ignore the direct tunneling from the source to the channel and/or the drain regions. As a consequence, the impurity-limited resistance is obtained by simply subtracting the contact (quantum) resistance from the total resistance, $R_s = R_{tot} - R_0$, and, thus, given by

$$R_s = \frac{\pi\hbar}{e^2} \frac{\langle R_A(E) \rangle}{1 - \langle R_A(E) \rangle} = R_0 \left\{ \langle R_A(E) \rangle + (\langle R_A(E) \rangle)^2 + \dots \right\} \quad (2)$$

with

$$\langle R_A(E) \rangle = \int_{-\infty}^{\infty} dE R_A(E) \left(-\frac{\partial f_{FD}(E)}{\partial E} \right). \quad (3)$$

Here, the contact (quantum) resistance is given by $R_0 = \pi\hbar/e^2$ and $R_A(E) (= 1 - T_A(E))$ is the reflection coefficient of the in-coming electrons with total energy E in the lowest subband A . We also assume that the chemical potential in the reservoirs is well above the bottom of the lowest subband in the channel. If R_s is truncated by the first term in the last expression of Eq. (2), the resulting expression corresponds to the usual Born approximation provided that the reflection probability $\langle R_A(E) \rangle$ is also approximated by the lowest order with respect to the interaction strength.

2.2. Transmission and reflection coefficients

The transmission and the reflection coefficients, $T_A(E)$ and $R_A(E)$, are calculated from the asymptotic forms of the scattered wave function by solving the Lippmann–Schwinger (LS) equation given by

$$|\psi^+(E)\rangle = |\phi(E)\rangle + \hat{C}_0^+(E) \hat{V} |\psi^+(E)\rangle, \quad (4)$$

where $|\psi^+(E)\rangle$ is the scattered state vector with total energy E , $|\phi(E)\rangle$ is the unperturbed state vector, \hat{V} is the scattering potential operator, and $\hat{C}_0^+(E)$ is the unperturbed (retarded) Green operator [12–15]. The unperturbed state vector $|\phi(E)\rangle$ is the eigenstate of the unperturbed Hamiltonian \hat{H}_0 , which is expressed by

$$\hat{H}_0 = -\frac{\hbar^2}{2m} \nabla^2 + U_{cyl}(\mathbf{R}) \quad (5)$$

under the effective mass approximation for the cylindrical wire. Here, m is the electron effective mass and $U_{cyl}(\mathbf{R})$ is the single-particle potential energy which confines the electrons inside the cylindrical wire. Assuming that the electrons are confined by the infinite potential barrier, the unperturbed eigenfunction in the lowest subband A is simply given by

$$\phi_A(\mathbf{R}) = \frac{1}{\sqrt{L}} e^{ikz} \zeta_A(\mathbf{r}) = \frac{1}{\sqrt{L}} e^{ikz} \frac{1}{\sqrt{\pi r_s J_1(x_{01})}} J_0\left(x_{01} \frac{r}{r_s}\right), \quad (6)$$

where $\mathbf{R} = (\mathbf{r}, z) = (r, \varphi, z)$ in the cylindrical coordinates, L is the wire (channel) length, $\zeta_A(\mathbf{r})$ is the subband wavefunction, $J_0(x)$ is the 0-th order Bessel function, and x_{01} is the first root of $J_0(x) = 0$. The total electron energy E in the lowest subband is then given by

$$E = \varepsilon_k + \varepsilon_{01} = \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{2m} \left(\frac{x_{01}}{r_s} \right)^2. \quad (7)$$

As for the scattering potential due to localized impurities, we employ the short-range δ -function potential defined by

$$V(\mathbf{R}) = \sum_{r=1}^{N_{imp}} (v_c a S) \delta^{(3)}(\mathbf{R} - \mathbf{R}_{0r}), \quad (8)$$

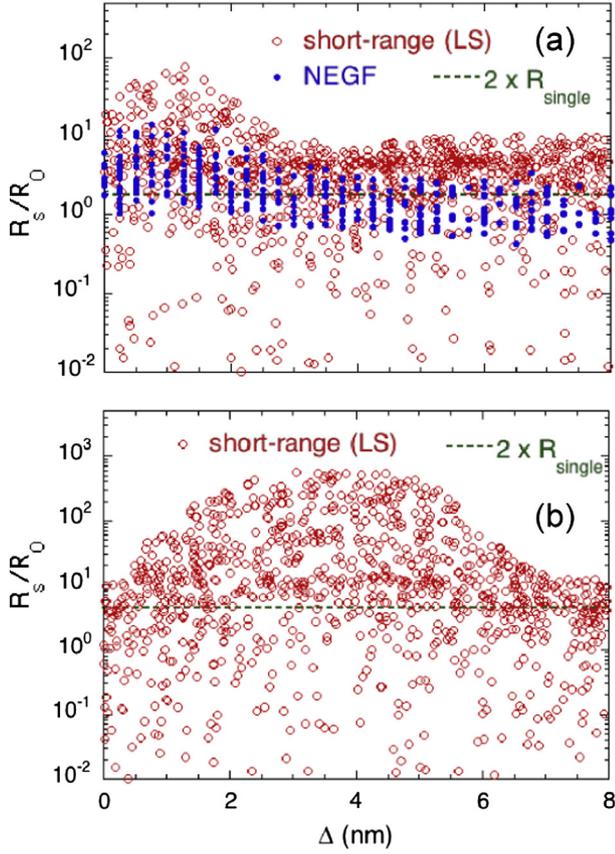


Fig. 2. Impurity-limited resistances of two impurities for 1000 different configurations as a function of the impurity separation Δ along the wire axis for (a) $T = 300$ K and (b) 30 K. The solid symbols in (a) represent the results from the NEGF simulations for acceptor impurities. The horizontal dashed lines show $2R_s^{\text{single}}$, with R_s^{single} being the single-impurity resistance averaged over the impurity configurations.

where N_{imp} is the number of impurities in the channel, v_c is the scattering potential energy, a ($= 0.5$ nm) is the characteristic length along the axis direction over which the scattering potential is effective, and S ($= \pi r_s^2$) is the cross-sectional area of the wire. $\mathbf{R}_{0r} = (\mathbf{r}_{0r}, z_{0r})$ is the position of the r -th impurity ($r = 1, 2, \dots, N_{\text{imp}}$). As already discussed in our previous paper [12], the expression employed here for the impurity scattering potential oversimplifies the reality: The present model represents only the singular part of the Coulomb potential due to ionized impurities and corresponds to the cases of extremely strong screening. Since the long-range part of the potential always smoothens the fluctuations in the transport characteristics, the present scattering potential somewhat exaggerates the phase interference effects among the localized impurities. Nevertheless, as we shall show in Fig. 2, in which the present analyses are compared with more elaborate NEGF simulations, the present model does represent properly the essential features associated with the phase interference among multiple localized impurities.

Thanks to the simple form of the impurity scattering potential, the exact transmission and reflection coefficients could be analytically derived from the LS equation and given, respectively, by

$$T_A(E) = |1 + I^-(E)|^2 \quad (9)$$

and

$$R_A(E) = |I^+(E)|^2, \quad (10)$$

where $I^\mp(E)$ is expressed as

$$I^\mp(E) = \left(e^{\mp ikz_{01}} \quad e^{\mp ikz_{02}} \quad \dots \quad e^{\mp ikz_{0N_{\text{imp}}}} \right) \left(-i\hat{\Gamma}(E) \right) \left(\frac{1}{1 + i\hat{\Sigma}(E)} \right) \begin{pmatrix} e^{ikz_{01}} \\ e^{ikz_{02}} \\ \vdots \\ e^{ikz_{0N_{\text{imp}}}} \end{pmatrix} \quad (11)$$

by employing the matrix representation with respect to impurity site indices. The matrix elements of $\hat{\Gamma}(E)$ and $\hat{\Sigma}(E)$ are then defined by

$$(\hat{\Gamma}(E))_{rs} = \frac{ma}{\hbar^2 k} v_c S |\xi_A(\mathbf{r}_{0r})|^2 \equiv \gamma_r(E) \delta_{r,s} \quad (12)$$

and

$$(\hat{\Sigma}(E))_{rs} = e^{ik|z_{0r} - z_{0s}|} \gamma_s(E) = e^{ik\Delta_{rs}} \gamma_s(E), \quad (13)$$

where $\Delta_{rs} = |z_{0r} - z_{0s}|$ and $\xi_A(\mathbf{r}_{0r})$ is the subband wavefunction of the r -th impurity site. It should be noted that $\gamma_r(E)$ is a dimensionless real number and represents the strength of the coupling between electron and impurity.

As a concrete example, the exact expression of the reflection coefficient $R_A(E)$ of two localized impurities is given by

$$R_A(E) = \frac{\Lambda(E)}{1 + \Lambda(E)} \quad (14)$$

with

$$\Lambda(E) = \gamma_1^2 + \gamma_2^2 + 2\gamma_1^2\gamma_2^2 + 2\gamma_1\gamma_2\{(1 - \gamma_1\gamma_2) \cos(2k\Delta) + (\gamma_1 + \gamma_2) \sin(2k\Delta)\}, \quad (15)$$

where $\Delta = \Delta_{12} = |z_{01} - z_{02}|$ and Δ is the separation between the two impurities along the wire axis direction. Notice that Eq. (14) is bounded above and its supremum is unity, as it should be.

3. Results and discussion

3.1. Variability and phase interference

We first show the impurity-limited resistance R_s of two localized impurities for 1000 different configurations inside the channel as a function of the impurity separation Δ for $T = 300$ and 30 K in Fig. 2. The scattering potential energy is set at $v_c = 183$ meV, corresponding to the screening length of $\lambda_{\text{sc}} = 2$ nm [12]. For comparison, similar results from more elaborate NEGF simulations [16] are also shown with solid symbols in Fig. 2(a).

In the NEGF simulations, we have carried out the simulations for more than 500 different impurity configurations. Since it is not possible to carry out the simulations coupled self-consistently with the Poisson equation so many times, the impurity scattering potential has been approximated with the Yukawa potential with the image charges associated with the interface with the gate. This approximation is not crucial because our main concern is to find out how the difference in shape of the scattering potential, namely a finite extension of the scattering potential in real space, affects the variability associated with the phase interference. In addition, the device structure employed in the NEGF simulations is a square nanowire with the side length of 3.5 nm. The difference in shape is again insignificant because the cross-sectional area is very similar in both cases.

The fluctuations associated with the subband wavefunction (and the screened Coulomb potential) are greatly suppressed, compared with the case of the short-range scattering potential in the LS approach, due to the long-range nature of the scattering potential. However, the variations of R_s along the wire axis direction are very similar to those calculated from the LS equation: R_s is large at small

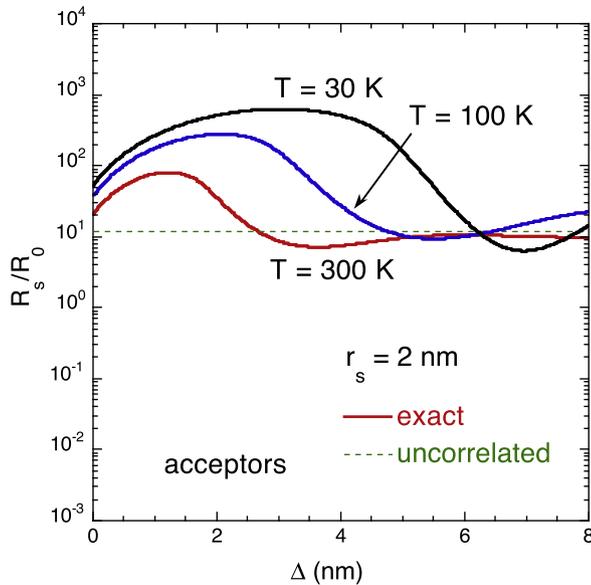


Fig. 3. Impurity-limited resistance of two impurities from the LS equation at $T = 30, 100, 300$ K as a function of the impurity separation Δ . Both impurities are located on the wire axis. The horizontal dotted line shows the uncorrelated value $2R_{single}$, with R_{single} being the single-impurity resistance located on the wire axis.

Δ , whereas R_s approaches to some constant value and, thus, the variation along the wire axis direction vanishes at large Δ .

In either case of the LS or NEGF, the value of the resistances R_s scatters over a few orders of magnitudes and such large fluctuations result in the two different physical origins: The fluctuations in R_s at fixed Δ are attributed to the variations in γ_r (the subband wavefunctions), whereas the fluctuations along the wire axis direction Δ are due to the trigonometric function dependence in Eq. (15) and, thus, due to the phase interference of electrons among the impurities. Therefore, the former has nothing to do with the phase correlation among multiple impurities²; each impurity could be regarded as nearly independent. Since the central-limit theorem could be applied in this case [17], the fluctuations with respect to the vertical direction in Fig. 2 would diminish under some particular impurity configuration as the channel length becomes longer.

On the other hand, the variations in R_s along the Δ direction do not generally vanish even for long channel wires, as noted by Kohn and Luttinger [11]. In order to demonstrate this point, we eliminate the fluctuations associated with the subband wavefunctions; we carry out similar calculations by placing two impurities on the wire axis for three different temperatures, $T = 30, 100,$ and 300 K. The calculation results are shown in Fig. 3 along with the uncorrelated value $2R_{single}$ for $T = 300$ K, where R_{single} is the resistance of the single-impurity located on the wire axis. We would like to mention that the Born approximation completely breaks down and the coupling between the electron and impurity is enhanced due to the confinement in nanostructures [12]. A large oscillatory behavior in R_s is observed in the first few nm at any temperature. This oscillation results from the trigonometric function dependence in the reflection coefficient $R_A(E)$ in Eq. (15) and represents the constructive phase interference among the two impurities. This interference effect becomes very strong at low temperature and would lead to the Anderson localization if the coupling between the electron and impurity is strong enough. However, at room temperature, this oscillation rapidly damps and R_s approaches $2R_{single}$, that is, the uncorrelated limit.

² There exists a slight phase dependence due to the trigonometric function involved in $R_A(E)$ of Eq. (15), though Δ is fixed.

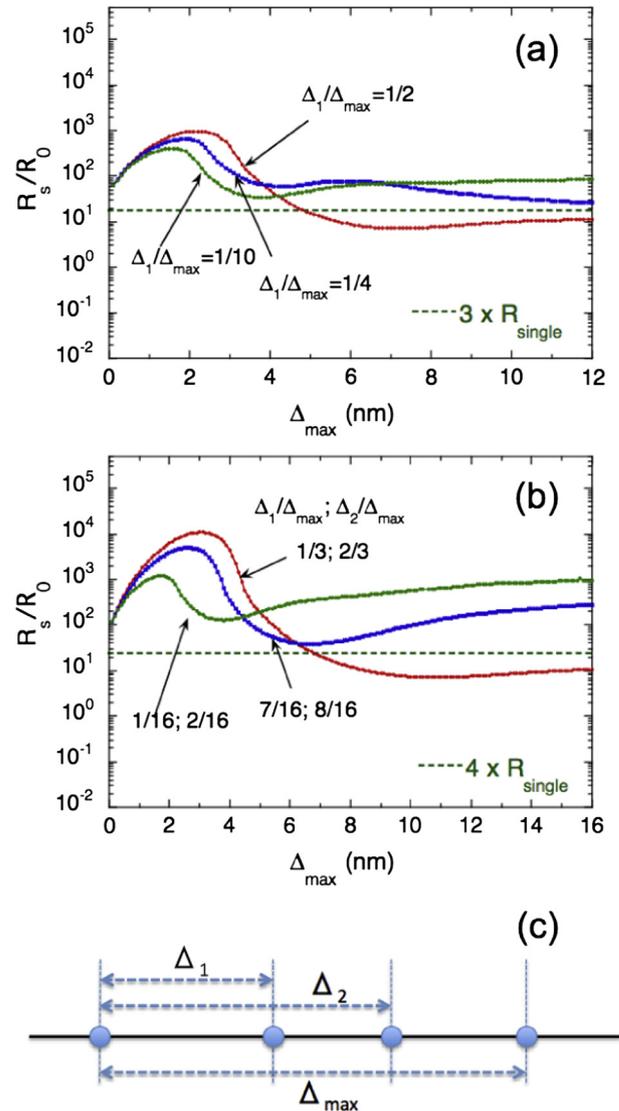


Fig. 4. Impurity-limited resistance of (a) three and (b) four impurities at $T = 300$ K as a function of the maximum separation among impurities Δ_{max} , as shown in (c), where the locations of impurities are schematically drawn. All impurities are located on the wire axis. The horizontal dotted lines show the uncorrelated values, $3R_{single}$ and $4R_{single}$, with R_{single} being the single-impurity resistance located on the wire axis.

The above findings also hold true for the cases of more than two impurities doped in the channel. Similar calculations of the impurity-limited resistance R_s are carried out by placing three or four impurities on the wire axis. The exact formulas of the reflection coefficients $R_A(E)$ are analytically derived from the LS equation. The calculation results for $T = 300$ K are shown in Fig. 4 as a function of the maximum separation among the impurities Δ_{max} . The locations of the impurities in between are assumed such that the ratios like Δ_1/Δ_{max} and Δ_2/Δ_{max} are fixed, where Δ_1 and Δ_2 are the locations of the impurities as drawn in Fig. 4(c). The normalized values of R_s of three different impurity configurations are shown for each case. The uncorrelated values are given by $3R_{single}$ and $4R_{single}$, and indicated with the horizontal dotted lines. Here, R_{single} is the single-impurity resistance located on the wire axis as before. Again, the variations in R_s along the wire axis direction are large at small Δ_{max} , whereas R_s nearly stays constant as Δ_{max} becomes large.

These results that R_s at large impurity separation approaches the uncorrelated value are rather surprising because no averaging

over the configurations of impurities nor energy dissipating scattering which randomizes the electron phase is included in the above analyses. In other word, the phase randomization is taking place *even under the fully coherent circumstances under some particular impurity configuration*. This is attributed to the self-average, as we shall discuss below.

3.2. Origin of self-average

The fact that the impurity-limited resistance of multiple impurities approaches the uncorrelated values at room temperature implies that the phase correlation among the impurities is somehow washed out as the impurity separation along the wire axis becomes large, namely, larger than some phase correlation length Δ_{phase} .

We find that the physical origin of this phase randomization along the wire axis direction is closely related to the broadness of the energy spectrum of in-coming electrons from the reservoirs (source and drain). That is, R_s is averaged by the in-coming electrons with many different kinetic energies (wavelengths) when the spectrum is broad (or equivalently, temperature of the reservoirs is high). Therefore, if the temperature is low enough such that the energy spectrum of the in-coming electrons is limited to be a very narrow range around the Fermi energy of the reservoirs, the phase coherence would last much longer distances and R_s deviates from the uncorrelated value even at large impurity separation. This is indeed confirmed from Fig. 3, in which the oscillation in R_s at $T = 30$ K is sustained over the entire channel region.

Notice that this is also consistent with the arguments given by Kohn and Luttinger [11] for “self-average.” They claim that some averaging procedure, in addition to averaging over the impurity configurations, is necessary to lead to self-average the transport properties under coherent circumstances. In their cases, the density of states for the final states after impurity scattering plays that role because their consideration is restricted to the electron gas in bulk. In the present case, the final states after scattering are rather limited in quasi-1D nanowires and it is not strong enough to lead to self-averaging. Instead, the broadness of the energy spectrum of the in-coming electrons from the reservoirs is attributed to another averaging process.

However, as is clear from Fig. 4, the values of R_s at large Δ_{max} for the cases of three and four impurities scatter to some extent around the uncorrelated values, contrary to the case of two impurities. This results from the fact that the transmission probability $T_A(E)$ is very close to zero at such large impurity separation ($\Delta_{max} \geq 10$ nm) and electrons hardly go through the channel region. As a result, R_s is greatly affected by a tiny fluctuation in $T_A(E)$ and scatter around the uncorrelated value. We should notice that at room temperature, phonon scattering is always inevitable even in nano-scale channels [18,19]. Since its mean-free-path is around 10 nm, the phase coherence is almost always destroyed at such large impurity separation. Then, the ensemble-average over various impurity configurations comes into play in the cases of long channel nanowires. In other words, thanks to the energy dissipating scattering, the variations shown at large Δ_{max} in Fig. 4 would vanish due to averaging over various impurity configurations as the channel length increases. This is indeed true as we shall show below.

3.3. Self-average versus ensemble average

Here, we would like to stress the difference between the self-average and the ensemble average of transport properties. The former averaging is taking place in a single nanowire under some fixed impurity configuration, whereas in the case of ensemble

average the average is taken literally over various impurity configurations.

In the present case, the phase correlation length Δ_{phase} where the constructive phase interference is significant is about 5 nm at room temperature, as seen in Figs. 3 and 4, and Δ_{phase} is much smaller than the channel length L (~ 8 –16 nm). Therefore, the phase correlation volume inside the nanowire is a fraction of the entire volume of the channel region. If one takes an ensemble average of the resistances over many nanowires where impurities are assumed to be distributed uniform, the ensemble average resistance becomes very close to the value of the uncorrelated resistance unless the scattering potential energy v_c is extremely large. That is, the ensemble average resistance would be a simple sum of R_{single} , where R_{single} is the ensemble average resistance of the single-impurity under the uniform impurity configurations. This is confirmed from Fig. 5, in which the ensemble average resistance of two impurities under various uniform impurity

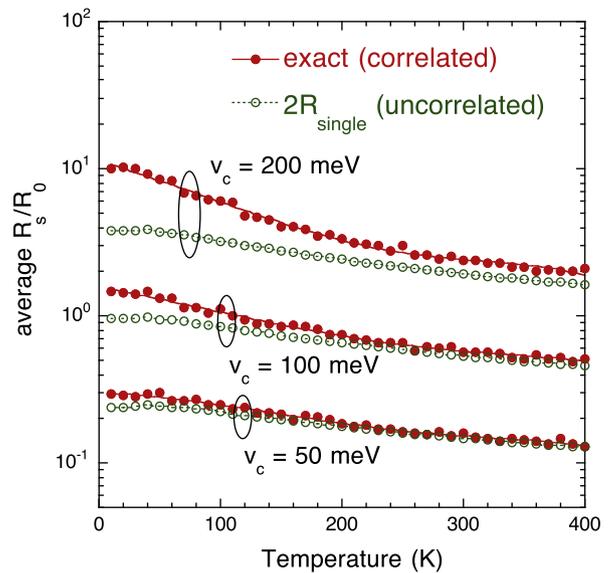


Fig. 5. Ensemble average resistance of two impurities under uniform impurity distributions as a function of temperature of the reservoirs for three different scattering potential energies, $v_c = 50, 100, 200$ meV. $2R_{single}$ represents the uncorrelated limit, where R_{single} is the ensemble average resistance of the single-impurity under uniform impurity configurations.

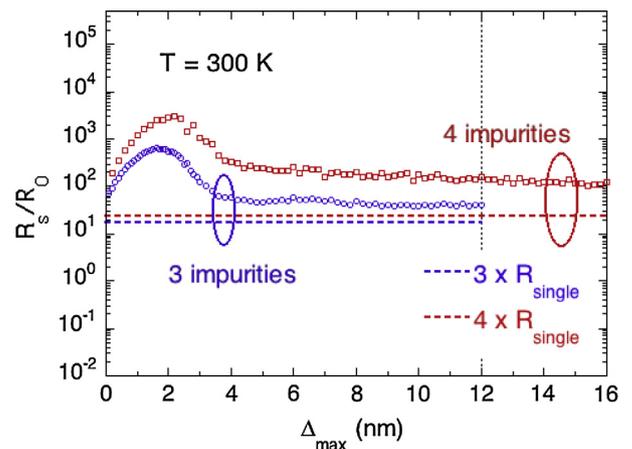


Fig. 6. Ensemble average resistances of three and four impurities at $T = 300$ K as a function of the maximum impurity separation Δ_{max} . Impurities in between are placed on the wire axis at random by fixing the maximum separation Δ_{max} among impurities. R_{single} represents the resistance of the single-impurity placed on the wire axis.

distributions is plotted as a function of temperature of the reservoirs for three different scattering potential energies, $v_c = 50, 100, 200$ meV. The uncorrelated average resistance, $2\bar{R}_{single}$, is also shown. The deviation between the two results, the exact and the uncorrelated ones, is large when the coupling strength v_c is large. However, as temperature increases, the deviation decreases even for $v_c = 200$ meV, which is close to the reality.

A similar scenario holds true even for the cases of three and four impurities: Fig. 6 shows the ensemble average resistance R_s of three or four impurities placed on the wire axis as a function of the maximum impurity separation Δ_{max} . The locations of the impurities in between are chosen at random with the fixed Δ_{max} . Since all impurities are placed on the wire axis, the average values of R_s are somewhat exaggerated. Nevertheless, the large variations at large Δ_{max} observed in Fig. 4 are greatly suppressed in both cases and the ensemble average resistance at room temperature becomes close to the uncorrelated value (within a factor of three or so). Hence, the classical Ohm's law is recovered. This is also consistent with our finding from more elaborate NEGF simulations, in which the ensemble average resistance is indeed proportional to the number of impurities [16].

4. Conclusions

We have investigated the variability in the impurity-limited resistance due to localized impurities in quasi-1D nanowires from the viewpoint of the phase interference. We have clarified the physical origin of the “self-average” of resistances under quasi-1D nanowire structures: Averaging is dominantly taking place through the broadness of the energy spectrum of in-coming electrons from the reservoirs. However, another averaging process is also required to achieve the “self-average” of the resistance under *long-channel* nanowire structures, namely, the ensemble average over impurity configurations. Since the energy-dissipating phonon scattering is always involved at room temperature, it breaks the phase coherence among the multiple impurities *at large impurity separations* (around 10 nm or more). Then, in addition to the averaging over the in-coming electrons with broad energy spectrum, the ensemble average also comes into play to self-average the resistance of multiple impurities in long channel nanowires. Thanks to this energy-dissipating scattering, the variations of R_s found at large impurity separation under fully coherent circumstances diminish. Then, the entire resistance becomes self-averaged and close to the uncorrelated value, namely, the series resistance of the single-impurity resistances. Hence, the classical Ohm's law is recovered.

Acknowledgment

This work was supported in part by Ministry of Education, Science, Sports, and Culture under Grant-in- Aid for Scientific Research (B) (No. 15H03983).

References

- [1] Appenzeller J, Knoch JJ, Bjork M, Riel H, Schmid H, Riess W. Toward nanowire electronics. *IEEE Trans Electron Dev* 2008;55(11):2827–45. doi: <http://dx.doi.org/10.1109/TFD.2008.2008011>.
- [2] Rurali R. Structural, electronic, and transport properties of silicon nanowires. *Rev Mod Phys* 2010;82(1):427–49. doi: <http://dx.doi.org/10.1103/RevModPhys.82.427>.
- [3] Sano N, Matsuzawa K, Mukai M, Nakayama N. Role of long-range and short-range coulomb potentials in threshold characteristics under discrete dopants in sub-0.1 μm Si-MOSFETs. In: IEEE international electron device meeting (IEDM2000); 2000. p. 275–8. <http://dx.doi.org/10.1109/IEDM.2000.904310>.
- [4] Sano N, Matsuzawa K, Mukai M, Nakayama N. On discrete random dopant modeling in drift-diffusion simulations: physical meaning of ‘atomistic’ dopants. *Microelectron Reliab* 2002;42(2):189–99. doi: [http://dx.doi.org/10.1016/S0026-2714\(01\)00138-X](http://dx.doi.org/10.1016/S0026-2714(01)00138-X).
- [5] Rurali R, Markussen T, Sune J, Brandbyge M, Jauho AP. Modeling transport in ultrathin si nanowires: charged versus neutral impurities. *Nano Lett* 2008;8(9):2825–8. doi: <http://dx.doi.org/10.1021/nl801409m>.
- [6] Persson MP, Mera H, Niquet YM, Delerue C, Diarra M. Charged impurity scattering and mobility in gated silicon nanowires. *Phys Rev B* 2010;82(11):115318. doi: <http://dx.doi.org/10.1103/PhysRevB.82.115318>.
- [7] Niquet YM, Mera H, Delerue C. Impurity-limited mobility and variability in gate-all-around silicon nanowires. *Appl Phys Lett* 2012;100(15):153119. doi: <http://dx.doi.org/10.1063/1.4704174>.
- [8] Martinez A, Aldegunde M, Seoane N, Brown A, Barker J, Asenov A. Quantum-transport study on the impact of channel length and cross sections on variability induced by random discrete dopants in narrow gate-all-around silicon nanowire transistors. *IEEE Trans Electron Dev* 2011;58(8):2209–17. doi: <http://dx.doi.org/10.1109/TFD.2011.2157929>.
- [9] Aldegunde M, Martinez A, Barker J. Study of discrete doping-induced variability in junctionless nanowire MOSFETs using dissipative quantum transport simulations. *IEEE Trans Electron Dev* 2012;33(2):194–6. doi: <http://dx.doi.org/10.1109/LED.2011.2177634>.
- [10] Sylvia S, Habib K, Khayer M, Alam K, Neupane M, Lake R. Effect of random, discrete source dopant distributions on nanowire tunnel FETs. *IEEE Trans Electron Dev* 2014;61(6):2208–14. doi: <http://dx.doi.org/10.1109/TFD.2014.2318521>.
- [11] Kohn W, Luttinger JM. Quantum theory of electrical transport phenomena. *Phys Rev* 1957;108(3):590–611. doi: <http://dx.doi.org/10.1103/PhysRev.108.590>.
- [12] Sano N. Impurity-limited resistance and phase interference of localized impurities under quasi-one dimensional nano-structures. *J Appl Phys* 2015;118(24):244302. doi: <http://dx.doi.org/10.1063/1.4938392>.
- [13] Newton RG. *Scattering theory of waves and particles*. New York: Dover; 2002. p. 228.
- [14] Taylor JR. *Scattering theory: the quantum theory of nonrelativistic collisions*. New York: Dover; 2000. p. 133.
- [15] DiVentra M. *Electrical transport in nanoscale systems*. Cambridge: Cambridge University Press; 2008. p. 119.
- [16] Sano N, Zuhidza MR, Kaneno Y, Honda S, Ueda A, Yoshida K. Correlation effects of localized impurities on electron transport under 1-d nano-structures. *J Phys: Conf Series* 2015;647(1):012028. doi: <http://dx.doi.org/10.1088/1742-6596/647/1/012028>.
- [17] Feller W. *An introduction to probability theory and its applications, vol. 1*. New York: Wiley; 1968.
- [18] Sano N. Kinetic study of velocity distributions in nanoscale semiconductor devices under room-temperature operation. *Appl Phys Lett* 2004;85(18):4208–10. doi: <http://dx.doi.org/10.1063/1.1812812>.
- [19] Sano N. Kinetics of quasiballistic transport in nanoscale semiconductor structures: is the ballistic limit attainable at room temperature? *Phys Rev Lett* 2004;93(24):246803. doi: <http://dx.doi.org/10.1103/PhysRevLett.93.246803>.



Drain current local variability from linear to saturation region in 28 nm bulk NMOSFETs



T.A. Karatsori^{a,b,*}, C.G. Theodorou^a, S. Haendler^c, C.A. Dimitriadis^b, G. Ghibaudo^a

^aIMEP-LAHC, INPG – Minatec, 3 Parvis Louis Néel, CS 50257, 38016 Grenoble, France

^bDepartment of Physics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

^cSTMicroelectronics, BP16, 38921 Crolles, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Mismatch
Local variability
Threshold voltage
Gain factor
Source-drain series resistance
Modeling
28 nm bulk
CMOS

ABSTRACT

In this work, we investigate the impact of the source – drain series resistance mismatch on the drain current variability in 28 nm bulk MOSFETs. For the first time, a mismatch model including the local fluctuations of the threshold voltage (V_t), the drain current gain factor (β) and the source – drain series resistance (R_{SD}) in both linear and saturation regions is presented. Furthermore, it is demonstrated that the influence of the source – drain series resistance mismatch is attenuated in the saturation region, due to the weaker sensitivity of the drain current variability on the series resistance variation. The experimental results were further verified by numerical simulations of the drain current characteristics with sensitivity analysis of the MOSFET parameters V_t , β and R_{SD} .

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Usually, process variations within CMOS technologies are categorized into global and local variations. Concerning the global variations, the device parameters change smoothly all over the wafer. On the other hand, for local variability or mismatch, each MOS transistor is affected differently even from its close neighbor. Mismatch is by nature an uncorrelated stochastic process, which is increasing due to the scaling down of CMOS technology. As a result, a crucial issue in MOSFETs and especially in advanced nano-scaled devices is the study of drain current local variability, as it affects the performance of analog but also digital circuits like SRAM cells.

According to the first mismatch studies, the main sources of the drain current, I_D mismatch are related to the local fluctuations of the threshold voltage, V_t and the current gain factor, β [1,2]. The source – drain (SD) series resistance mismatch and its impact on the drain current variability of FDSOI devices has been reported in recent works [3–5].

In this work, we study the SD series resistance mismatch on the drain current variability in bulk NMOS transistors processed with

28 nm Gate-first technology and extend for the first time the drain current local variability model in the saturation region as well.

2. Devices and experimental details

The devices measured in this work are bulk n-MOS transistors, issued from 28 nm planar CMOS technology with channel width (W) varying from 10 down to 0.08 μm and channel length (L) varying from 5 down to 0.03 μm . The devices were fabricated by ST Microelectronics in France and present also pocket implants. A sample of 70 pairs of identical MOS transistors (namely, MOS1 and MOS2), electrically independent with symmetric connections, spaced by the minimum allowed distance and laid out in an identical environment, was necessary for matching measurements. Drain current measurements in both linear ($V_D = 30$ mV) and saturation regions ($V_D = 1$ V) were performed with Agilent B1500 Semiconductor Device Analyzer. Fig. 1(a)–(d) present typical I_D - V_G characteristics in both linear and saturation regions, illustrating the single device variability over full wafer for the nominal device.

As in nominal dimension devices, the drain current difference between two paired transistors might reach several decades, the drain current mismatch, $\Delta I_d/I_d$, is no longer evaluated with a linear difference but using the log difference as [4],

* Corresponding author at: Polygone scientifique MINATEC, 3 Parvis Louis Néel, CS 50257, 38016 Grenoble Cedex 1 (Office 350), France.

E-mail address: karatsot@minatec.inpg.fr (T.A. Karatsori).

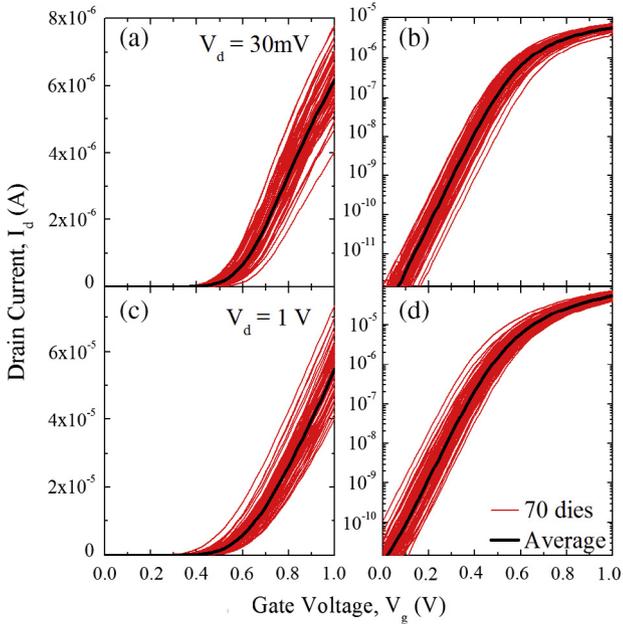


Fig. 1. Experimental $I_D(V_G)$ variability characteristics and their mean value in linear (a and c) and logarithmic scale (b and d) at low (a and b) and high (c and d) drain voltage of an ensemble of 70 bulk n-MOSFETs with channel width $W = 0.08 \mu\text{m}$ and channel length $L = 0.03 \mu\text{m}$.

$$\Delta I_D / I_D = \ln(I_{D2} / I_{D1}), \quad (1)$$

where I_{D1} and I_{D2} refer to the drain current of MOS1 and MOS2 pair.

3. Drain current mismatch modeling

In order to build a general drain current mismatch model, we have to take into account all the main sources of drain current local variability through the sensitivity of the drain current to the MOSFET parameters, i.e. V_t , β and R_{SD} . Therefore, considering the first order Taylor approximation, the drain current variation can be expressed as the sum of the principal contributions as [3],

$$\frac{dI_D}{I_D} = \left(\frac{1}{I_D} \cdot \frac{\partial I_D}{\partial V_t} \right) \cdot dV_t + \left(\frac{1}{I_D} \cdot \frac{\partial I_D}{\partial \beta} \right) \cdot d\beta + \left(\frac{1}{I_D} \cdot \frac{\partial I_D}{\partial R_{SD}} \right) \cdot dR_{SD} \quad (2)$$

After calculating the partial derivatives of I_D with respect to V_t , β and R_{SD} , the expression describing the drain current mismatch in the linear region is obtained as [3],

$$\sigma^2(\Delta I_D / I_D) = \left(\frac{g_m}{I_D} \right)^2 \cdot \sigma^2(\Delta V_t) + (1 - G_D \cdot R_{SD})^2 \cdot \sigma^2\left(\frac{\Delta \beta}{\beta}\right) + G_D^2 \cdot \sigma^2(\Delta R_{SD}) \quad (3)$$

where $\sigma(\Delta V_t)$, $\sigma(\Delta \beta / \beta)$ and $\sigma(\Delta R_{SD})$ are the standard deviations of the threshold voltage, the gain factor $\beta = W \cdot C_{ox} \cdot \mu_0 \cdot V_D / L$ and the SD series resistance mismatch respectively, g_m is the transconductance and G_D is the channel conductance in the linear region.

Furthermore, a useful approach concerning the mismatch study in linear operation is based on the Y-Function defined as $Y = I_D / \sqrt{g_m}$, which is independent of R_{SD} . Indeed, a Y-Function based methodology has been proposed in order to extract directly the V_t and β local variations, using the mismatch of the Y-Function [3]. According to [4], the Y-Function mismatch can be modeled from weak to strong inversion by,

$$\sigma^2(\Delta Y / Y) = \left(\frac{\beta \cdot \sigma^2(\Delta V_t)}{4 \cdot \beta \cdot n^2 \cdot (k \cdot T / q)^2 + Y^2} \right) + \frac{1}{4} \cdot \sigma^2\left(\frac{\Delta \beta}{\beta}\right) \quad (4)$$

The limitation of all above equations is that they are only valid in the linear operation regime. In the present work, we extended the drain current local variability model described in [4] in order to include the saturation region in the calculation of the drain current mismatch. Using again the sensitivity analysis as in Eq. (2) and making the assumption that $R_S = R_D = R_{SD} / 2$, we can calculate the partial derivative of I_D with respect to R_{SD} as,

$$\frac{1}{I_D} \frac{\partial I_D}{\partial R_{SD}} = g_m / 2 + g_d \quad (5)$$

where g_d is the output conductance.

As a result, combining Eqs. (2) and (5), it is easy to show that the drain current local variability from weak to strong inversion in both linear and saturation regions can be described by,

$$\sigma^2\left(\frac{\Delta I_D}{I_D}\right) = \left(\frac{g_m}{I_D}\right)^2 \cdot \sigma^2(\Delta V_t) + ((1 - (g_m / 2 + g_d) \cdot R_{SD})^2) \cdot \sigma^2\left(\frac{\Delta \beta}{\beta}\right) + (g_m / 2 + g_d)^2 \cdot \sigma^2(\Delta R_{SD}) \quad (6)$$

Note that, as compared to Eq. (3) which is valid in the linear region, Eq. (6) is generalized by including also the saturation region as manifested by the transconductance contribution that may dominate the output conductance term for large drain voltage.

4. Results and discussion

4.1. Experimental results

Fig. 2(a) presents the drain current local variability of n-MOS paired transistors of the nominal geometry, measured in the linear region. Note that the drain current deviation below threshold voltage exceeds 2 decades, making essential the use of the logarithm in Eq. (1). The respective Y-parameter mismatch presented in Fig. 2(b) shows a similar trend.

Moreover, the drain current mismatch at high drain voltage is illustrated in Fig. 2(c) and (d) for a small and a large area device, respectively. As expected, according to Pelgrom's Law, the drain current mismatch is significantly lower in large area devices.

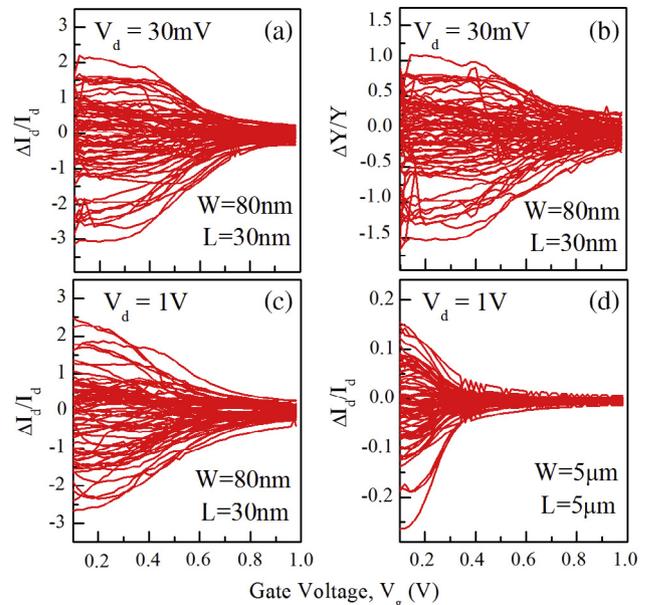


Fig. 2. Typical $\Delta I_D / I_D(V_G)$ (a) and $\Delta Y / Y(V_G)$ (b) curves in linear region for n-MOS devices with $W = 0.08 \mu\text{m}$ and $L = 0.03 \mu\text{m}$. The corresponding $\Delta I_D / I_D(V_G)$ characteristics at saturation region for a small (c) but also for a large area device (d).

In Fig. 3(a) and (b), the normalized standard deviation of the drain current mismatch, $\sigma(\Delta I_D/I_D)$, is plotted as a function of the gate voltage, V_G for various geometries in the linear and saturation regions, respectively. Note that, at low drain voltage, there are clearly cases at strong inversion where an increase of $\sigma(\Delta I_D/I_D)$ with V_G is observed. This increase, which is not observed in all geometries, was attributed to SD series resistance variability in FDSOI devices [4]. Therefore, it appears that this phenomenon also exists in bulk devices. On the other hand, we observe that, for the same geometries, no such behavior is clearly observed in the saturation region (see Fig. 3(b)).

In order to verify our model, we used Eq. (6) to fit the experimental data with 3 fitting parameters, the threshold voltage mismatch, $\sigma(\Delta V_t)$, the current gain factor mismatch, $\sigma(\Delta\beta/\beta)$ and the R_{SD} mismatch, $\sigma(\Delta R_{SD})$, in all operation regions. The value of R_{SD} was extracted with the Y-Function method using several gate lengths [6]. The results are displayed in Fig. 4(a) and (b). Concerning the saturation region, we extracted values for $\sigma(\Delta V_t)$, $\sigma(\Delta\beta/\beta)$ and $\sigma(\Delta R_{SD})$, which are consistent with those extracted in the linear region. As it is shown in Fig. 4(b), we achieved good agreement between experimental and model results. This indicates that the influence of ΔR_{SD} can be significantly attenuated in the saturation region. This feature can be understood through the last term of Eq. (6) which relates the drain current sensitivity with ΔR_{SD} , indicating that I_D is at least twice less sensitive to R_{SD} in the saturation region, where g_d is almost equal to 0 (see Fig. 5). This observation is confirmed in Fig. 6(a) and (b), where the individual matching parameter $iA_{AVg}(V_G)$ (Eq. (7)) is presented for various geometries in the linear and saturation regions, respectively.

$$iA_{AVg} = \left[\sigma\left(\frac{\Delta I_D}{I_D}\right) / \left(\frac{g_m}{I_D}\right) \right] \cdot \sqrt{W \cdot L} \quad (7)$$

The plateau observed in Fig. 6 at low gate voltages nearly corresponds to the individual matching parameter iA_{AVt} (Eq. (8)). Note that the abnormal behavior of $iA_{AVg}(V_G)$ observed at low gate voltages in long channel devices is due to the fact that the devices are pocket implanted [7].

$$iA_{AVt} = \sigma(\Delta V_t) \cdot \sqrt{W \cdot L} \quad (8)$$

As we can see in more detail in Fig. 7, $\sigma(\Delta V_t)$ has almost the same value for both linear and saturation regions. The difference observed between the two regions at high V_G values is due to the $\sigma(\Delta R_{SD})$ difference, while the slight increase of iA_{AVg} in strong inversion is due to $\sigma(\Delta\beta/\beta)$. From Fig. 7, it is also clear that the

parameter iA_{AVg} is smaller at $V_D = 1V$, since the impact of ΔR_{SD} on the drain current variability is less in the saturation region (Eq. (6)).

Figs. 8 and 9 present the individual matching parameters iA_{AVt} and $iA_{\Delta\beta/\beta}$, respectively, as a function of the gate length (Eqs. (8) and (9)).

$$iA_{\Delta\beta/\beta} = \sigma(\Delta\beta/\beta) \cdot \sqrt{W \cdot L} \quad (9)$$

The values corresponding to the iA_{AVt} parameter are ranging between 2 and 6.5 mV· μm , in agreement with [8], increasing slightly with the gate length. Moreover, $iA_{\Delta\beta/\beta}$ ranges from 0.4 to 0.6%· μm , verifying that our fitting with including the gain factor mismatch is correct. Last but not least, the standard deviation of the SD series resistance mismatch versus the channel width is presented in Fig. 10. As can be seen, the R_{SD} and the $\sigma(\Delta R_{SD})$ follow the same trend and more specifically their values decrease as the channel width increases. Furthermore, a dependence on the gate length is observed at fixed width. Finally, it was found that the normalized series resistance local variability, $\sigma(\Delta R_{SD})/R_{SD}$, is of the order of 5–20%, which is similar to FDSOI technologies [4].

4.2. Simulation results

To further verify the findings presented above, we performed numerical simulations of drain current mismatch characteristics in both linear and saturation regions. In order to accurately reproduce the drain current local variability behavior, we used a MOSFET compact model based on Lambert W-function and recalled below [9],

$$I_D(V_G, V_D) = \frac{\frac{W}{L} \int_0^{V_D} \mu_{eff}(V_G, U_c) \cdot Q_i(V_G, U_c) \cdot dU_c}{1 + \frac{1}{L} \int_0^{V_D} \frac{\mu_{eff}(V_G, U_c)}{v_{sat}} \cdot R(V_G, U_c) \cdot dU_c} \quad (10)$$

where the inversion charge Q_i is given by,

$$Q_i(V_G, U_c) = C_{ox} n \frac{kT}{q} LW(e^{q \frac{V_G - V_t - U_c}{nkT}}) \quad (11)$$

with U_c being the quasi Fermi potential along the channel, C_{ox} being the gate oxide capacitance, $k \cdot T/q$ the thermal voltage, n the sub-threshold ideality factor and v_{sat} the saturation carrier velocity. The factor R is related to inversion capacitance and gate oxide capacitance as [9],

$$R(V_G, U_c) = \frac{C_{inv}}{C_{inv} + C_{ox}} \quad (12)$$

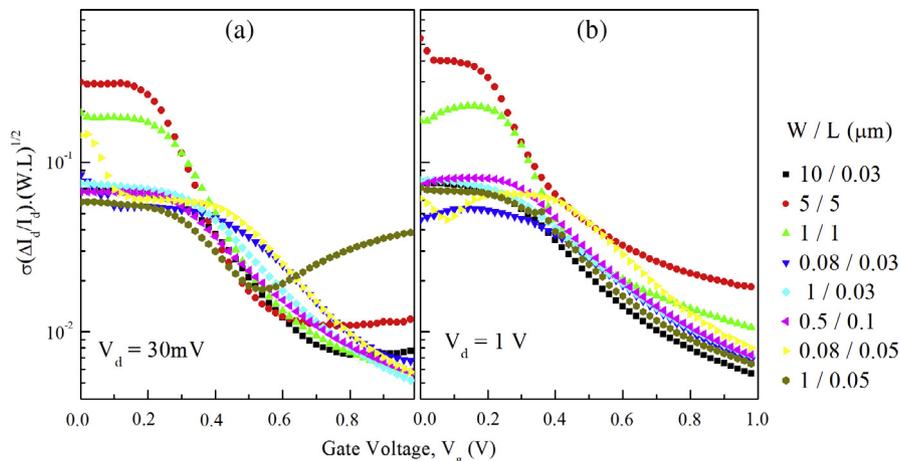


Fig. 3. Normalized standard deviation of the drain current mismatch versus gate voltage, V_G , for different geometries in linear (a) and saturation region (b).

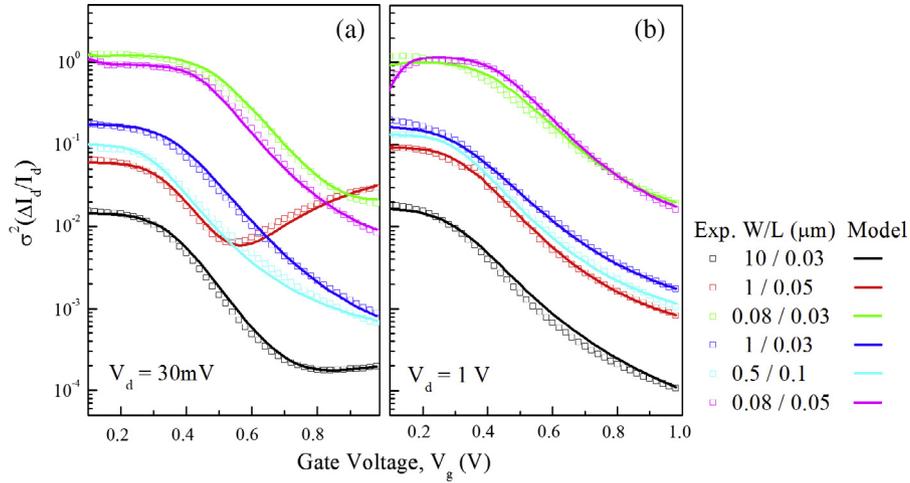


Fig. 4. Experimental results (symbols) of $\sigma^2(\Delta I_D / I_D)$ versus gate voltage, V_G , for small and large area devices in linear (a) and saturation (b) region and Model (lines).

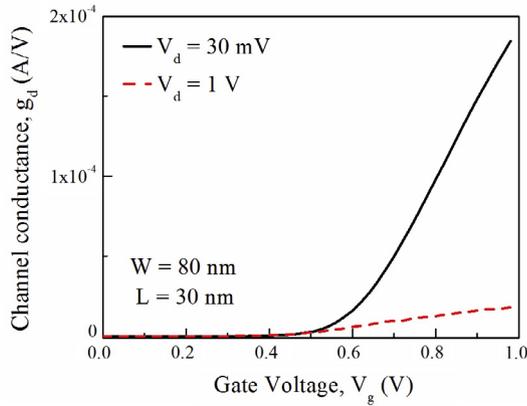


Fig. 5. Channel conductance, g_d as a function of the gate voltage, V_G , in linear (solid line) and saturation region (dashed line) calculated from the nominal device ($W = 0.08 \mu\text{m}$ and $L = 0.03 \mu\text{m}$).

where C_{inv} is the inversion charge capacitance. This factor allows to activate the saturation velocity effect in strong inversion ($R \approx 1$) and to cancel it in weak inversion ($R \approx 0$). The effective mobility μ_{eff} is related as usual to the inversion charge by the first order approximation,

$$\mu_{eff} = \frac{\mu_0}{1 + \theta_1 \cdot (Q_i / C_{ox})}, \quad (13)$$

where μ_0 is the low field mobility and θ_1 is the first order mobility attenuation coefficient.

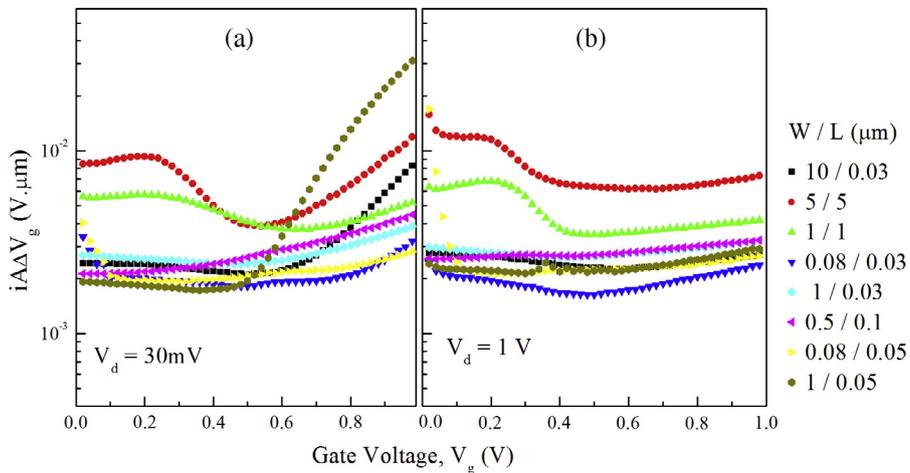


Fig. 6. Individual matching parameter $iA_{\Delta V_g}$ versus gate voltage, V_G , extracted by experimental data in linear (a) and saturation region (b) for different geometries.

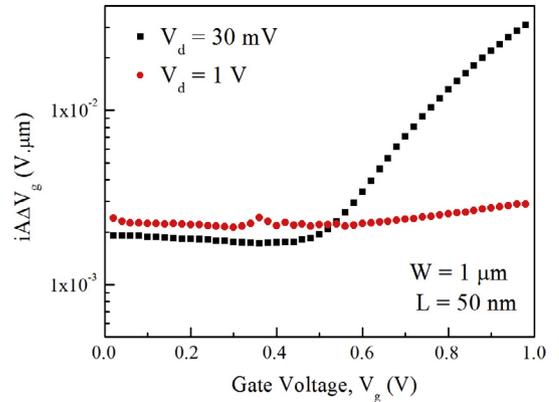


Fig. 7. Individual matching parameter, $iA_{\Delta V_g}$ versus gate voltage, V_G in linear (black symbols) and saturation region (red symbols) for an n-MOS with $W = 1 \mu\text{m}$ and $L = 0.05 \mu\text{m}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

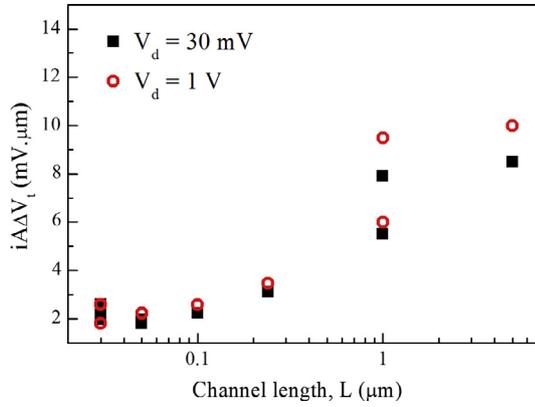


Fig. 8. $i\Delta V_t$ versus channel length, L in linear (black symbols) and saturation region (red symbols) for 28 nm BULK nMOSFETs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

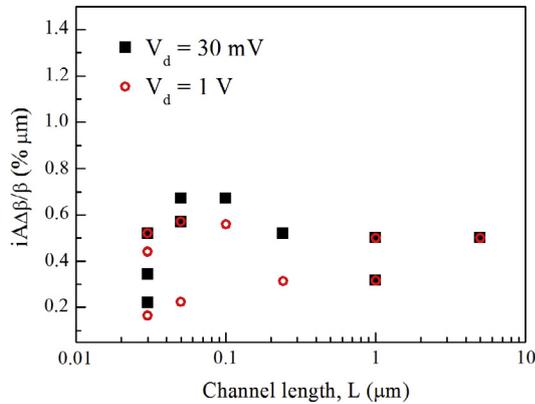


Fig. 9. $i\Delta\beta/\beta$ versus channel length, L in linear (black symbols) and saturation region (red symbols) for 28 nm BULK nMOSFETs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

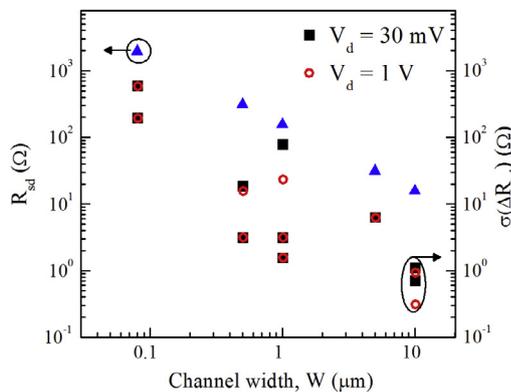


Fig. 10. Source – Drain, SD series resistance (blue symbols) and its mismatch both in linear (black symbols) and saturation region (red symbols) versus channel width, W for 28 nm BULK n-MOSFETs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The effect of the SD series resistance has then been taken into account through the gate and drain voltage drops as $V_G = V_{G0} - (R_{SD}/2) \cdot I_D$ and $V_D = V_{D0} - R_{SD} \cdot I_D$, V_{G0} and V_{D0} being the intrinsic gate and drain to source voltages, respectively.

Then, based on the above MOSFET model, the standard deviation of the drain current variability has been calculated numeri-

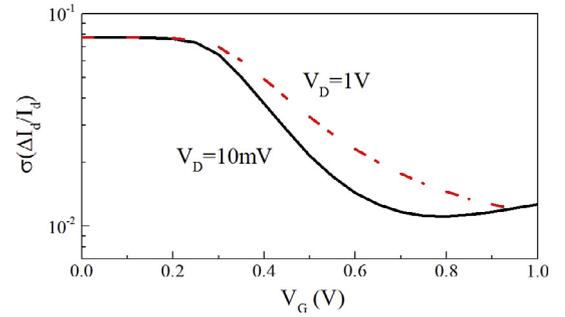


Fig. 11. Simulation results of $\sigma(\Delta I_D/I_D)$ versus gate voltage, V_G for n-MOS devices with channel length $L = 1 \mu\text{m}$ and channel width $W = 1 \mu\text{m}$ for linear (solid line) and saturation (dashed line) regions.

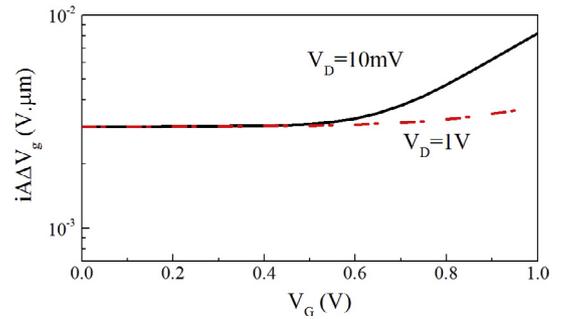


Fig. 12. Simulation results of parameter $i\Delta V_g$ versus gate voltage, V_G for n-MOS with channel length $L = 1 \mu\text{m}$ and channel width $W = 1 \mu\text{m}$ for linear (solid line) and saturation (dashed line) regions.

cally using the sensitivity equation (2) for I_D with respect to V_t , β and R_{SD} for any gate and drain voltage. Fig. 11 shows typical simulated variations of $\sigma(\Delta I_D/I_D)$ with V_G in the linear and saturation regions. It confirms that the ΔR_{SD} mismatch has a smaller impact on the drain current variability in the saturation region, thus verifying the experimental results behavior. This finding is also supported by the simulated individual matching parameter characteristics, $i\Delta V_g(V_G)$, which is also lower in saturation region (see Fig. 12).

5. Conclusions

The impact of the SD series resistance mismatch on the drain current variability has been investigated for 28 nm Bulk MOSFETs. A mismatch model that takes into consideration the R_{SD} local variability was developed and used to extract all mismatch parameters, including $\sigma(\Delta V_t)$, $\sigma(\Delta\beta/\beta)$ and $\sigma(\Delta R_{SD})$, in the linear and saturation regions. It has been demonstrated that the impact of R_{SD} on the drain current variability is reduced in the saturation region due to the lower drain current sensitivity from the series resistance variation. Finally, as in FDSOI devices, the SD series resistance mismatch, $\sigma(\Delta R_{SD})$, was found to scale down with gate width as R_{SD} , and the normalized series resistance local variability parameter $\sigma(\Delta R_{SD})/R_{SD}$ takes similar values as in FDSOI, demonstrating very good access resistance control in bulk technology.

Acknowledgments

This work has been partly supported by the ENIAC places2be and ECSEL Waytogo Fast projects.

References

- [1] Pelgrom MJM, Duinmaijer ACJ, Welbers APG. Matching properties of MOS transistors. *IEEE J Solid-State Circ* 1989;24:1433–9.
- [2] Croon JA, Rosmeulen M, Decoutere S, Sansen W. An easy-to-use mismatch model for the MOS transistor. *IEEE J Solid State Circ* 2002;37:1056–64.
- [3] Rahhal L, Bajolet A, Diouf C, Cros A, Rosa J, Planes N, et al. New methodology for drain current local variability characterization using Y Function method. In: *Proc IEEE Intern Conf Microelectronic Test Structures (ICMTS'13)*. Osaka (Japan). p. 99–103.
- [4] Ioannidis EG, Theodorou CG, Haendler S, Josse E, Dimitriadis CA, Ghibaudo G. Impact of source-drain series resistance on drain current mismatch in advanced fully depleted SOI n-MOSFETs. *IEEE Electron Dev Lett* 2015;36:433–5.
- [5] Ioannidis EG, Haendler S, Josse E, Planes N, Ghibaudo G. Characterization and modeling of drain current local variability in 28 and 14 nm FDSOI nMOSFETs. *Solid State Electron* 2016;118:4–11.
- [6] Ghibaudo G. New method for the extraction of MOSFET parameters. *Electron Lett* 1988;24:543–5.
- [7] Cathignol A, Bordez S, Cros A, Rochereau K, Ghibaudo G. Abnormally high local fluctuations in heavily pocket-implanted bulk long MOSFET. *Solid State Electron* 2009;53:127–33.
- [8] Rahhal L, Bajolet A, Manceau JP, Rosa J, Ricq S, Lassere S, et al. Mismatch trends in 20 nm gate-last bulk CMOS technology. In: *Proc IEEE Intern Conf Ultimate Integration on Silicon (ULIS'14)*. Stockholm (Sweden). p. 133–6.
- [9] Shin M, Shi M, Mouis M, Cros A, Josse E, Mukhopadhyay S, et al. Experimental and theoretical investigation of magnetoresistance from linear regime to saturation in 14 nm FDSOI MOS devices. *IEEE Trans Electron Dev* 2015;62:3–8.



Benchmarks of a III-V TFET technology platform against the 10-nm CMOS FinFET technology node considering basic arithmetic circuits



S. Strangio^{a,*}, P. Palestri^a, M. Lanuzza^b, D. Esseni^a, F. Crupi^b, L. Selmi^a

^a DPIA, Università degli Studi di Udine, Via delle Scienze 206, 33100 Udine, Italy

^b DIMES, Università della Calabria, Via P. Bucci, 41C, I-87036 Arcavacata di Rende (CS), Italy

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

III-V
TFET
Full adders
Ripple carry adders

ABSTRACT

In this work, a benchmark for low-power digital applications of a III-V TFET technology platform against a conventional CMOS FinFET technology node is proposed. The analysis focuses on full-adder circuits, which are commonly identified as representative of the digital logic environment. 28T and 24T topologies, implemented in complementary-logic and transmission-gate logic, respectively, are investigated. Transient simulations are performed with a purpose-built test-bench on each single-bit full adder solution. The extracted delays and energy characteristics are post-processed and translated into figures-of-merit for multi-bit ripple-carry-adders. Trends related to the different full-adder implementations (for the same device technology platform) and to the different technology platforms (for the same full-adder topology) are presented and discussed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The sub-threshold swing (SS) represents the key device parameter to improve the energy efficiency of digital circuits. Among various device proposals for future low power applications, the tunnel field-effect-transistor (TFET) beat the conventional MOSFET and is considered as promising solution to achieve sub-60 mV/dec operation at 300 K in standard CMOS compatible processes [1–13]. Therefore, many efforts are being devoted to the fabrication of TFETs with high performance electrical characteristics [1–5]. In this context, full-quantum simulators [6–8] are widely used as modeling tools to guide the design of such innovative devices, whereas mixed device/circuit simulations are exploited for early analysis at circuit level [9–13]. A virtual III-V TFET technology platform has been recently proposed by Baravelli et al. based on 3D full-quantum simulations [7,8]. An early benchmark against a future CMOS FinFET platform [14,15], based on single device and inverter operation, has been shown [8]. In particular, since the full quantum modeling approach used in [8] does not allow to perform circuit simulations, the inverter operation and the related figures-of-merit (e.g. voltage transfer characteristics - VTC -, V_{OUT}/V_{IN} gain, intrinsic rise and fall times, etc.) have been estimated (i.e. considering the device drain current characteristics and assuming equiv-

alent effective capacitive loads instead of the intrinsic device capacitance characteristics).

The purpose of the present paper is to extend such a benchmark, by considering various 24T and 28T full-adders (FA) blocks as vehicle circuits. Figures-of-merit such as delay and average energy per cycle are extracted and discussed for both TFET and CMOS FinFET implementations. The present paper is an extended version of the work presented at EUROSIOI-ULIS 2016 Conference [13], where only preliminary results on the standard 28T full-adder topology were discussed.

2. Simulation methodology

Fig. 1 sketches the device structures considered in this work, which are: (1) the complementary square cross-section InAs/AlGaSb TFET nanowires proposed in [8]; (2) the 10-nm node CMOS FinFETs described in [14].

Our analysis uses a multi-scale simulation approach, ranging from device simulations to circuit simulations. As regards the TFETs, the TCAD simulator *Sentaurus SDEVICE* [16] has been calibrated to reproduce the full-quantum simulation of the AlGaSb/InAs hetero-structure [7,8]. At the circuit level, the look-up table (LUT) compact models implemented in Verilog-A enabled time-efficient simulations. As regards the CMOS FinFETs, we used the Predictive-Technology-Models of Multi-Gate transistors (PTM-MG) projected to the 10-nm node available at [15].

* Corresponding author.

E-mail address: seb88str@gmail.com (S. Strangio).

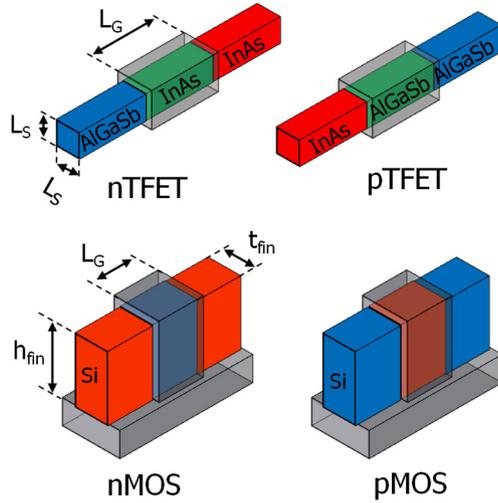


Fig. 1. n- and p-type TFET and FinFET (CMOS) device architectures. The red and blue colors indicate the n- and p-doping types, respectively (green: intrinsic semiconductor, transparent-grey: oxide). TFET dimensions are: $L_G = 20$ nm, nanowire cross section (L_S) = 7 nm, EOT = 1 nm. FinFET dimensions are: $L_G = 14$ nm, $t_{fin} = 9$ nm, $h_{fin} = 21$ nm, EOT = 0.7 nm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. Devices and calibration of TFET models

The III-V TFET technology platform consists of n-type and p-type AlGaSb/InAs TFET nanowires. Due to quantum confinement in the 7×7 nm² square cross-section of the nanowire, the default setup of the TCAD model parameters is not adequate. Thus, the energy-gap (E_C) and the electron affinity (χ) have been chosen so as to reproduce the same band alignment as in [7]. The dynamic nonlocal-path Band-to-Band Tunneling (BtBT) model parameters for the direct tunneling process ($A_{path,dir}$ and $B_{path,dir}$, see [16]) were also recalculated by using the effective masses from bulk GaSb and InAs [17]. Finally, the effective valence and conduction band density of states (N_V and N_C) have been increased compared to the default value for bulk crystals to improve the matching of the I-V curves between TCAD and full-quantum results. The calibrated parameters are summarized in Table 1.

Fig. 2 shows the I_D - V_{GS} transfer-characteristics of p-type and n-type TFET and FinFET devices. For reference, the TFET I_D - V_{GS} obtained from full-quantum simulations [8] are also included (symbols). For the same off-current (I_{OFF}) of 35 pA, the pTFET features a lower on-current (about 1/4) than the nTFET at $V_{DD} = 400$ mV. Despite this asymmetry, we keep a 1/1 ratio between the size of the nTFETs and pTFETs in the following circuit analysis. Differently from TFETs, the I_D - V_{GS} of the n-type and p-type FinFETs (red lines) are essentially symmetric.

Table 1
Calibrated parameters used in the TCAD simulations of the AlGaSb/InAs TFET templates.

Parameter	Al _{0.05} Ga _{0.95} Sb	InAs
<i>Band-gap parameters (including quantization effects)</i>		
Energy gap E_C (eV)	1.04	0.59
Electron affinity χ (eV)	4.01	4.9
<i>Dynamic non-local BtBT model parameters</i>		
A_{path} (cm ⁻³ s ⁻¹)	$1.51 \cdot 10^{20}$	$1.44 \cdot 10^{20}$
B_{path} (V/cm)	$9.54 \cdot 10^6$	$2.94 \cdot 10^6$
<i>Effective conduction and valence band density of states</i>		
N_C (cm ⁻³)	$1.26 \cdot 10^{18}$	$5.22 \cdot 10^{17}$
N_V (cm ⁻³)	$1.8 \cdot 10^{19}$	$6.6 \cdot 10^{18}$

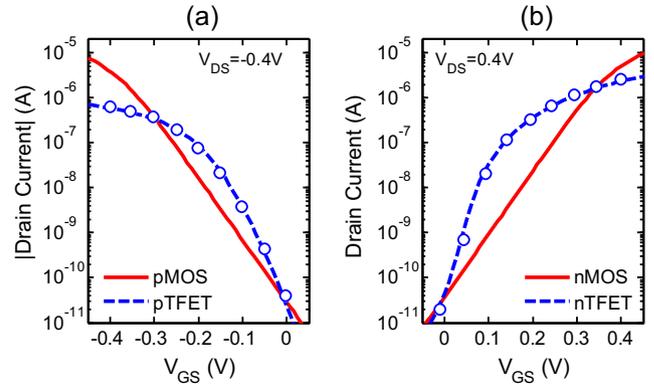


Fig. 2. Transfer-characteristics (I_D - V_{GS}) at $|V_{DS}| = 400$ mV for the transistors simulated in this work. Off-currents are matched at 35 pA at $|V_{DS}| = 400$ mV and $|V_{GS}| = 0$ V.

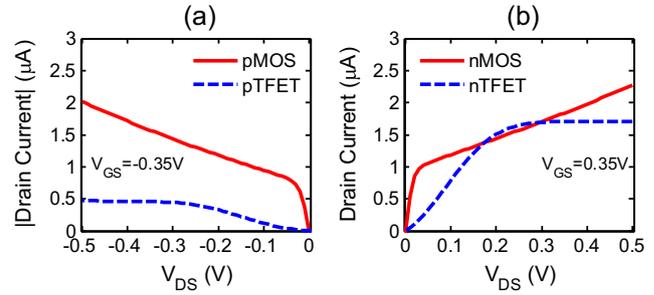


Fig. 3. Output-characteristics (I_D - V_{DS}) at $|V_{GS}| = 350$ mV for the transistors simulated in this work.

Fig. 3 reports the output characteristics I_D - V_{DS} at $V_{GS} = 350$ mV for all the considered devices. The TFETs feature a delayed turn-on behavior (superlinear I_D - V_{DS} [18]), but also a better saturation than FinFETs at $|V_{DS}|$ above 300 mV.

2.2. Circuit simulations

TFETs have uni-directional I_{DS} - V_{DS} characteristics and thus source and drain terminals cannot be exchanged as in conventional MOSFETs. The source is thus marked in the TFET symbol, see Fig. 4a. For circuit simulations we employ a Verilog-A model (sketched in Fig. 4b) implemented within the Cadence environment. LUTs including the drain current I_D , the gate-to-source capacitance C_{gs} and the gate-to-drain capacitance C_{gd} were set up

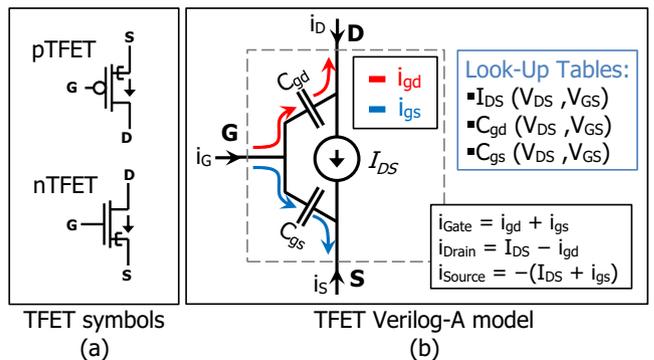


Fig. 4. (a) n-type and p-type TFET symbols definition. (b) Sketch of the Verilog-A model based on the look-up tables with the values of I_D , C_{gs} and C_{gd} as a function of V_{DS} and V_{GS} .

using the TCAD, by simulating the device at bias points in the V_{GS} and V_{DS} range (between -800 mV and 800 mV, with 10 mV voltage step).

As an example of circuit simulation with the proposed approach, we consider a self-loading inverter, that is an inverter loaded by an identical inverter (Fig. 5a). This condition has been already studied in [8], by assuming an equivalent constant load capacitance representative of both the input capacitance C_{in} of the loading inverter and of the Miller capacitance due to the device C_{gd} of the driving inverter itself. Differently from [8], however, here we simulate the inverter by taking into account the actual bias dependence of the device capacitances, i.e. by employing current and capacitance LUTs for all the devices, as sketched in Fig. 5a. Fig. 5b and c show the simulated voltage signals for both the rise and fall transitions of the output voltage, respectively. As explained in [13], since we consider the Miller capacitances as bias-dependent capacitors connected between input and output and not only as an effective load, the simulated output waveforms show a plateau as well as a voltage under/overshoot, which are peculiar behaviors due to the input-to-output capacitive coupling. These features were not included in the simplified estimation of [8], thus resulting in an underestimation of the rise and fall times [13].

3. Benchmark of TFETs vs FinFETs considering full-adders

In this section, the vehicle circuits used for the benchmark are presented along with the benchmarking protocol. Then, the simulation results are presented and discussed.

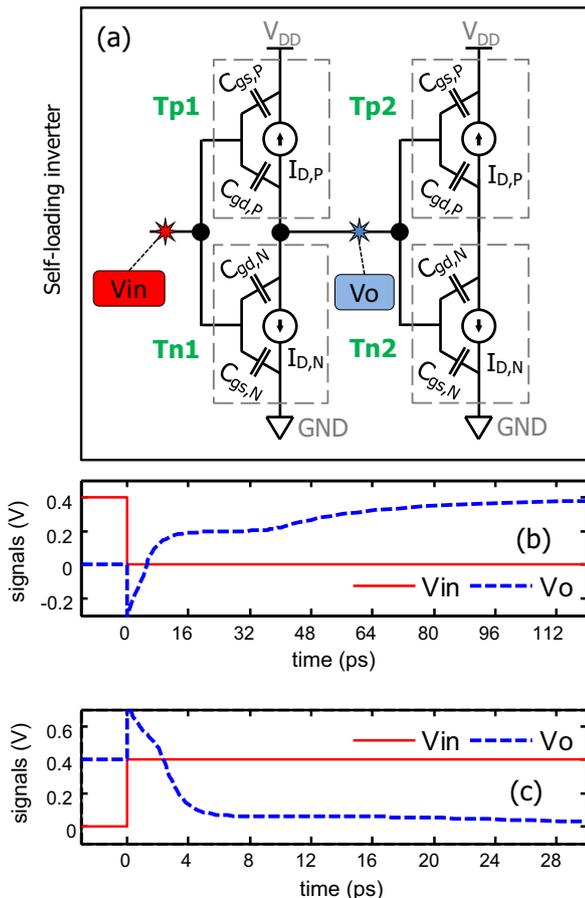


Fig. 5. Schematic of the self-loading inverter simulated in Verilog-A (a). Transient simulations reporting the rise (b) and fall (c) transitions. Rise time = 92 ps; fall time = 25 ps.

3.1. Full-adder topologies

The transistor level designs of the full-adder topologies are reported in Fig. 6 (TFET implementations only). We consider the standard and the mirror implementations with 28 transistors (Fig. 6a and b, respectively), and the transmission-gate implementation with driving capacitance ($24T$ -tgd in the following), that is with inverter stages acting as output buffers, implemented with 24 transistors (Fig. 6c). Concerning the $24T$ -tgd, due to the unidirectional conduction of TFETs, the correct operation of the adder is ensured only if the n- and p-type TFETs implementing the transmission-gates are properly connected. In particular, the circuit schematic is such that the nTFET and the pTFET of each transmission-gate has the source (drain) connected toward the input (output) pin. In fact, although both devices are active during the transmission phase (i.e. when $V_{GS}^{nTFET} = V_{DD}$ and $V_{GS}^{pTFET} = 0$), only the n-type (p-type) device is efficient in transmitting the '0' ('1') logic value. Thanks to this arrangement of the transistors, the nTFET transmits the logic '0' by allowing the current to flow in the direction opposite to the data flow (i.e. from outputs toward inputs), whereas the pTFET permits the transmission of the logic '1' (with the current flowing in the same direction as the data flow). In case of bidirectional transistors (as for FinFETs) there is no need to take care of the device orientation due to symmetric drain/source terminals.

Although other full-adder implementations are possible [19,20], they will not be considered here because they are amenable to operate in the ultra-low voltage regime (e.g. the transmission-gate topology without output buffers) which is the focus of the present paper.

3.2. Test bench and simulation protocol

The block diagram in Fig. 7 has been conceived as a general test-bench for the simulation of full-adders under normal operating conditions. In fact, the block under test (central box) is placed in a framework including realistic driving and loading blocks. The $\overline{A_{wfm}}$, $\overline{B_{wfm}}$ and $\overline{C_{wfm}}$ signals consist of random binary waveforms with a length of 100 bits, where each bit is held for a bit-time (T_{bit}) of 100 ns. This relatively relaxed timing condition has been chosen in order to allow the circuit to operate down to ultra-low V_{DD} values.

For each transition of the sum waveform S_{wfm} , the related delay is computed, taking as a reference the specific input signal that triggers the transition (i.e. the latest signal to switch among A_{wfm} , B_{wfm} and C_{wfm}). The delay of the carry out signal Co is calculated for different cases: propagation of '1' or '0' (P1 and P0), generate carry (G), delete carry (D). In the "propagate" mode (i.e. when $A \neq B$), Co_{wfm} follows C_{wfm} . In the "generate" and "delete" modes (i.e. when $A = B$), we have $Co = A = B$, regardless of the value assumed by C . Thus, for each transition of Co , the delay is computed taking as a reference either C (when the Co transition takes place in the "propagate" mode) or the latest signal to flip between A and B ("generate" or "delete" modes). Although it is possible to define various delay metrics as discussed above, we will here consider the propagation delay ($t_{propagation}$). This corresponds to the worst case delay between $Co(P1)$ and $Co(P0)$ and it is the bottleneck for a multi-bit full-adder implemented as a Ripple-Carry-Adder (RCA) [20], where the theoretical minimum clock period T_{CLK} is given approximately by $N \times t_{propagation}$, with N being the length of the bit words.

The total energy is calculated by integrating the product $V_{DD} \cdot i_{DD,wfm}$ over the simulation time, being $i_{DD,wfm}$ the waveform of the overall current flowing through the full-adder under test. This is then normalized to extract the average energy per bit cycle

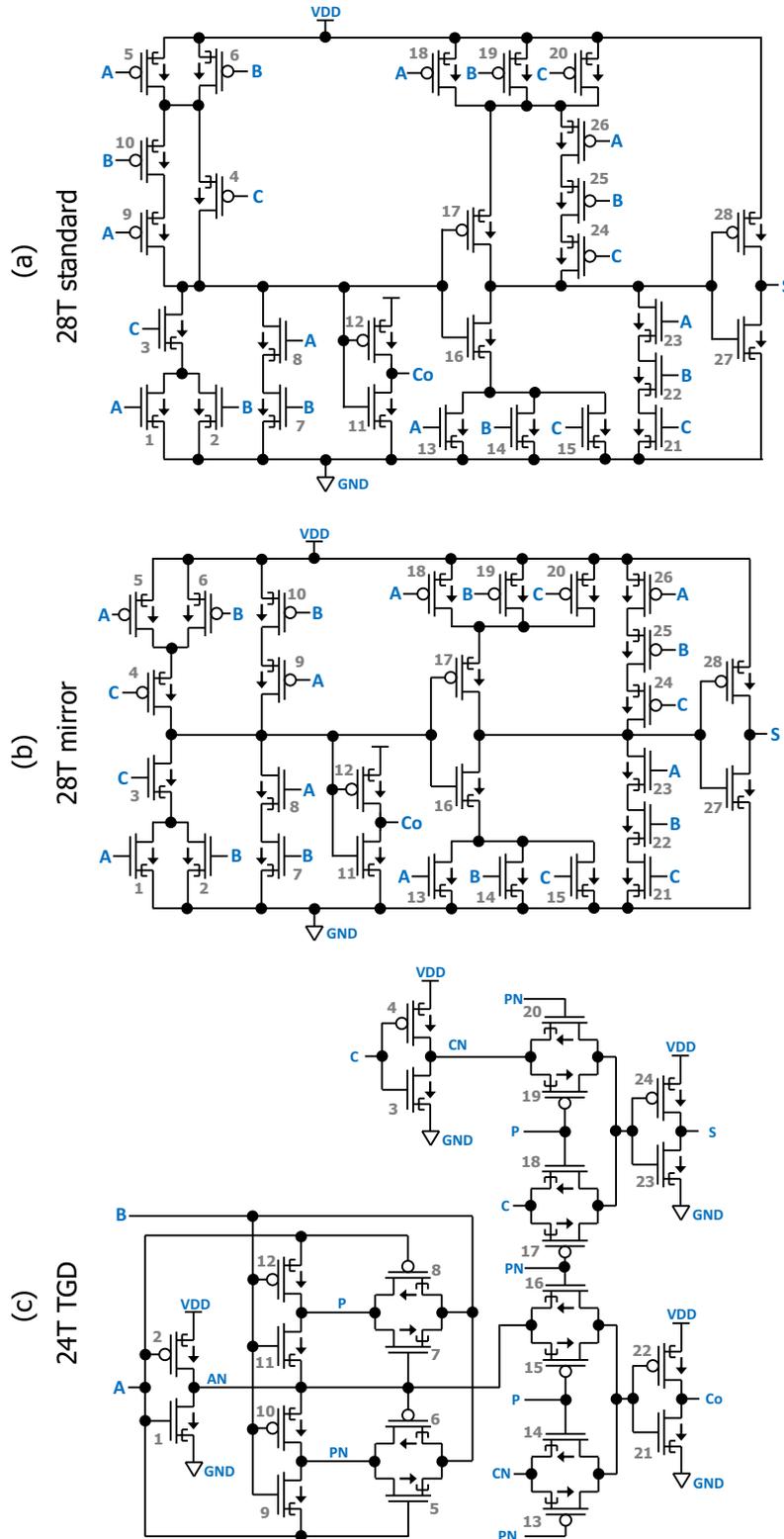


Fig. 6. Schematic of the full-adder implementations considered in this work: (a) 28T standard, (b) 28T mirror, (c) 24T transmission-gate with output buffers (inverters).

(i.e. per $T_{bit} = 100$ ns in this case) as $E_{tot}@[T_{bit}=100ns]$ per bit cycle = $\frac{V_{DD}}{100} \int_0^{T_{Sim}} i_{DD} dt$, where T_{Sim} is the overall simulation time ($T_{Sim} = 100 \cdot T_{bit} = 10$ μ s). At the same time, the steady state current is sampled at the end of each cycle ($I_{DD,DC}^{(i)}$, 'i' being the index of the cycle) in order to calculate the average static power as $P_{Stat,avg} = \frac{V_{DD}}{100} \sum_1^{100} I_{DD,DC}^{(i)}$ and the static energy as $E_{Stat}@[T_{bit}=100 ns]$ per bit cycle = $P_{Stat,avg} \cdot T_{bit}$. Consequently, the average dynamic energy

(E_{Dyn} per bit cycle) can be also extracted, as the difference between the overall energy per cycle and the average static energy per cycle (note that the dynamic energy is essentially independent on T_{bit}).

In order to evaluate the minimum energy per clock cycle needed by a 32-bit RCA for various V_{DD} values, results extracted from such single-bit blocks were post-processed and translated

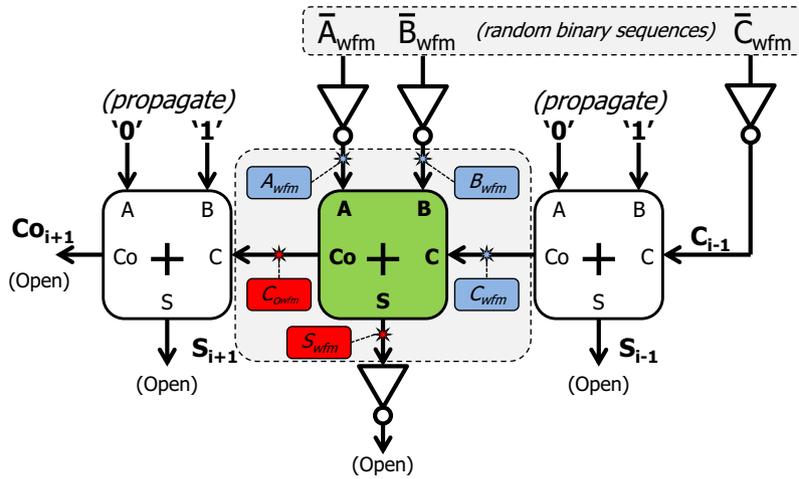


Fig. 7. Simulated test-bench reproducing a normal operating condition for the full-adder under test (central block). Input signals: $\overline{A_{wfm}}$, $\overline{B_{wfm}}$ and $\overline{C_{wfm}}$. The voltage signals A_{wfm} , B_{wfm} , C_{wfm} , S_{wfm} , $C_{o_{wfm}}$, as well as the waveform of the overall current absorbed by the full-adder under test are post processed in order to obtain the figures-of-merit of interest.

into figures-of-merit for multi-bit adders. A clock period T_{CLK} equal to the minimum time that ensures a correct functioning is assumed. T_{CLK} is estimated as $\gamma \cdot N \cdot t_{propagation}$ for each V_{DD} , where γ is a correction factor to accommodate the worst case propagation delay, and N is the number of bits of the input operands (in this study: $\gamma = 2$, $N = 32$). Then, for each V_{DD} , besides weighting the dynamic energy and the static power to the 32 blocks of the 32-bit RCAs ($E_{Dyn}^{32bit-RCA} = 32 \cdot E_{Dyn}^{1bit-FA}$ and $P_{Stat,avg}^{32bit-RCA} = 32 \cdot P_{Stat,avg}^{1bit-FA}$, respectively), the average static power is multiplied by the minimum T_{CLK} (instead of the fixed T_{bit} of 100 ns used for the simulations) to obtain the corresponding static energy ($E_{Stat}^{32bit-RCA} = P_{Stat,avg}^{32bit-RCA} \cdot T_{CLK}$).

4. Results

Fig. 8 reports the propagation delays of all the considered full-adder blocks. For the same technology platform, the 28T mirror implementation turned out to be the fastest, whereas the 24T-tgd is the slowest (see insets in linear scale). However, irrespective of the particular circuit implementation, the TFET-based circuits show less performance degradation when scaling V_{DD} compared to FinFETs; this allows the TFET solutions to become faster than their FinFET counterparts for V_{DD} below ~ 350 mV. Beside the trivial consequence that TFET circuits can operate at a higher clock frequency for such reduced V_{DD} , the smaller performance degradation

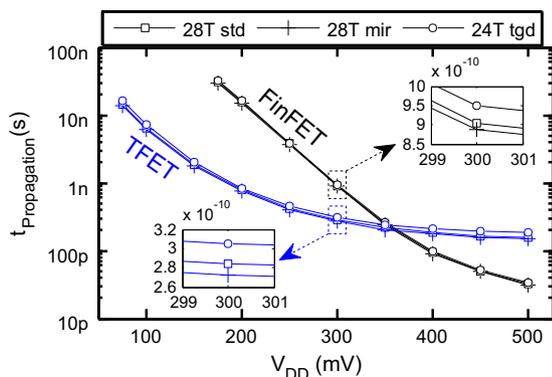


Fig. 8. Propagation delays of the three full-adder topologies, implemented either with TFETs or with FinFETs.

with V_{DD} scaling with respect to FinFETs has also implications from the energy point of view.

In fact, in the energy balance of a circuit, both static and dynamic energy components can be relevant. In particular, at high V_{DD} values the dynamic energy per cycle ($E_{dynamic} = \alpha \cdot C_{L,eq} \cdot V_{DD}^2$, where α is the activity and $C_{L,eq}$ is the equivalent capacitive load of the circuit) is dominant; on the other hand, when the V_{DD} is scaled down, an increasingly relaxed T_{CLK} is needed due to larger delays, thus the static energy component ($E_{Static \text{ per cycle}} = T_{CLK} \cdot V_{DD} \cdot I_{leak}$) becomes dominant.

Fig. 9 reports the estimated static, dynamic and total energy per cycle for 32-bit RCAs (each corresponding to 32 blocks of the topologies in **Fig. 6**, with either TFETs or FinFETs). The static energy in **Fig. 9a** follows qualitatively the same trends as the $t_{propagation}$ in **Fig. 8** (aside from its further dependence on V_{DD}); in fact the increase of T_{CLK} with the V_{DD} scaling is much more pronounced than the linear dependence of E_{Static} on V_{DD} . For both the TFET and FinFET technologies, the 24T-tgd solution features slightly larger static energy consumption than the 28T implementations, despite the reduced number of transistors. A systematic analysis of the leakage paths suggests the following explanation for this finding: for certain input conditions (e.g. $(A,B,C) = ('0','0','0')$), in the 24T-tgd adders there is a larger number of off-state transistors which are biased with $|V_{DS}| = V_{DD}$, whereas in the 28T implementation the $|V_{DS}|$ is smaller because of the voltage partitioning between off-state transistors in series thanks to the staking effect [21].

Looking at the dynamic energy components (**Fig. 9b**), the 24T solutions feature slightly lower active energy consumption per cycle, due to reduced effective load capacitance of the circuit.

The total energy per cycle for all the considered 32-bit RCAs is reported in **Fig. 9c**. For both technology platforms, the RCAs implemented with 32 24T-tgd blocks feature, compared to the RCAs implemented with 32 28T blocks, a slight energy saving at large V_{DD} (i.e. where the dynamic energy dominates) but also a slight energy penalty at reduced V_{DD} (i.e. where the static energy dominates). It should be noticed that, irrespective of the selected circuit topology, the minimum energy point (MEP) for all the FinFET implementations is close to 1000 aJ/cycle for a V_{DD} close to 250 mV, whereas for the TFET implementations it is about 140 aJ/cycle for a smaller V_{DD} close to 150 mV.

These results indicate that, from an energy efficiency point of view, using steep slope devices is more efficient than changing the circuit architecture.

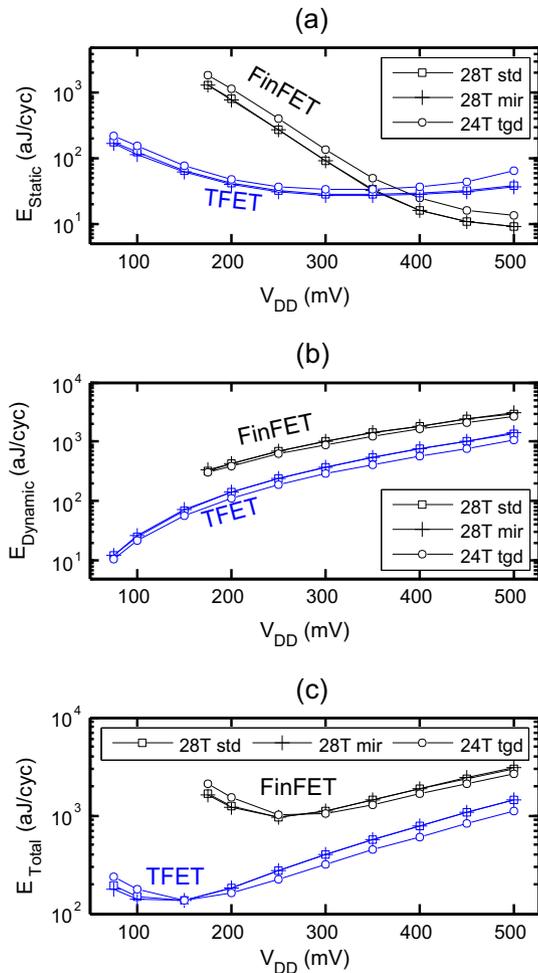


Fig. 9. 32-bit ripple-carry-adder: (a) static energy per cycle, (b) dynamic energy per cycle and (c) total energy per cycle as a function of V_{DD} .

5. Conclusions

A III-V TFET technology platform has been benchmarked against the predictive models for the 10 nm node CMOS FinFETs, considering three topologies of full-adder as test circuits. The propagation delay has been selected as the main performance figure-of-merit for single-bit full-adders. Irrespective of the particular circuit topology, full-adders implemented with TFETs are faster than the corresponding solutions implemented with FinFET technology for V_{DD} below 350 mV. Considering 32-bit ripple carry adders implemented with a chain of 32 single-bit full-adder blocks, the TFET circuits allow energy saving at any V_{DD} in the considered range; in particular, the minimum energy per cycle of each TFET implementation is ~ 140 aJ/cycle (at $V_{DD} = 150$ mV), well below the minimum energy point of ~ 1000 aJ/cycle values for the FinFET implementations (at $V_{DD} = 250$ mV). The difference in performance and energy produced by the circuit topology is practically negligible compared to the differences corresponding to that due to the transistor technology.

Acknowledgment

The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement No. 619509 (project E2SWITCH).

References

- [1] Ionescu A, Riel H. Tunnel field-effect transistors as energyefficient electronic switches. *Nature* 2011;479(7373):329–37.
- [2] Dewey G, Chu-Kung B, Boardman J, Fastenau JM, Kavalieros J, Kotlyar R, Liu WK, Lubyshev D, Metz M, Mukherjee N, Oakey P, Pillarisetty R, Radosavljevic M, Then HW, Chau R. Fabrication, characterization, and physics of III-V heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing". In: IEEE International Electron Devices Meeting (IEDM). <http://dx.doi.org/10.1109/IEDM.2011.6131666>. pp. 33.6.1–33.6.4, 5–7 Dec.
- [3] Mohata DK et al. Demonstration of MOSFET-like on-current performance in arsenide/antimonide tunnel FETs with staggered hetero-junctions for 300 mV logic applications. Washington, DC: IEEE International Electron Devices Meeting (IEDM); 2011. <http://dx.doi.org/10.1109/IEDM.2011.6131665>. pp. 33.5.1–33.5.4.
- [4] Guangle Zhou, Li R, Vasen T, Qi M, Chae S, Lu Y, Zhang Q, Zhu H, Kuo J-M, Kosel T, Wistey M, Fay P, Seabaugh A, Huili Xing. Novel gate-recessed vertical InAs/GaSb TFETs with record high ION of $180 \mu\text{A}/\mu\text{m}$ at $V_{DS} = 0.5$ V. IEEE International Electron Devices Meeting (IEDM); 2012. <http://dx.doi.org/10.1109/IEDM.2012.6479154>. pp. 32.6.1–32.6.4, 10–13 Dec. 2012.
- [5] Luong GV, Strangio S, Tiedemann A, Lenk S, Trellenkamp S, Bourdelle KK, Zhao QT, Mantl S. Experimental demonstration of strained Si nanowire GAA n-TFETs and inverter operation with complementary TFET logic at low supply voltages. *Solid-State Electron* 2016;115(Part B):152–9. January, ISSN 0038–1101.
- [6] Luisier M, Klimeck G. Performance comparisons of tunneling field-effect transistors made of InSb, Carbon, and GaSb-InAs broken gap heterostructures. IEEE International Electron Devices Meeting (IEDM); 2009. p. 1–4. <http://dx.doi.org/10.1109/IEDM.2009.5424280>. 7–9 Dec.
- [7] Baravelli E, Gnani E, Grassi R, Gnudi A, Reggiani S, Baccarani G. Optimization of n- and p-type TFETs Integrated on the Same InAs/AlxGa1-xSb Technology Platform. *IEEE Trans Electron Dev* 2014;61(1):178–85. <http://dx.doi.org/10.1109/TFD.2013.2289739>.
- [8] Baravelli E, Gnani E, Gnudi A, Reggiani S, Baccarani G. TFET inverters with n-/p-devices on the same technology platform for low-voltage/low-power applications. *IEEE Trans Electron Dev* 2014;61(2):473–8. <http://dx.doi.org/10.1109/TFD.2013.2294792>.
- [9] Morris DH, Avci UE, Rios R, Young IA. Design of low voltage tunneling-FET logic circuits considering asymmetric conduction characteristics. *IEEE J Emerg Select Topics Circuits Syst* 2014;4(4):380–8. <http://dx.doi.org/10.1109/JETCAS.2014.2361054>.
- [10] Avci UE, Morris DH, Young IA. Tunnel field-effect transistors: prospects and challenges. *IEEE J Electron Dev Soc* 2015;3(3):88–95. <http://dx.doi.org/10.1109/JEDS.2015.2390591>.
- [11] Strangio S, Palestri P, Esseni D, Selmi L, Crupi F, Richter S, Qing-Tai Zhao, Mantl S. Impact of TFET unidirectionality and ambipolarity on the performance of 6T SRAM cells. *IEEE J Electron Dev Soc* 2015;3(3):223–32. <http://dx.doi.org/10.1109/JEDS.2015.2392793>. May.
- [12] Lanuzza M, Strangio S, Crupi F, Palestri P, Esseni D. Mixed tunnel-FET/MOSFET level shifters: a new proposal to extend the tunnel-FET application domain. *IEEE Trans Electron Dev* 2015;62(12):3973–9. <http://dx.doi.org/10.1109/TFD.2015.2494845>.
- [13] Strangio S, Palestri P, Lanuzza M, Esseni D, Crupi F, Selmi L. Benchmarks of a III-V TFET technology platform against the 10-nm CMOS technology node considering 28T Full-Adders. In: 2016 Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS). Vienna, Austria: Wien; 2016. p. 139–42. <http://dx.doi.org/10.1109/ULIS.2016.7440072>.
- [14] Sinha S, Yeric G, Chandra V, Cline B, Yu Cao. Exploring sub-20 nm FinFET design with predictive technology models. In: 2012 49th ACM/EDAC/IEEE Design Automation Conference (DAC). p. 283–8. <http://dx.doi.org/10.1145/2228360.2228414>. 3–7 June 2012.
- [15] URL: <http://ptm.asu.edu>.
- [16] TCAD Sentaurus Device U.G., version G-2012.06; 2012.
- [17] URL: <http://www.ioffe.ru/SVA/NSM/Semicond/>.
- [18] De Michielis L, Lattanzio L, Ionescu AM. Understanding the superlinear onset of tunnel-FET output characteristic. *IEEE Electron Dev Lett* 2012;33(11):1523–5.
- [19] Weste Neil, Harris David. *CMOS VLSI Design: a circuits and systems perspective*. 4th ed. USA: Addison-Wesley Publishing Company; 2010.
- [20] Alioto M, Palumbo G. Analysis and comparison on full adder block in submicron technology. *IEEE Trans Very Large Scale Integr VLSI Syst* 2002;10(6):806–23. <http://dx.doi.org/10.1109/TVLSI.2002.808446>.
- [21] Magnone P, Crupi F, Alioto M, Kaczer B, De Jaeger B. Understanding the potential and the limits of germanium pMOSFETs for VLSI circuits from experimental measurements. *IEEE Trans Very Large Scale Integr VLSI Syst* 2011;19(9):1569–82.



Study of line-TFET analog performance comparing with other TFET and MOSFET architectures



Paula Ghedini Der Agopian^{a,b,*}, João Antonio Martino^a, Anne Vandooren^c, Rita Rooyackers^c, Eddy Simoen^c, Aaron Thean^c, Cor Claeys^{c,d}

^a LSI/PSI/USP – University of São Paulo, Av. Prof. Luciano Gualberto, trav. 3 n° 158, 05508-010 Sao Paulo, Brazil

^b UNESP – Universidade Estadual Paulista, São João da Boa Vista, Brazil

^c imec, Kapeldreef 75, B-3001 Leuven, Belgium

^d E.E. Dept, KULeuven, Leuven, Belgium

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Line-TFET

Intrinsic voltage gain

Different device architectures

ABSTRACT

In this work the Line-TFET performance is compared with MOSFET and Point-TFET devices, with different architectures (FinFET and GAA:Gate-All-Around) at both room and high temperatures. This analysis is based on the experimental basic analog parameters such as transconductance (g_m), output conductance (g_D) and intrinsic voltage gain (A_V). Although the Line-TFETs present worse A_V than the point-TFETs, when they are compared with MOSFET technology, the line-TFET shows a much better intrinsic voltage gain than both MOSFET architectures (FinFET and GAA). Besides the A_V , the highest on-state current was obtained for Line-TFETs when compared with other two TFET architectures, which leads to a good compromise for analog application.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Tunnel-FETs have been proposed by the international community as an alternative for MOSFETs, when focusing on extremely small technology nodes, due to their high speed switching capability [1] that allows to improve the energy efficiency of switches. However, as the low on-current is a problem of homo-junction TFETs, several works report research on different materials and different geometries aiming to reach a smaller subthreshold swing and an on-current improvement [2,3].

A planar Line-TFET is an alternative structure that increases the electric field at the source-pocket junction and consequently improves the on-current and reduces the subthreshold swing when compared with the point-TFET silicon devices [4–6].

Although the main focus of the tunnel-FET is the digital switch, some recent work has also pointed out the great potential of these devices for analog applications [7–10].

In this work the planar heterojunction Line-nTFET is experimentally analyzed through the basic analog parameters, focusing mainly on intrinsic voltage gain. A comparison of the intrinsic voltage gain of Line-TFETs with devices with different architectures

like FinFET (MOSFET and TFETs) [7] and Gate-all-around (MOSFET and TFET) [11] is also performed, for temperatures ranging from room up to 150 °C. For vertical GAA-TFETs, different source compositions (Si and $\text{Si}_{0.73}\text{Ge}_{0.27}$) will be also considered.

2. Device characteristics

The studied Line-nTFETs are Si/SiGe heterojunction devices fabricated on silicon-on-insulator wafers at imec/Belgium. The p-type $\text{Si}_{0.55}\text{Ge}_{0.45}$ source extends under the gate and a thin intrinsic silicon pocket layer (~5 nm) is on top. The source and the drain regions are separated by a nominally undoped Si channel.

The gate stack is composed by a 1 nm interfacial SiO_2 layer followed by 1.8 nm of HfO_2 , 2 nm of TiN and p-doped amorphous silicon. The channel width (W) ranges from 110 nm to 200 nm and two different gate lengths (L) (1 μm and 130 nm) were evaluated.

Fig. 1 presents a schematic structure of a Line-TFET and more details on this structure/fabrication can be found in [6].

3. Analysis and discussion

Since the source of Line-TFETs extends under the gate region, this architecture promotes tunneling in the same electric field direction, which is more efficient than the conventional point-tunneling. Besides, the position of the tunneling source/channel

* Corresponding author at: LSI/PSI/USP – University of São Paulo, Av. Prof. Luciano Gualberto, trav. 3 n° 158, 05508-010 Sao Paulo, Brazil.

E-mail address: agopian@lsi.usp.br (P.G.D. Agopian).

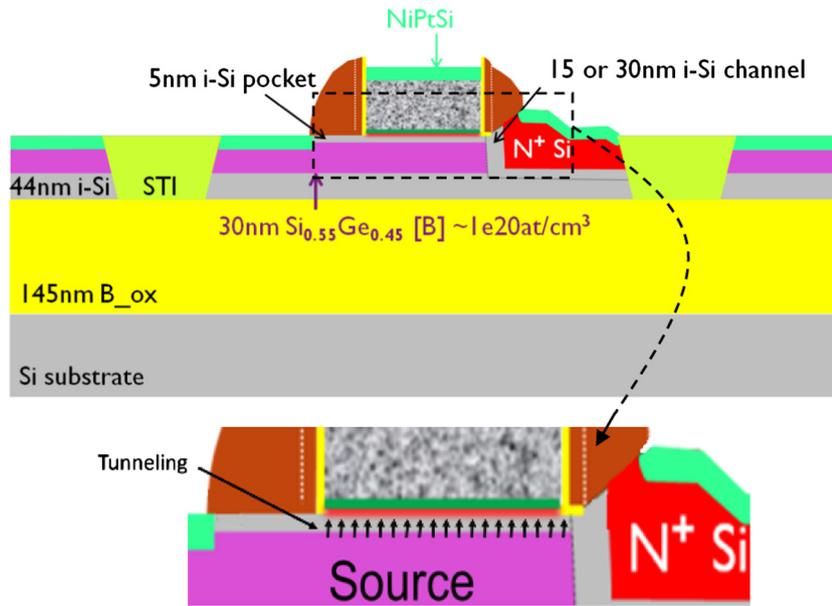


Fig. 1. Line-TFET structure.

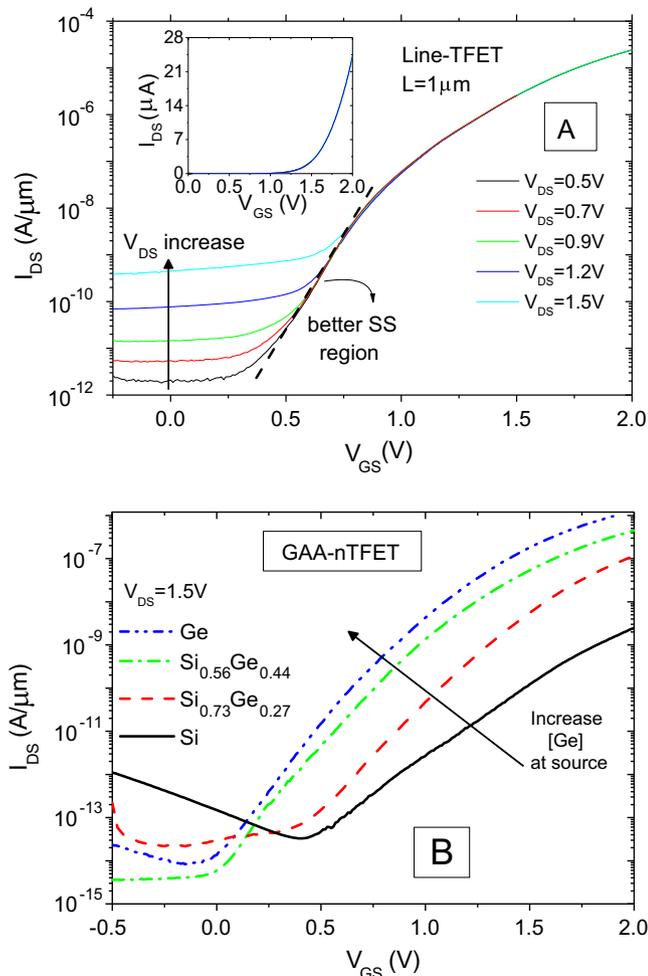


Fig. 2. Experimental normalized transfer curves of a Line-TFET for different drain bias (A) and of a point-TFET for different source compositions [12] (B).

junction makes the total tunneling proportional to the $L \times W$ dimensions.

Fig. 2 presents the transfer characteristic normalized by the channel width of a single Line-TFET for different drain bias (A) and for point-TFETs with different source composition (B) as a function of gate voltage. From Fig. 2, it is possible to observe that Line-TFETs (2A) reaches a higher on-state current than point-TFETs (2B), considering the same channel width. This ON-current improvement, promoted by the Line-TFET structure, is a result of the alignment of the electric field with the tunneling direction that in turns, results in a strong energy band bending, increasing the band-to-band current. It is also possible to observe from the Line-TFET transfer characteristics (Fig. 2A) that the smaller the drain bias (V_{DS}), the steeper the drain current (I_{DS}) in the sub-threshold region due to the off current reduction. The SS improvement becomes even more pronounced with increasing gate length and consequently the tunneling area as reported in [6].

Besides the high on state current, when the output characteristic is evaluated (Fig. 3), the line-TFET also presents a good plateau in the saturation like region, showing to be a promising device for

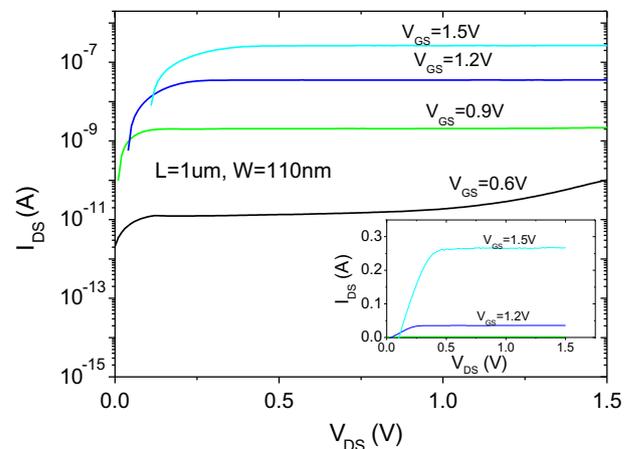


Fig. 3. Experimental output characteristic of a Line-TFET for different gate bias.

analog applications, as was already reported for point-TFETs [7,8]. However, at low gate bias, it is clear that the saturation like region does not present a plateau, i.e., the output characteristic is degraded and becomes inappropriate for this kind of application. Since the drain current level changes several orders of magnitude with the gate bias increase, the drain current was also plotted in a linear scale (inset), in order to better observe the plateau region, confirming the aforementioned.

Considering that the intrinsic voltage gain (A_V) is one of the most important figures of merit for analog applications and it can be calculated by the transconductance (g_m) over output conductance (g_D) ratio, these parameters were evaluated for different bias, channel lengths and channel widths, aiming to optimize the Line-TFET A_V performance.

The g_m and g_D analysis were performed for different channel lengths and different channel widths in order to select the best combination of the transistor dimensions to optimize the A_V value as can be seen in Fig. 4.

From Fig. 4A, it is possible to notice that although g_m increases with the gate bias due to the higher overlap between bands, similarly as occurs for point-TFETs, when the comparison between line and point TFETs focus on the g_m dependence with channel length, the line-TFETs presents a direct dependence on L , while point devices usually are independent on it. The higher g_m for longer channel length occurs due to the larger tunneling junction area underneath the gate as shown in Fig. 4 (source/Si pocket).

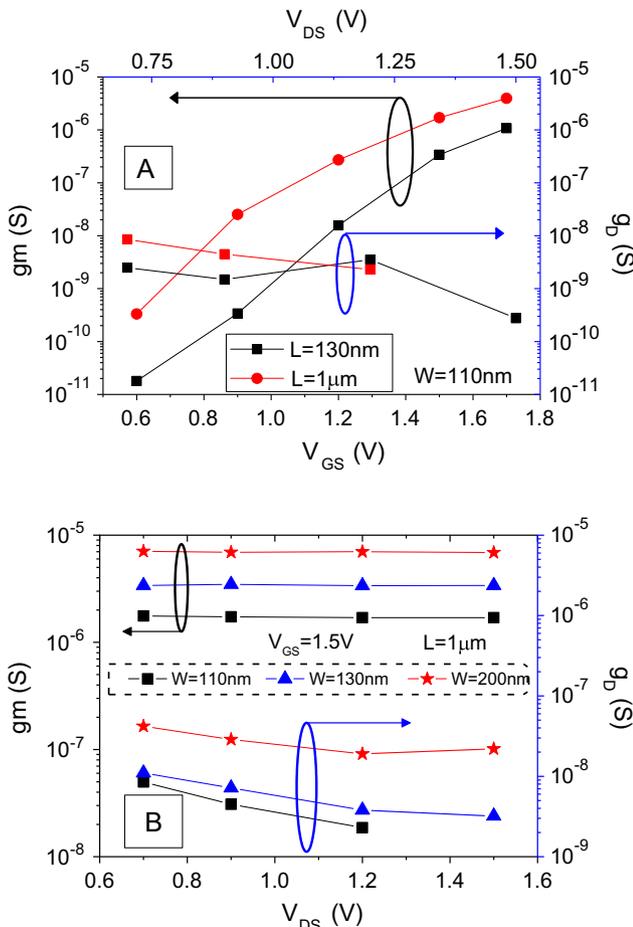


Fig. 4. Transconductance (Left/Bottom axis) and output conductance (Right/Top axis) for different channel lengths (A) and for different channel widths as a function of V_{DS} (B).

However, evaluating the g_D values, a smaller gate length dependence is observed. Besides this, as V_{DS} increases, the TFET devices operate more in the “saturation like” region, resulting in a better (smaller) g_D .

The dependence of these two parameters (g_D and g_m) on channel width (W) were also evaluated for high V_{GS} (1.5 V) as a function of drain bias (Fig. 4B). Although the drain bias does almost not affect the transconductance, it increases with the channel width, as expected, due to the junction area increase. Focusing on output conductance (g_D), it was observed that it depends on both the drain bias and the channel width, but in the opposite way. While the higher drain bias contributes to the g_D improvement (the TFET operates more in the “saturation like” region), the drain current increases with channel width resulting in a g_D degradation (increase).

Keeping in mind that the transconductance presented a significant improvement for longer channel device, while the variation on output conductance is not so important, the obtained intrinsic voltage gain (A_V) for longer devices was higher than the shorter one. Fig. 5 shows the experimental A_V for long line-TFETs ($L=1\mu\text{m}$) with different channel widths operating at different drain bias. From Fig. 5 it is possible to observe that the output characteristic improvement associated with a V_{DS} increase leads to an optimization of the bias operation point for the analog performance of all Line-TFETs. However, when the channel width (W) was evaluated, the response of a transistor with W of 130 nm, shows a reduction of two times on transconductance and a strong reduction on g_D compared with their counterpart with $W=200\text{nm}$, that in turns, results in a best A_V value for devices with $W=130\text{nm}$ and $L=1\mu\text{m}$.

Since the line-TFET with $W=130\text{nm}$ and $L=1\mu\text{m}$ shows to be slightly better and the best bias condition was defined above for the line-TFET architecture and it was already performed for TFET-FinFET and GAA-FinFET architectures in [7,8], respectively, from now on, their analog performance is compared among these three different architectures. Besides this analysis, a comparison of this planar Line-TFET was also performed for different technologies (MOS and TFET).

The first comparison, shown in Fig. 6, is focused on g_m and g_D of these three different architectures. Comparing the TFETs fabricated in a planar (line) and FinFET structure (Fig. 6A) it is possible to see that both transconductance and output conductance obtained for planar devices are very high because the line TFETs becomes to be dominated by the earlier start of the Band-to-band tunneling phenomena. However, considering the vertical GAA structure

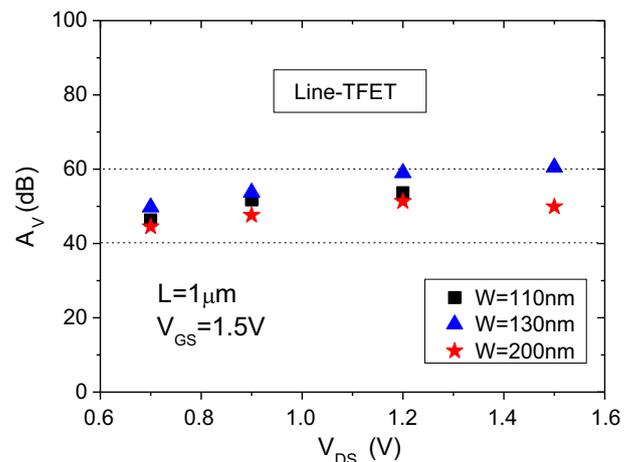


Fig. 5. Experimental A_V for Line-TFETs for different V_{DS} and channel widths at $V_{GS}=1.5\text{V}$.

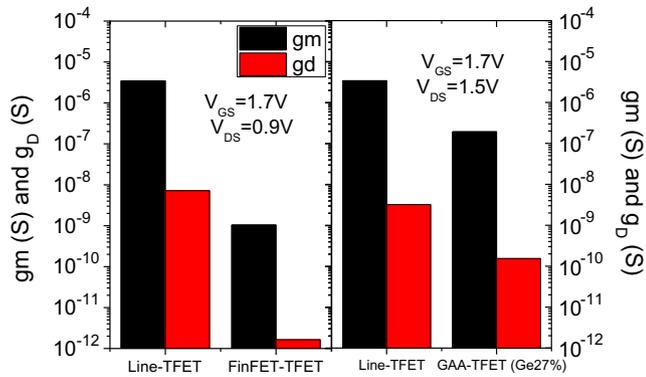


Fig. 6. Experimental comparison of g_m and g_D at optimum point for different architectures: Line-TFET \times FinFET (A) and Line-TFET \times GAA-TFET (B).

(Fig. 6B), only the transconductance stands out due to its improvement.

Making the analog evaluation for different temperatures (from room to 150 °C), the A_V performance among the planar Line-TFET and transistors fabricated with the FinFET structure (tunnel-FET and MOSFET), shows a smaller A_V for Line-TFETs than for the TFET with the FinFET structure (Fig. 7). It occurs because line-TFETs are more dependent of BTBT and as a consequence a higher g_D was obtained.

However, when line-TFETs are compared with the conventional FinFETs (MOSFET technology), the A_V of Line-TFETs is at least 30 dB higher for all temperatures because the BTBT tunneling current is less dependent of V_{DS} than the drift current.

Considering the gate-all around transistors, like nanowire devices, it is well known that for smaller diameter, the coupling between gate and channel is higher, resulting in a predominance of BTBT rate along all the source/channel junction area [13]. As a consequence it can be obtained a better subthreshold swing behavior, higher g_m and consequently higher transistor efficiency (g_m/I_{DS}) at weak conduction regime as reported by [14], which increases the intrinsic voltage gain increases only in this operation region. However, transistors with higher diameter is more TAT dependent (less BTBT dependence), which degrades the weak inversion, but it is less drain voltage dependent, resulting in a better output conductance, Early voltage and consequently the intrinsic voltage gain at strong conduction as already reported in [7,8].

Focusing on the performance of intrinsic voltage gain in the strong conduction regime, the same analog comparison was per-

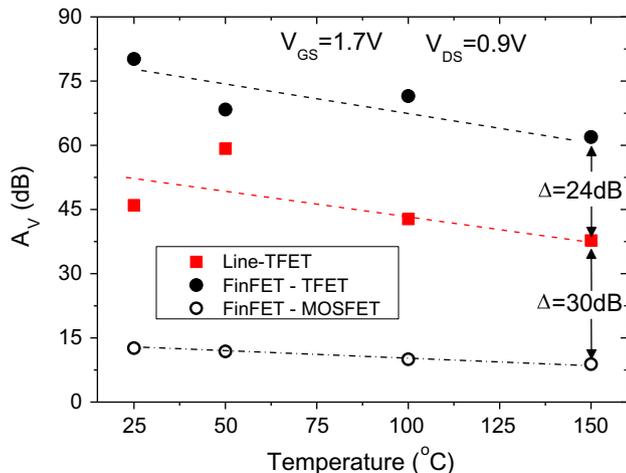


Fig. 7. Experimental comparison of A_V as a function of temperature among Line-TFET, FinFET-TFET and FinFET MOSFET, for FinFET-TFET optimal bias point ($V_{GS} = 1.7$ V and $V_{DS} = 0.9$ V).

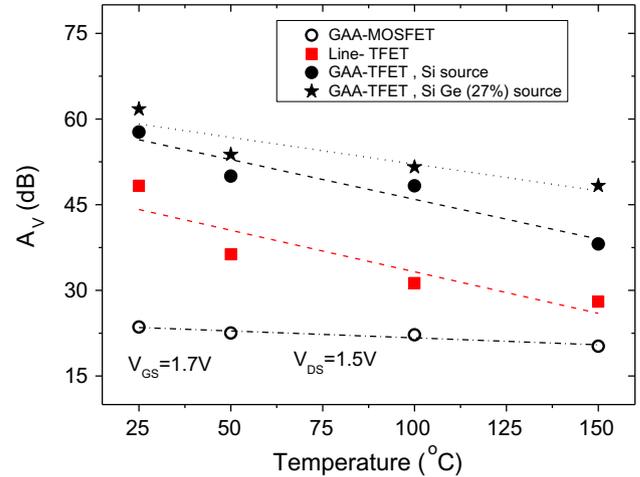


Fig. 8. Experimental comparison of A_V as a function of temperature among Line-TFET, GAA-TFET (with different source compositions) and GAA-MOSFET, for GAA-TFET optimal bias point ($V_{GS} = 1.7$ V and $V_{DS} = 1.5$ V).

formed, but now considering Line-TFETs and vertical GAA structures (Si-MOSFETs and TFETs with Si and $Si_{0.73}Ge_{0.27}$ sources) as can be seen in Fig. 8. Since the line-TFETs reaches a higher BTBT current while GAA structures are more TAT dependent, the GAA-TFETs present higher A_V values than the Line TFETs, independent on the source composition.

However when the Line-TFET is compared with a GAA MOSFET, the Line TFET seems to be better again.

Although Line-TFETs do not reach A_V values as high as for GAA-TFETs and Fin-TFETs, when a high on-state current is required, the planar Line-TFET can be considered as an alternative, since it reaches on-state currents much higher than the other TFET structures studied in this paper.

4. Conclusion

This paper presents an analysis of the intrinsic voltage gain of Line-TFETs and makes a comparison with devices fabricated with vertical GAA and FinFET structures for both TFET and MOSFET technologies. The results show that this planar line-TFET architecture does not present the highest A_V values when compared with the two other vertical TFET architectures (FinFET-TFET and GAA-TFET), however it reaches a very high on-state current, which till now was a road block for another TFET structures. Therefore it is possible to conclude that Line-TFETs can be a good alternative to replace MOSFETs since it reaches the highest on-state currents and a better intrinsic voltage gain than the advanced MOSFET architectures (at least 30 dB higher than FinFETs and 10 dB higher when compared with GAA-MOSFETs).

Acknowledgments

The authors would like to thank CNPq and FAPESP for the financial support during the execution of this work. Part of the work has been performed within the frame of imec's Core Partner program on Logic Devices.

References

- [1] Ionescu AM, Riel H. Tunnel field-effect transistors as energy-efficient electronics switches. *Nature* 2011;479:329–37.
- [2] Verhulst A et al. Complementary silicon-based heterostructure tunnel-FETs with high tunnel rates. *IEEE Electron Dev Lett* 2008;29(12):1398–401.
- [3] Nah J et al. Ge-Si_xGe_{1-x} core-shell nanowire tunneling field-effect transistors. *IEEE Trans Electron Dev* 2010;57(8):1883–8.

- [4] Verreck D et al. Quantum mechanical performance predictions of p-n-i-n versus pocketed line tunnel field-effect transistors. *IEEE Trans Electron Dev* 2013;60(7):2128–34.
- [5] Schmidt M et al. Line and point tunneling in scaled Si/SiGe heterostructure TFETs. *IEEE Electron Dev Lett* 2014;35(7):699–701.
- [6] Walke AM et al. Fabrication and analysis of a Si/Si_{0.55}Ge_{0.45} heterojunction line tunnel FET. *IEEE Trans Electron Dev* 2014;61(3):707–15.
- [7] Agopian PGD et al. Experimental comparison between trigate p-TFET and p-FinFET analog performance as a function of temperature. *IEEE Trans Electron Dev* 2013;60:2493–7.
- [8] Agopian PGD et al. Influence of the source composition on the analog performance parameters of vertical nanowire-TFETs. *IEEE Trans Electron Dev* 2015;62:16–22.
- [9] Zhao Q-T et al. Strained Si and SiGe nanowire tunnel FETs for logic and analog applications. *J Electron Dev Soc* 2015;3(3):103–14.
- [10] Sedighi B et al. *IEEE Trans Circ Syst-I: Regular Pap* 2015;62(1).
- [11] Agopian PGD et al. Comparison between vertical silicon NW-TFET and NW-MOSFET from analog point of view. *EuroSOI-ULIS* 2015:233.
- [12] Martino JA et al. The impact of the Ge concentration in the source for vertical tunnel-FETs. *ECS Trans* 2015;66(4):79–86. doi: <http://dx.doi.org/10.1149/06604.0079ecst>.
- [13] Sivieri VB, Agopian PGD, Martino JA. Impact of diameter on TFET conduction mechanisms. In: 30th Symposium on microelectronics technology and devices (SBMicro), p. 1–4. doi: <http://dx.doi.org/10.1109/SBMicro.2015.7298146>.
- [14] Schulte-Braucks C et al. Experimental demonstration of improved analog device performance in GAA-NW-TFETs. In: 44th European Solid State Device Research Conference (ESSDERC), p. 178–81. 10.1109/ESSDERC.2014.6948789.



Confinement orientation effects in S/D tunneling



C. Medina-Bailon*, C. Sampedro, F. Gámiz, A. Godoy, L. Donetti

Nanoelectronics Research Group, Departamento de Electrónica, Universidad de Granada, Campus Fuentenueva S/N, 18071 Granada, Spain
 CITIC Universidad de Granada, 18071 Granada, Spain

ARTICLE INFO

Article history:

Available online 20 October 2016

The review of this paper was arranged by Viktor Sverdlov

ABSTRACT

The most extensive research of scaled electronic devices involves the inclusion of quantum effects in the transport direction as transistor dimensions approach nanometer scales. Moreover, it is necessary to study how these mechanisms affect different transistor architectures to determine which one can be the best candidate to implement future nodes. This work implements Source-to-Drain Tunneling mechanism (S/D tunneling) in a Multi-Subband Ensemble Monte Carlo (MS-EMC) simulator showing the modification in the distribution of the electrons in the subbands, and, consequently, in the potential profile due to different confinement direction between DGSOIs and FinFETs.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The study of alternative technical approaches for electronic devices is necessary to fulfill the requirements of power consumption, delay time and scalability demanded by ITRS [1]. Currently, there are two main work trends on the simulation of semiconductor processes and devices. The first one is the study of quantum effects in the nanometric dimensions of the conventional devices. The second one is mainly focused on novel engineering solutions to create improved device architectures.

The inclusion of quantum effects in the transport direction is mandatory when the dimensions of the electronic devices are reduced. In particular, Source-to-Drain tunneling (S/D tunneling) allows electrons to go through the potential barrier instead of rebound from it. When this quantum effect is taken into account, the height of the potential barrier is modified increasing the sub-threshold current. Moreover, this phenomenon introduces noise because the number of affected electrons has a random nature. A ballistic non-equilibrium Green's Function (NEGF) approach has demonstrated that S/D tunneling is a scaling limit due to the reasons mentioned above [2]. In addition, it will distort the MOSFET operation at transistor channel lengths around 3 nm [3]. This phenomenon is of special interest when the operation regime is near-threshold because the leakage current increases and V_{th} decreases [4].

At the same time, different technological architectures are proposed to overcome the limitations of conventional planar devices

[5,6]. For this reason, new transistor architectures based on multiple gates [7] are replacing standard technology as a way to keep short channel effects (SCEs) under control. Furthermore, the increased electrostatic confinement provided by multiple gates relaxes the manufacturing constraints in comparison to conventional planar devices. For example, a channel thickness (T_{Si}) is required to be one fourth of the channel length to guarantee acceptable short-channel effects in SOI technology. However, extremely thinner T_{Si} can represent a critical parameter in the fabrication of electronic devices as they are scaling down. This critical T_{Si} of a double gate transistor is approximately twice as wide as T_{Si} of a single-gate device with the same short-channel properties. It therefore alleviates the fabrication problem. If we consider a double gate device, these gates can be oriented horizontally, Double-Gate Silicon-On-Insulator (DGSOIs), or vertically, FinFETs. Ideally, both channels are activated simultaneously and feature identical characteristics. The gates are parallel to the standard wafer orientation for DGSOIs whereas they are perpendicular in FinFETs as depicted in Fig. 1. It should be highlighted that the FinFET is a 3D structure whereas our MS-EMC simulator makes use of a 2D description. However, it was demonstrated that FinFETs with a big enough aspect ratio show similar behavior in all transport regimes when 2DMS-EMC, which consider infinite fin height, and other 3D codes are used [8].

This work presents a meticulous comparison between DGSOIs and FinFETs by means of a Multi-Subband Ensemble Monte Carlo (MS-EMC) simulator when S/D tunneling mechanism is taken into account. It will be shown the influence of the orientation on the S/D tunneling and, consequently, on the device characteristics.

The outline of this work is as follows. Section 2 gives an overview of the code developed to carry out our research, where the

* Corresponding author at: Nanoelectronics Research Group, Departamento de Electrónica, Universidad de Granada, Campus Fuentenueva S/N, 18071 Granada, Spain.

E-mail address: cmedba@ugr.es (C. Medina-Bailon).

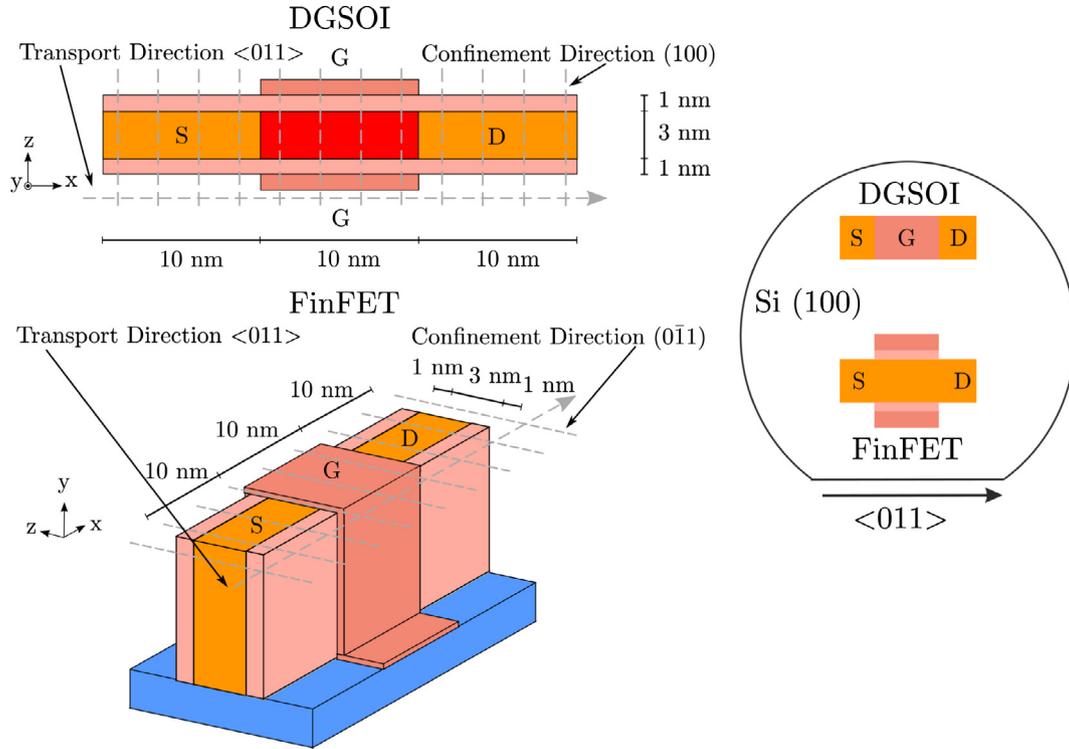


Fig. 1. DGSOI and FinFET structures analyzed in this work. 1D Schrödinger equation is solved for each grid point in the transport direction and BTE is solved by the MC method in the transport plane.

starting point simulation frame and the S/D tunneling algorithm are accurately described. Subsequently, the results and discussions are summarized in Section 3. Finally, the conclusions of this paper are summed up in Section 4.

2. Simulation set-up

Our MS-EMC simulator is based on the mode-space approach of quantum transport [9]. The device structure is divided into slices along the confinement direction where the 1D Schrödinger equation is solved, whereas the 2D Boltzmann Transport Equation (BTE) is solved in the transport plane as showed in Fig. 1. Both equations are coupled to the 2D Poisson equation to keep the self-consistency of the solution. This simulator has already demonstrated its capabilities studying different advanced nanodevices [10–13]. The main advantage of this tool against NEGF approach is the reasonable computational time when scattering mechanisms and quantum effects on the ultrascaled devices are taken into account.

In addition, the fundamentals of the free-flight technique of an electron used in Monte Carlo algorithms are based on the stochastic and ergodicity processes. It calculates the positions of each electron in the transport direction after a random flight time which finishes because of the random choice of a scattering event. After each flight, the new position and transport properties of the electrons are calculated. Depending on the carrier location and energy, our algorithm estimates the probability of undergoing a tunnel process. For this reason, another advantage of the MS-EMC simulator is the ability to switch on and off the tunneling process as it is included in a separate routine after each iteration.

The model employed here to include the S/D tunneling is an extension of the non-local band-to-band tunneling (BTBT) algorithm [14]. In that work, the same classical path and tunneling probability were considered, whereas the starting and ending point in the tunneling path belong to Valence and Conduction Band, respectively. The main advantage of this method is that, once it

has been implemented in the simulator, it is possible to extend it from the study of BTBT to that of S/D tunneling because the description of both mechanisms is based on the same assumptions.

In this work, the performance of DGSOI and FinFET devices is analyzed when S/D tunneling is included in order to determine its impact. The considered confinement direction of these devices on standard wafers changes between (100) for planar DGSOIs and (011) for vertical FinFETs, and $\langle 011 \rangle$ for the transport direction as depicted in Fig. 1. The differences in the confinement direction modify the electron distribution in the subbands, and, consequently, the potential profile. The carrier transport effective mass is also modified [15]. Table 1 summarizes the masses of each device and Table 2 shows their numerical values. Where $m_l = 0.916m_0$ and $m_t = 0.198m_0$ are the longitudinal and transversal effective masses in silicon, respectively, m_0 is the electron free-mass, m_x is the transport mass, m_z is the confinement mass, and Δ_2 and Δ_4 represent the degeneration factors of each valley. Moreover, the lower energy subband changes from Δ_2 in DGSOI to Δ_4 in FinFET.

These devices have been parametrized for gate lengths ranging from 5 nm to 20 nm. The rest of the technological parameters remains constant, channel thickness $T_{Si} = 3$ nm, gate oxide with Equivalent Oxide Thickness $EOT = 1$ nm and metal gate work function of 4.385 eV.

The position and energy of each electron are calculated after each free-flight as described above. In a semiclassical approximation, if the total energy of this electron is lower than the potential

Table 1
Effective mass in silicon for DGSOI and FinFET devices studied in this work where m_x is the transport mass and m_z is the confinement mass.

Device	Valley	m_x	m_y	m_z
DGSOI (100)(011)	Δ_2	m_t	m_t	m_l
	Δ_4	$\frac{2m_l m_t}{m_l + m_t}$	$\frac{m_l + m_t}{2}$	m_t
FinFET (011)(011)	Δ_2	m_t	m_l	m_t
	Δ_4	$\frac{m_l + m_t}{2}$	m_t	$\frac{2m_l m_t}{m_l + m_t}$

Table 2

Numerical values of effective mass in silicon for DGSOI and FinFET devices studied in this work where m_x is the transport mass and m_z is the confinement mass.

Device	Valley	m_x	m_y	m_z
DGSOI	Δ_2	0.198	0.198	0.916
(100)(011)	Δ_4	0.326	0.557	0.198
FinFET	Δ_2	0.198	0.916	0.198
(0 $\bar{1}$ 1)(011)	Δ_4	0.557	0.198	0.326

barrier at this position, the electron must undergo a backscattering. When S/D tunneling is taken into account, there is a probability for the electron to go through the barrier. There are two steps to determine that probability at a specific energy.

Firstly, the tunneling probability of the electron T_{dt} is calculated using the WKB approximation [16]:

$$T_{dt}(E) = \exp \left\{ -\frac{2}{\hbar} \int_a^b \sqrt{2m_{tr}^*(E_i(x) - E)} dx \right\} \quad (1)$$

where a and b are the starting and ending points, E and m_{tr}^* are the energy and transport effective masses of the electron, respectively, and $E_i(x)$ the energy of the i -th subband. This approximation has already been used to study this phenomenon in other electron devices [17]. Our MSB-EMC simulator offers a detailed description of the subband structure. Consequently, T_{dt} has been calculated for each electron keeping in mind the minimum energy of its subband instead of the Conduction Band [18].

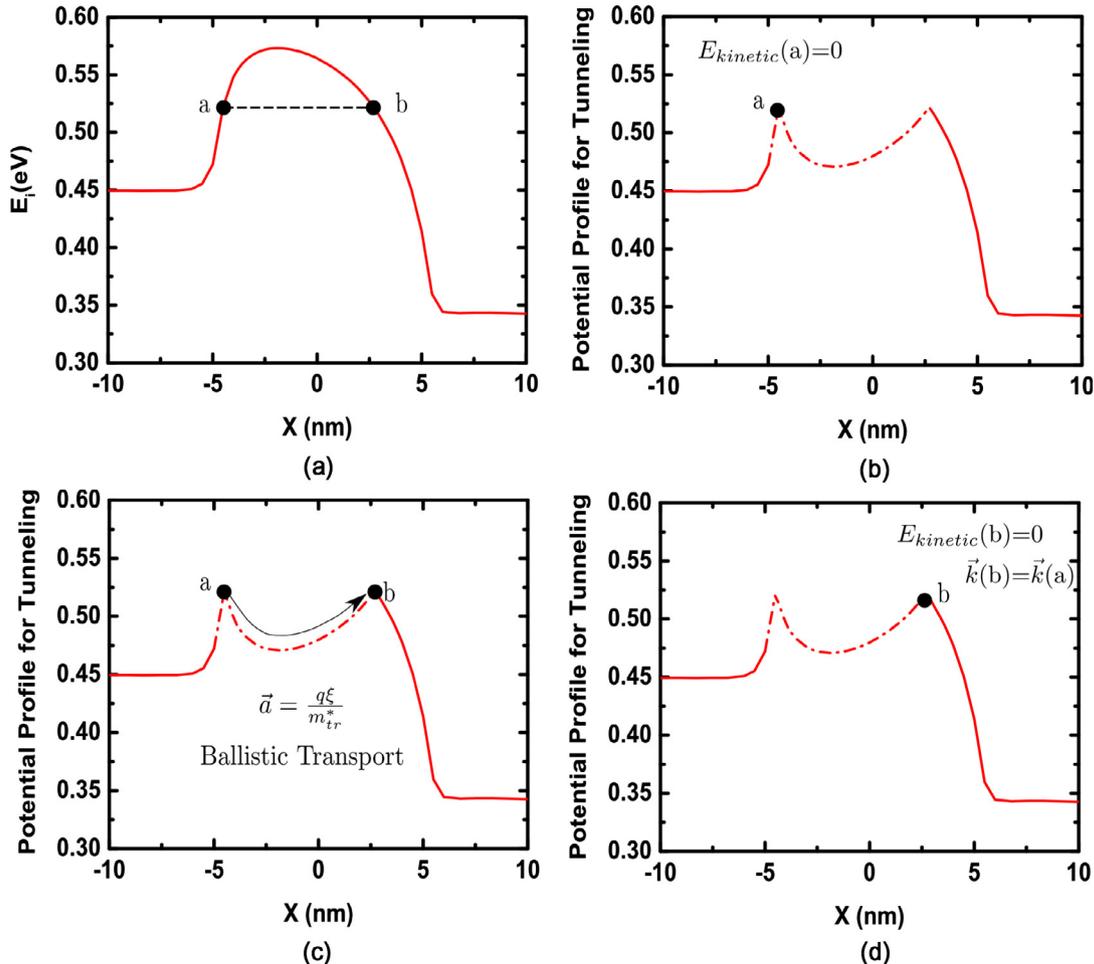


Fig. 2. Representation of the tunneling model: the potential barrier (a) is inverted and the particle is placed at the starting point a (b), it follows a classical path obeying Newton's second law of motion (c) until it reaches the ending point b (d).

In that point, several assumptions have been considered after each integration step to enhance the calculation of T_{dt} and to reduce the computational effort. The exact starting and ending points in the tunneling path are calculated to evaluate T_{dt} . In addition, a maximum tunneling rejection length is also introduced (L_{max}). If the tunneling length of an electron from the starting point to a specific integration step is higher than L_{max} , the calculation of

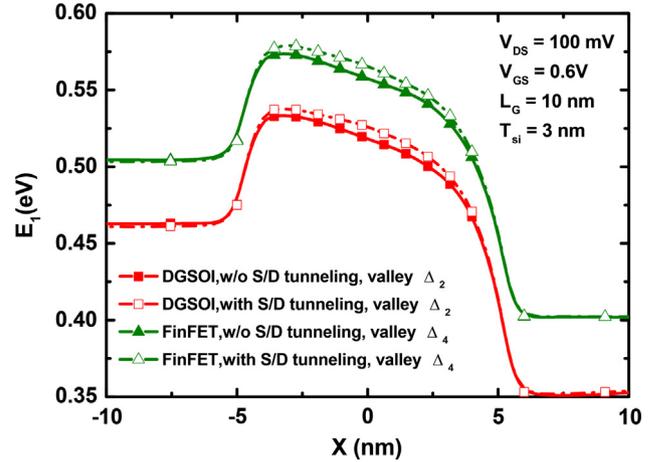


Fig. 3. Energy profile of the lowest energy subband in the 10 nm device for DGSOI (valley Δ_2) and FinFET (valley Δ_4) with and w/o S/D tunneling with $V_{GS} = 0.6$ V and $V_{DS} = 100$ mV.

T_{dt} stops. L_{max} has been chosen herein at $L_{max} = 10$ nm because T_{dt} decreases substantially for higher lengths. It remains constant regardless of the channel length.

Secondly, a rejection technique is used to determine whether the particle will tunnel or not. A uniform distributed random number r_{dt} is generated and compared to T_{dt} . On the one hand, if $r_{dt} > T_{dt}$, the electron will turn back with $v_x = -v_x$. On the other hand, if $r_{dt} \leq T_{dt}$, the electron will go through the barrier.

Subsequently, if the electron undergoes a tunnel process, it is required to find the most probable tunneling path to completely determine its new position. The motion inside the barrier obeys Newton mechanics considering an inverted potential profile and ballistic transport [19]. This classical trajectory could be found by the following steps [4] as shown in Fig. 2. Firstly, a pseudo-particle is placed at the starting point a with zero kinetic energy (Fig. 2(b)). It is assumed that this particle is going to exit the barrier

with the same transport properties. Consequently, its flight direction is maintained before starting its motion. It is also marked to force a ballistic transport inside the barrier. Then, it accelerates in this system according to Newton’s second law of motion (Fig. 2 (c)):

$$\vec{a} = \frac{q\zeta}{m_{tr}^*} \quad (2)$$

where ζ is the electric field. Lastly, it reaches the ending point b (Fig. 2(d)). Thereafter, the particle recovers its transport properties.

3. Results and discussion

A set of simulations at low bias condition has been performed to determine the importance of S/D tunneling on each device. The modifications in the energy profile of the lower energy subbands and the carrier transport effective mass caused by the difference in the confinement directions are shown in Figs. 3 and 4, respec-

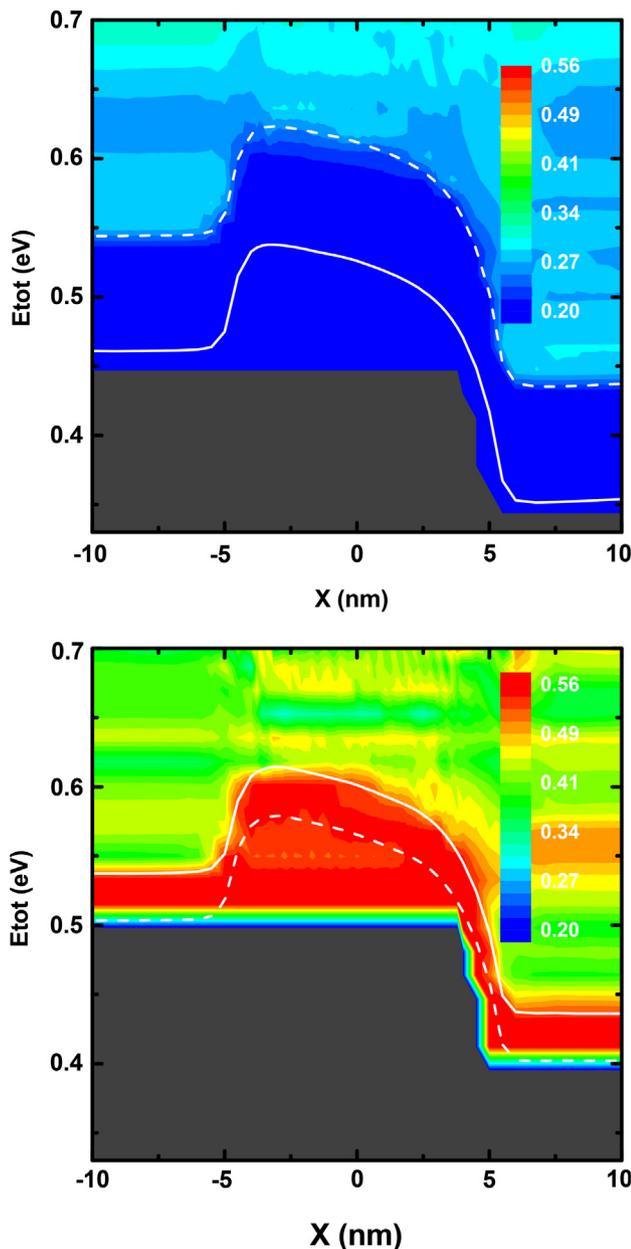


Fig. 4. Average effective mass of the electron distribution with the lower energy subband of the valley Δ_2 (solid) and of the valley Δ_4 (dashed) as a function of the total energy and the total population in the 10 nm device including S/D tunneling for DGSOI (top) and FinFET (bottom) with $V_{GS} = 0.6$ V and $V_{DS} = 100$ mV.

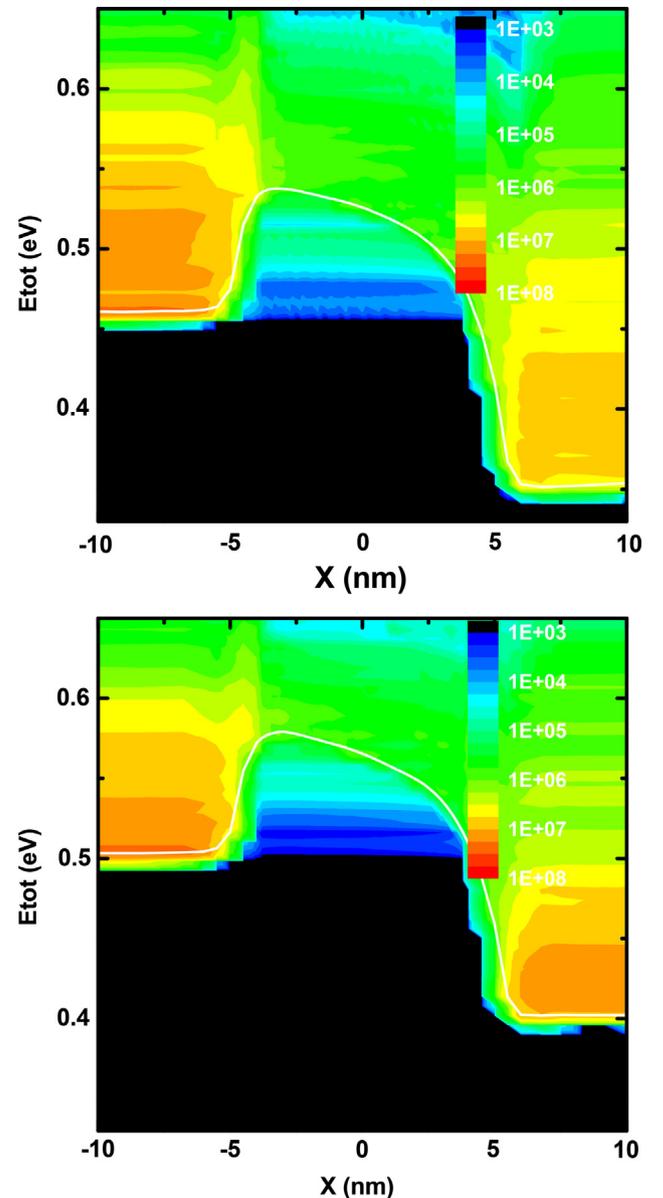


Fig. 5. Electron distribution in arbitrary units in the lower energy subband as a function of total energy in the 10 nm device including S/D tunneling for DGSOI (top) and FinFET (bottom) with $V_{GS} = 0.6$ V and $V_{DS} = 100$ mV.

tively. Both devices present similar energy profiles but the lower energy subband changes from Δ_2 in DGSOIs to Δ_4 in FinFETs (Fig. 3). This change modifies the distribution of the population and the effective transport mass in the subbands. Moreover, Fig. 3 shows the increase of the potential barrier when S/D tunneling is considered because of the existence of electrons located inside the potential barrier.

The average effective transport mass of the electrons as a function of the total energy and the total population which undergoes this tunnel process is higher in the FinFET than in the DGSOI as depicted in Fig. 4. These values correspond to m_x of the fundamental valleys in Table 2. It is also represented in Fig. 4 the lower energy profile of the less populated valleys: Δ_2 in FinFET, and Δ_4 in DGSOI. The average effective mass in these non-fundamental subbands decreases for the FinFET whereas it increases for the DGSOI. Despite this, the average effective mass continues being higher for the less populated valley in FinFET than in DGSOI.

As a result, assuming similar energy profile (Fig. 3), which means similar tunneling length at a specific starting point a , the higher is the value of m_{tr}^* in the fundamental valley in Eq. (1) for the FinFET orientation, the smaller is T_{dt} . Besides, the value of m_{tr}^* in the non-fundamental valley is higher in the DGSOI than in the FinFET, whereas the energy profile remains constant between both

valleys. However, the reduction of the population in this valley decreases the number of particles involved in S/D tunneling. For these reasons, the FinFET reduces its effectiveness of the tunnel phenomenon compared to the DGSOI one.

The higher T_{dt} , the higher the probability of an electron undergoing S/D tunneling for the same energy. It therefore increases the number of particles affected by S/D tunneling for the DGSOI than for the FinFET. This effect is shown in Fig. 5 where the electron distribution in arbitrary units from the fundamental subband as a function of total energy is represented.

Electrons with reduced energy must go through longer tunneling paths. When its length is similar to 10 nm, which corresponds to L_G in Fig. 5, the population decreases substantially. That is the reason why the maximum tunneling rejection length has been chosen at 10 nm.

The same effect is also shown in the percentage of electrons near the potential barrier affected by S/D tunneling respect to the total number of electron with lower energy than the top of the barrier in the same region, which is higher for DGSOI (Fig. 6 top) than for FinFET (Fig. 6 bottom). In addition, there is a maximum of this percentage for the FinFET due to a reduced height of the potential barrier. When V_{GS} increases, the height of the potential barrier decreases causing the enhancement of the thermionic current. It therefore induces the reduction of the number of electrons near the potential barrier with lower energy. It is necessary to highlight that the change in the channel length modifies

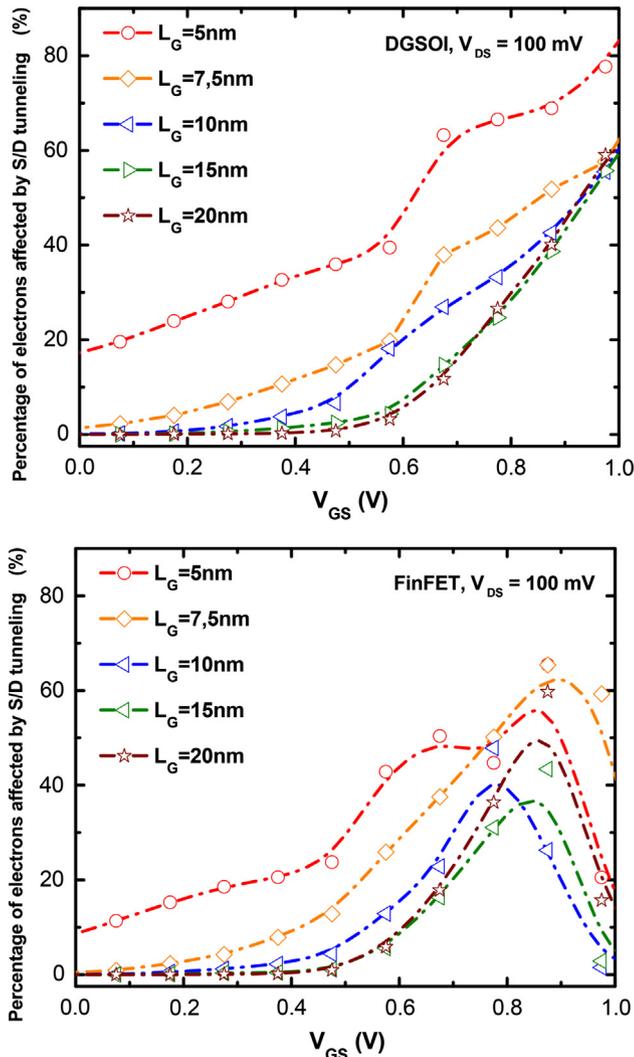


Fig. 6. Percentage of electrons near the potential barrier affected by S/D tunneling respect to the total number of electron with lower energy than the top of the barrier in the same region as a function of L_G at low drain bias for DGSOI (top) and for FinFET (bottom).

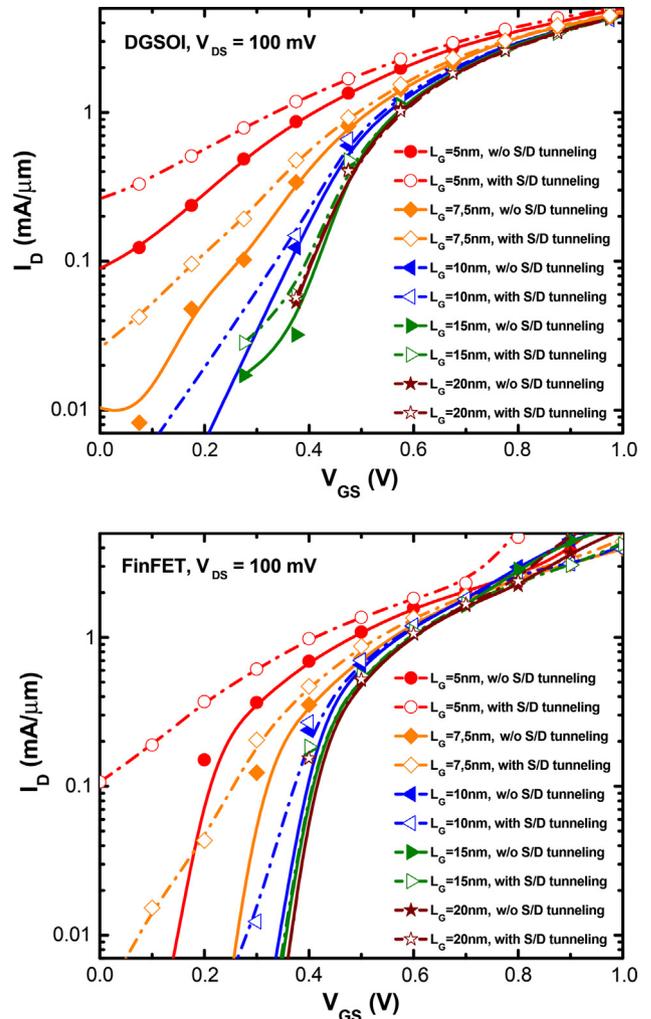


Fig. 7. I_D vs. V_{GS} as a function of L_G at low drain bias with and w/o considering S/D tunneling for DGSOI (top) and FinFET (bottom).

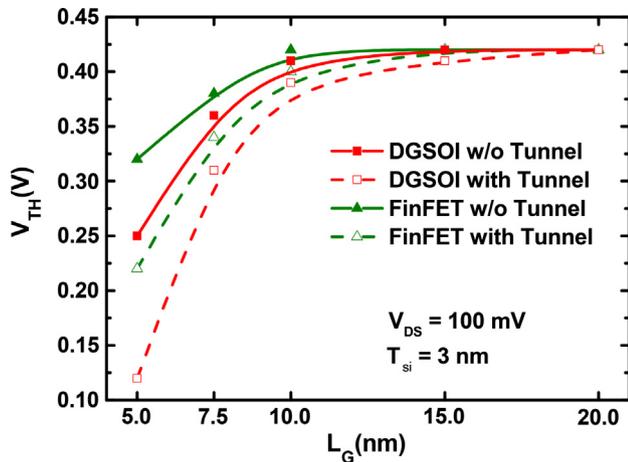


Fig. 8. Threshold voltage (V_{th}) for FinFETs and DGSOIs as a function of L_G at low bias condition and $T_{si} = 3$ nm with and w/o considering S/D tunneling.

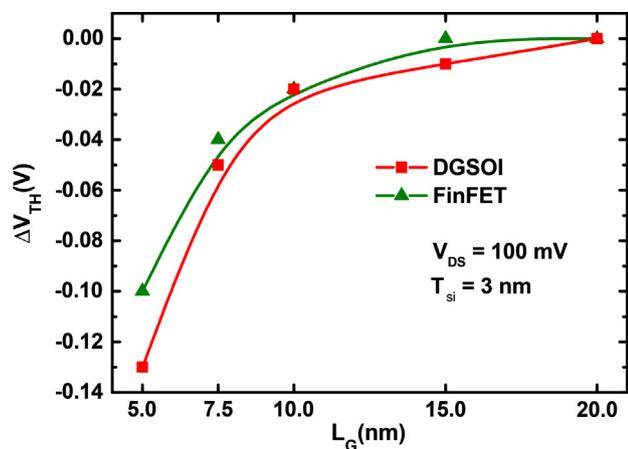


Fig. 9. Difference between the threshold voltage (ΔV_{th}) of a simulation considering S/D tunneling and w/o taking it into account as a function of L_G for DGSOIs and FinFETs at low drain bias condition and $T_{si} = 3$ nm.

the tunneling length and, consequently, T_{dt} . For this reason, the number of particles that suffer S/D tunneling increases when the devices are scaling down. By way of contrast, the maximum percentage appears in DGSOI devices but it is shifted to higher gate voltages (not shown).

This quantum effect produces a noticeable modification of the $I_D - V_{GS}$ characteristics (Fig. 7). Despite the increase of the potential barrier when S/D tunneling is included (Fig. 3), a higher current level is observed. The number of electrons that flows from source to drain is higher because of the possibility of tunneling through the barrier. This increase is also exacerbated when the devices are scaled down. As it is shown, the influence of the S/D tunneling is lower in the FinFET (Fig. 7 bottom) compared to DGSOI (Fig. 7 top).

The inclusion of tunneling introduces an important reduction in the threshold voltage (V_{th}) as it is shown in Fig. 8. Due to the reduced number of particles affected by this phenomenon in the FinFET compared to the DGSOI, the shift of the V_{th} in the vertical device than in the horizontal one. This effect is amplified in both devices as the channel length is reduced.

The impact of the S/D tunneling on the electrostatics can be observed in Fig. 9 where the threshold voltage variation (ΔV_{th}) between a simulation with and without taking it into account is

shown. This difference is also aggravated for reduced L_G because the influence of this quantum effect in the electrostatics is lower in the FinFET.

4. Conclusions

This work presents the implementation of S/D tunneling in a MSB-EMC simulator for the study of its impact in DGSOIs and FinFETs. Our simulations show important differences fully caused by the change in the confinement directions in both DGSOIs and FinFETs when S/D tunneling is taken into account due to the electron distribution and the variation of transport effective mass. Nevertheless, FinFET devices show less degradation in their subthreshold characteristics, and therefore are better candidates to implement future nodes, especially for ultra-low power applications.

Acknowledgment

The authors are grateful for the support given by the Spanish Ministry of Science and Innovation (TEC2014-59730-R), H2020 - REMINDER (687931), and H2020 - WAYTOGO-FAST (662175).

References

- [1] The international Technology Roadmap For Semiconductors (ITRS); 2013. <<http://www.itrs.net/>>.
- [2] Wang JW, Lundstrom M. Does source-to-drain tunneling limit the ultimate scaling of MOSFETs? In: Digest, international electron devices meeting; 2002. p. 707–10.
- [3] Hiroshi I. Future of nano CMOS technology. *Solid-State Electron* 2015;112:56–67.
- [4] Medina-Bailon C, Sampedro C, Gámiz F, Godoy A, Donetti L. Impact of S/D tunneling in ultrascaled devices, a multi-subband ensemble monte carlo study. In: 2015 International conference on simulation of semiconductor processes and devices (SISPAD). p. 21–4.
- [5] Kuhn KJ. Considerations for ultimate CMOS scaling. *IEEE Trans Electron Dev* 2012;59(7):1813–28.
- [6] Bohr M. The evolution of scaling from the homogeneous era to the heterogeneous era. In: 2011 IEEE international electron devices meeting (IEDM). p. 1–6.
- [7] Cristoloveanu S. How many gates do we need in a transistor? 2007 International semiconductor conference, vol. 1. p. 3–10.
- [8] Sampedro C, Donetti L, Gámiz F, Godoy A. 3D multi-subband ensemble Monte Carlo simulator of FinFETs and nanowire transistors. In: 2014 International conference on simulation of semiconductor processes and devices (SISPAD). p. 21–4.
- [9] Venugopal R, Ren Z, Datta S, Lundstrom MS, Jovanovic D. Simulating quantum transport in nanoscale transistors: real versus mode-space approaches. *J Appl Phys* 2002;92(7):3730–9.
- [10] Sampedro C, Gámiz F, Godoy A, Valín R, García-Loureiro A, Ruiz FG. Multi-subband Monte Carlo study of device orientation effects in ultra-short channel DGSOI. *Solid-State Electron* 2010;54(2):131–6.
- [11] Sampedro C, Gámiz F, Godoy A, Valín R, García-Loureiro A, Rodríguez N, et al. Multi-subband ensemble Monte Carlo simulation of bulk MOSFETs for the 32 nm-node and beyond. *Solid-State Electron* 2011;65–66(1):88–93.
- [12] Sampedro C, Gámiz F, Donetti L, Godoy A. Reaching sub-32 nm nodes: ET-FDSOI and BOX optimization. *Solid-State Electron* 2012;70:101–5.
- [13] Sampedro C, Gámiz F, Godoy A. On the extension of ET-FDSOI roadmap for 22 nm node and beyond. *Solid-State Electron* 2013;90:23–7.
- [14] Shen C, Yang L-T, Samudra G, Yeo Y-C. A new robust non-local algorithm for band-to-band tunneling simulation and its application to tunnel-FET. *Solid-State Electron* 2011;57(1):23–30.
- [15] Rahman A, Lundstrom MS, Ghosh AW, Rahman A, Lundstrom MS, Ghosh AW. Generalized effective-mass approach for n-type metal-oxide-semiconductor field-effect transistors on arbitrarily oriented wafers. *J Appl Phys* 2005;97:053702. doi: <http://dx.doi.org/10.1063/1.1845586>.
- [16] Griffiths DJ. The WKB approximation. In: Introduction to quantum mechanics. New Jersey: Prentice Hall; 1995. p. 274–97 [ch. 8].
- [17] Sun GD, Liu XY, Han RQ. Monte Carlo simulation of Schottky contact with direct tunneling model. *Semicond Sci Technol* 2003;18:576–81.
- [18] Revelant A, Palestri P, Selmi L. Multi-subband semi-classical simulation of n-type tunnel-FETs. In: 2012 13th International conference on ultimate integration on silicon (ULIS). p. 187–90.
- [19] Huang Z, Feuchtwang TE, Cutler PH, Kazes E. Wentzel-Kramers-Brillouin method in multidimensional tunneling. *Phys Rev A* 1990;41(1):32–41.



Process modules for GeSn nanoelectronics with high Sn-contents



C. Schulte-Braucks^{a,*}, S. Glass^a, E. Hofmann^a, D. Stange^a, N. von den Driesch^a, J.M. Hartmann^{b,c}, Z. Ikonic^d, Q.T. Zhao^a, D. Buca^a, S. Mantl^a

^a Peter-Gruenberg-Institute 9 (PGI-9) and JARA-FIT, Forschungszentrum Juelich GmbH, 52428 Juelich, Germany

^b University of Grenoble Alpes, F38000, France

^c CEA, LETI, MINATEC Campus, F-38054 Grenoble, France

^d Institute of Microwaves and Photonics, School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT Leeds, United Kingdom

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

GeSn
MOSFET
High-k/metal gate
NiGeSn

ABSTRACT

This paper systematically studies GeSn n-FETs, from individual process modules to a complete device. High-k gate stacks and NiGeSn metallic contacts for source and drain are characterized in independent experiments. To study both direct and indirect bandgap semiconductors, a range of 0–14.5 at.% Sn-content GeSn alloys are investigated. Special emphasis is placed on capacitance-voltage (C-V) characteristics and Schottky-barrier optimization. GeSn n-FET devices are presented including temperature dependent I-V characteristics. Finally, as an important step towards implementing GeSn in tunnel-FETs, negative differential resistance in Ge_{0.87}Sn_{0.13} tunnel-diodes is demonstrated at cryogenic temperatures. The present work provides a base for further optimization of GeSn FETs and novel tunnel FET devices.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recently, GeSn alloys have emerged as promising group IV semiconductors for electronic [1] as well as photonic [2,3] applications. The breakthrough in epitaxial growth of high-Sn content and strain relaxed layers, enabled fundamental direct bandgap group IV alloys grown on Si [4,5]. The direct bandgap property is a requirement for efficient Si based light emitters. However, such alloys may also serve as performance boosters in nanoelectronics. The small effective mass and associated reduction of intra-valley scattering yields high mobility Γ -electrons. Performance of GeSn based n-type Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) should then be superior to those of their pure Ge counterparts. In addition, the possibility of combining direct band-to-band tunneling and low bandgap should yield efficient tunnel field effect transistors (TFETs).

Mobility calculations, using the 8-band k.p method for the Γ -valley band structure and effective mass (including nonparabolicity) for the L-valley band structure, predict a significant mobility enhancement as soon as the population of Γ -valley is sufficiently large. The calculations take acoustic phonon, deformation potential, alloy disorder, ionized impurity, and inter-valley scattering into account. Modulation of the Γ -valley population

can be achieved either by changing Sn-content or layer strain. For Sn contents below ~ 9 at.% GeSn alloys are indirect bandgap semiconductors. Hence, the electron mobility is dominated by electrons occupying the L-valley. For larger Sn contents, above the indirect to direct bandgap transition, the Γ -valley becomes increasingly populated and the electron mobility is boosted significantly. The calculated Sn-dependent Γ -valley population and effective (weighted-average) mobility is shown in Fig. 1(a). GeSn pseudomorphically grown on Ge is Sn-content dependently biaxially compressively strained. However, growing thicker GeSn layers leads to strain relaxation or even tensilely strained GeSn when combining different Sn-contents [6,7]. Decreasing compressive strain has the same effect as increasing Sn content, leading to an increase of Γ -valley population and of Γ -electron mobility according to Fig. 1(b). In contrast, L-electron mobility is of the order of $4 \times 10^3 \text{ cm}^2/\text{Vs}$ for all Sn-contents and strain values presented here. The large difference between Γ - and L-electron mobility comes from a much larger effective mass of the L-electrons. However, just above the indirect to direct transition, the Γ -electron mobility is still limited by strong inter-valley Γ -L scattering, which gives a large relative contribution to total scattering due to a large L-valley density of states (while the L-electrons are less affected, because of a smaller density of states of Γ). Together with a small fraction of Γ -electrons, this implies that in alloys with 8.5 at.% Sn, at the direct to indirect transition, the mobility is always dominated by the L-electrons. However, when Γ -L spacing increases,

* Corresponding author.

E-mail address: c.schulte-Braucks@fz-juelich.de (C. Schulte-Braucks).

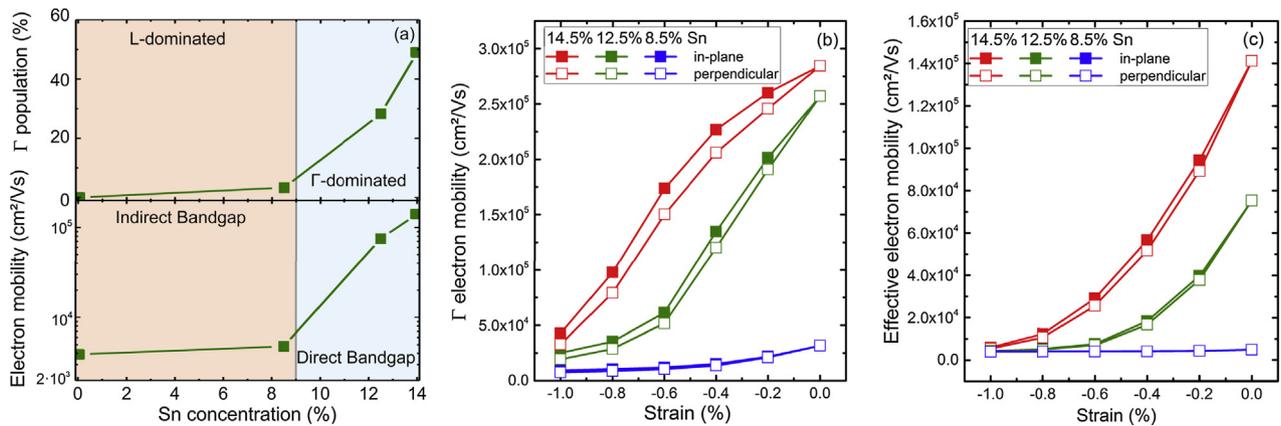


Fig. 1. (a) Calculated Γ -valley population (top) and effective electron mobility, at 300 K and 10^{14} cm^{-3} electron density, vs Sn-content (bottom) at zero strain. (b) Γ and (c) effective electron mobilities for various GeSn alloys, dependent on biaxial strain.

by decreasing strain or increasing Sn-content, inter-valley scattering is reduced, and Γ -population becomes significant. Consequently, not only the Γ -electron mobility but also the effective electron mobility, displayed in Fig. 1(c), strongly increases. It is also worth noting that biaxial strain induces a non-negligible anisotropy of the Γ -valley, and in case of compressive strain the in-plane mobility, relevant for MOSFETs, is larger (by up to $\sim 20\%$) than perpendicular mobility.

Preliminary works on p- and n- MOSFETs [8,9] and even TFETs [10] based on GeSn alloys have been reported, however, the Sn-contents and strain values were far below the indirect to direct transition. The low solid solubility of Sn in Ge $< 1 \text{ at.}\%$ and the non-equilibrium growth restricts the thermal budget to temperatures $< 350 \text{ }^\circ\text{C}$ for Sn-contents above 10 at.% making process integration challenging.

In this work we discuss advances on low temperature process modules for GeSn-FET devices with Sn-contents up to 13 at.%, including high-k/metal gate stack deposition and low resistivity metallic NiGeSn contact formation. Emphasis is placed on the fabrication and characterization of metal-semiconductor-metal (MSM) diodes for Schottky-barrier height (SBH) extraction and Schottky-barrier tuning by dopant segregation (DS). GeSn n-FETs are fabricated using these modules and, as a step towards novel devices, *p-i-n* tunneling diode characterization is presented.

2. Experimental

Due to the low solid solubility of Sn in Ge ($< 1 \text{ at.}\%$) growth conditions for GeSn with up to 13 at.% Sn are far from equilibrium. An industry compatible AIXTRON TRICENT RP-CVD epitaxial reactor was employed to grow these layers on 200 mm Ge buffered Si (100) wafers [11]. All process temperatures were kept below $350 \text{ }^\circ\text{C}$ in order to avoid Sn-diffusion and segregation. As a first key module, MOS-capacitors (MOScaps) with high-k/metal gate stacks on GeSn were investigated. After a wet HF-HCl surface preparation, 6 nm HfO_2 high-k dielectric was deposited at low temperature by atomic layer deposition (ALD) followed by 40 nm sputter deposited TiN metallization both using 200 mm, industry compatible reactors. MOScaps with Sn-contents between 0 at.% (Ge-substrate) and 12.5 at.% were fabricated. Standard CMOS technology, such as photo lithography and reactive ion etching, was used to define the structures. The fabrication ended with a lift-off process after the deposition of 150 nm Al for contacts followed by forming gas annealing at $300 \text{ }^\circ\text{C}$. A set of Capacitance-Voltage (C-V) characteristics at different frequencies measured on

TiN/ HfO_2 / $\text{Ge}_{0.915}\text{Sn}_{0.085}$ capacitors is shown in Fig. 2(a). The good GeSn/ HfO_2 interface quality is evidenced by the small frequency dependent flat-band voltage shift and the small frequency dispersion in accumulation. Typical for low bandgap semiconductors, the C-V curves feature a strong minority carrier inversion response even at high frequencies $> 100 \text{ kHz}$. As a consequence, a reliable extraction of the interface state density (D_{it}) using the conduction method at room temperature becomes difficult [12]. However, the minority carrier inversion response is reduced at lower temperatures. We have thus used the low temperature conductance method as described in work by Nicollian and Brews [13] at $T < 120 \text{ K}$ to extract D_{it} values of $2 \times 10^{12} \text{ cm}^{-2} \text{ eV}^{-1}$ at midgap for GeSn capacitors with different Sn contents (Fig. 2(b)). A study focusing on the process development and characterization of ternary SiGeSn MOScaps has been published recently [14].

A second fundamental module is contact formation. Metal-semiconductor-metal (MSM) diodes based on NiGeSn/GeSn Schottky contacts were fabricated using an oxide mask. After native oxide removal, 10 nm of Ni were deposited by sputter deposition and $\sim 23 \text{ nm}$ NiGeSn was formed by rapid thermal annealing for 10 s in N_2/H_2 forming gas atmosphere. Unreacted Ni was removed by sulfuric acid (96% aq.). The van-der-Pauw method [15] has been used to measure the sheet resistance of the so formed NiGeSn films. The lowest sheet resistance was obtained by stano-germanidation at $325 \text{ }^\circ\text{C}$ [16]. The low-resistive NiGeSn-phase could be maintained over the complete available Sn-content range from 0 to 12.5 at.%. The sheet resistance of NiGeSn for several Sn-contents is shown in Fig. 3(d). Furthermore, a smooth NiGeSn/GeSn interface was obtained as shown by the cross-sectional Transmission-Electron-Microscopy (TEM) image in the inset of Fig. 3(d).

Current transport properties across a metal-semiconductor contact are determined by the Schottky-barrier. Previous studies have investigated the electron Schottky-barrier on NiGeSn/ $\text{Ge}_{0.958}\text{Sn}_{0.042}$ [17] and hole Schottky-barrier on NiGeSn/ $\text{Ge}_{0.86}\text{Si}_{0.07}\text{Sn}_{0.07}$ [18]. Here, we determine the NiGeSn/GeSn hole Schottky-barrier from MSM diodes with different contact areas and for several Sn-concentrations using the activation-energy method. The advantage of this method is that the electrically active contact area does not need to be known, e.g. current crowding does not affect the Schottky-barrier extraction. The temperature dependent I-V characteristics were measured in a liquid nitrogen cooled cryostat under high vacuum where the temperature range from 400 K to 100 K is covered in 10 K incremental steps. From Arrhenius plots of the current characteristics for different voltages (Fig. 3(a)) the Schottky-barrier height (SBH) was extracted. According to

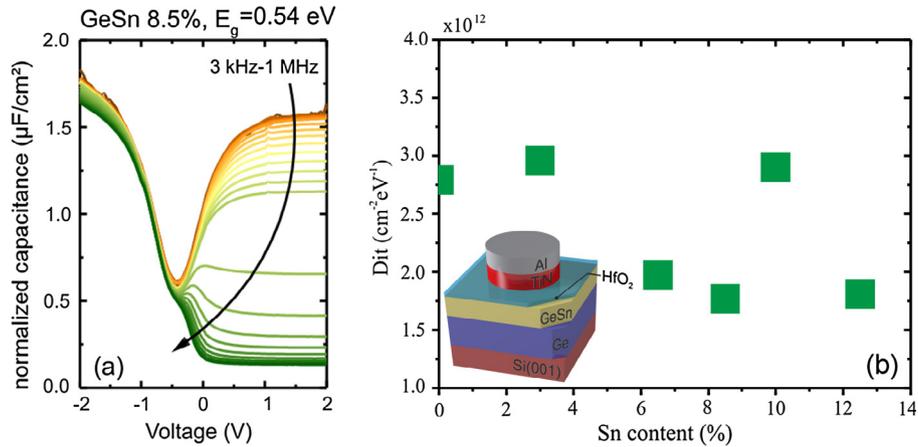


Fig. 2. (a) CV-characteristics of TiN/6 nm HfO₂/Ge_{0.915}Sn_{0.085} MOSCap for a set of frequencies. (b) Dit at midgap for several Sn-contents extracted at temperatures below 120 K.

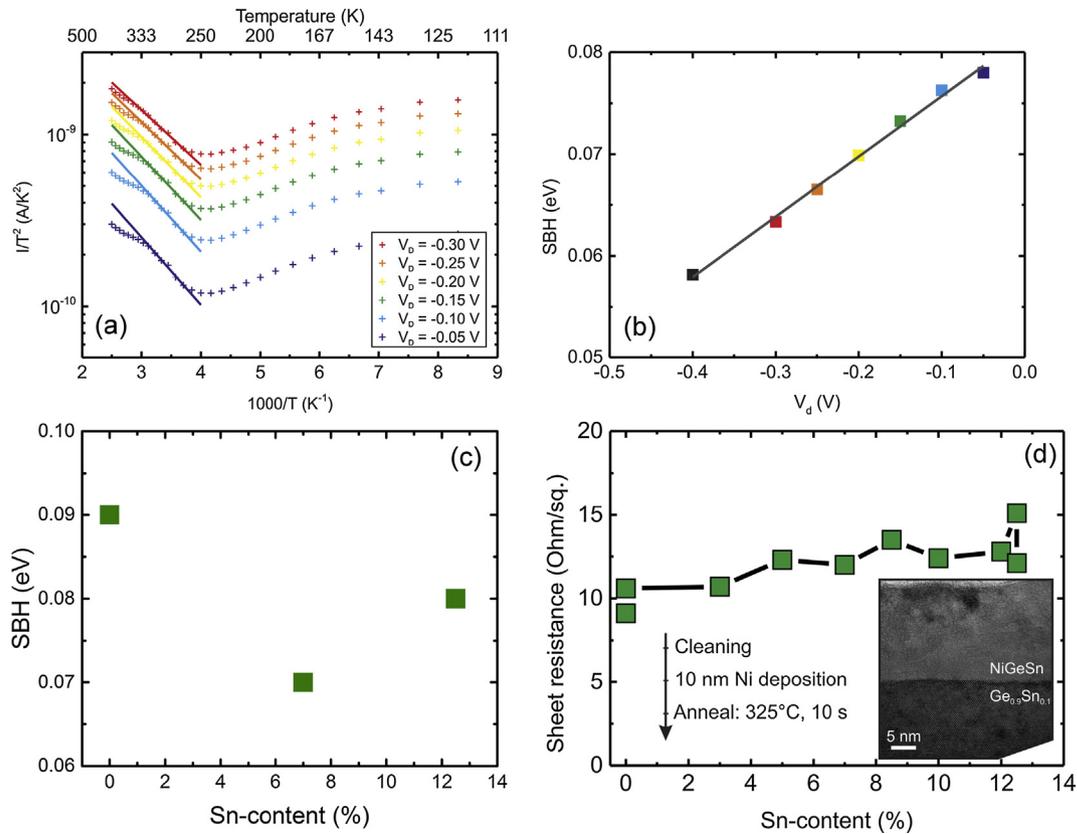


Fig. 3. Investigation of NiGeSn Schottky-contacts. (a) Arrhenius plot of current characteristics for a 12.5 at.% Sn sample. The linear region is fitted and SBH is extracted from the slope. (b) SBH vs. applied bias from (a). The SBH for 0 V is extracted by linear extrapolation. (c) Extracted 0 V SBH for various Sn contents. (d) Sheet resistance of NiGeSn fabricated on several GeSn layers. The inset depicts a TEM micrograph of a NiGeSn/GeSn contact.

thermionic-emission-diffusion theory, the voltage dependent SBH can be extracted from the slope s of the linear region in the $\ln|I/T^2|$ plot via

$$SBH = -s \cdot \frac{k}{e},$$

where k is Boltzmann's constant and e the electron charge. Outside this linear region, the current is determined by the series resistance (high temperatures) or the shunt resistance (low temperatures). The primary source of the former is the GeSn resistivity while the latter is impacted by parasitic currents. As the MSM diode consists of two back-to-back Schottky diodes, the current corresponds to the

reverse bias I-V characteristic at all times. The magnitude of the current is given by the lower Schottky-barrier – hole or electron barrier – in undoped semiconductors. In our case the hole Schottky-barrier is observed, as the nominally intrinsic GeSn layers are actually p-type. The main reasons for the voltage dependence of the SBH observed in Fig. 3(b) are image force and static lowering due to the applied voltage. By linearly extrapolating to 0 V these effects are suppressed and the hole Schottky-barrier is obtained. Fig. 3(c) shows the results for NiGeSn/GeSn Schottky contacts with Sn contents of 0 at.%, 7 at.% and 12.5 at.%. Throughout the entire Sn content range, the hole Schottky-barrier remains below 0.10 eV, making NiGeSn an ideal contact for p-type devices. However, this

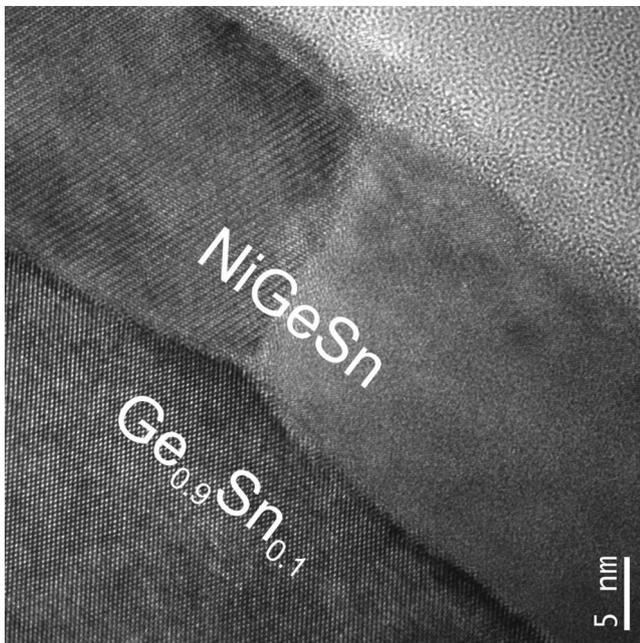


Fig. 4. High-resolution TEM image of the NiGeSn/GeSn interface after BF_2 implantation with 10 kV and $1 \times 10^{15} \text{ cm}^{-2}$ proving the good crystalline quality of the GeSn after NiGeSn-formation.

might imply very high Schottky-barriers for electrons leading to high S/D resistances for n-type GeSn and demanding further investigation on n-type GeSn-contacts over a wide Sn-content range.

For a metal-semiconductor interface with high carrier concentrations, the tunneling current component through the barrier is increased, which reduces the experimentally observed SBH. A well-known method to modify the SBH using this effect is dopant segregation [19]. Dopants are implanted shallowly into the contact windows before metallization. During the following stano-germanidation step the entire implanted region is consumed. Therefore, semiconductor quality is conserved as the damaged area is fully converted to stano-germanide. The high-crystalline quality can be seen in the high-resolution TEM-image in Fig. 4.

Due to the different solubility of dopants in metal and semiconductor the snow plough effect leads to dopant diffusion through the metallic region into the semiconductor at the interface. This results in a sharp doping profile with a high dopant concentration at the metal-semiconductor interface. The dopant segregation effect in GeSn for both n- and p-type dopants is presented below. Phosphorous (P), arsenic (As) and boron (BF_2) were implanted into GeSn test-structures with a dose of $1 \times 10^{15} \text{ cm}^{-2}$ at energies of 7,

13 and 10 keV, respectively. In GeSn, P and As act as n-type dopants, while B is a p-type dopant. The implanted region was then converted into NiGeSn as described above. Subsequently, doping profiles were measured by means of Time of Flight Secondary-Ion-Mass-Spectrometry (ToF-SIMS). Whereas there is no peak visible in the doping profile for P, a snow plough effect is observed for both As and B leading to a peak in the As/B-concentration at the NiGeSn/GeSn interface (Fig. 5). The differences in DS for the n-type dopants As and P might be attributed to differences in solubility and diffusion. Nonetheless, as DS is possible both for n- and p-type dopants, this effect can be used to modify the SBH.

In order to investigate the impact of DS, NiGeSn contacts were fabricated on in-situ phosphorus doped $\text{Ge}_{0.875}\text{Sn}_{0.125}$ (GeSn:P) with a $2.7 \times 10^{18} \text{ cm}^{-3}$ n-type carrier concentration. DS was then performed with As or B using the process described above. Since activated As provides electrons in GeSn, DS increases the majority carrier concentration at the NiGeSn/GeSn:P interface. For p-type B, DS yields the opposite. An increase in majority carrier concentration at the interface allows for a higher tunneling component through the barrier. Consequently, the effective SBH observed by the charge carriers is reduced. Fig. 6 shows I-V characteristics measured from one NiGeSn contact to the next, for samples without DS or with As or B DS. As expected, the I-V curves become more and more Ohmic when increasing the electron concentration at the NiGeSn/GeSn:P interface (e.g. by switching from B DS to no DS to As DS samples).

Combining the above described process modules, GeSn n-MOSFETs were fabricated with Sn-contents of 0 at.%, 7 at.% and 12.5 at.% using ion implanted source/drain (S/D) contacts after forming a gate stack with TiN/HfO_2 . Transfer curves of $\text{Ge}_{0.93}\text{Sn}_{0.07}$ n-FETs for a series of temperatures are shown in Fig. 7.

At room temperature, the device shows a low $I_{\text{on}}/I_{\text{off}}$ ratio while a reduced on current at lower temperatures is due to the poor n+/p junctions in the S/D regions. The limited process temperatures used here in order to avoid Sn diffusion (max. 300 °C), was not enough to recrystallize the amorphized regions created by ion implantation, leading to very poor junctions with low activation and high access resistances. This is even more critical for high Sn-content devices, as shown in Fig. 8 at 80 K. Apart from the un-healed implantation damage, the unintentional background doping of GeSn increases with the Sn-content. Furthermore, the bandgap is decreased. Both factors lead to increased S/D-leakage and gate induced drain-leakage (GIDL) which is caused by band-to-band tunneling and increases exponentially with the reduced bandgap. This is also visible in the temperature dependence of the transfer characteristics in Fig. 7. The S/D leakage strongly decreases for temperatures below 200 K. The solution for maintaining crystalline GeSn is the use of in-situ doping and selective

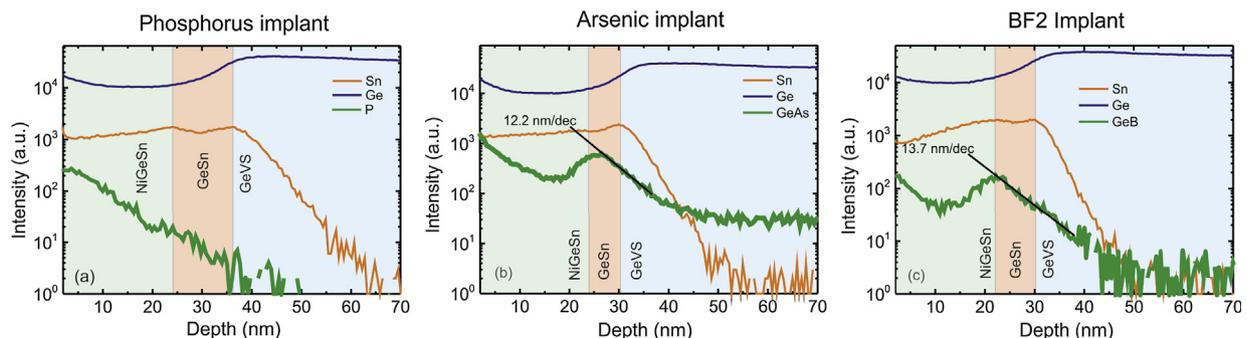


Fig. 5. SIMS-profiles of NiGeSn/GeSn contacts with 10 at.% Sn: For Phosphorous implantation no segregation was observed (a), whereas there is a clear peak in the Arsenic (b) and Boron (c) profiles. For higher sensitivity As and B are measured as GeAs and GeB.

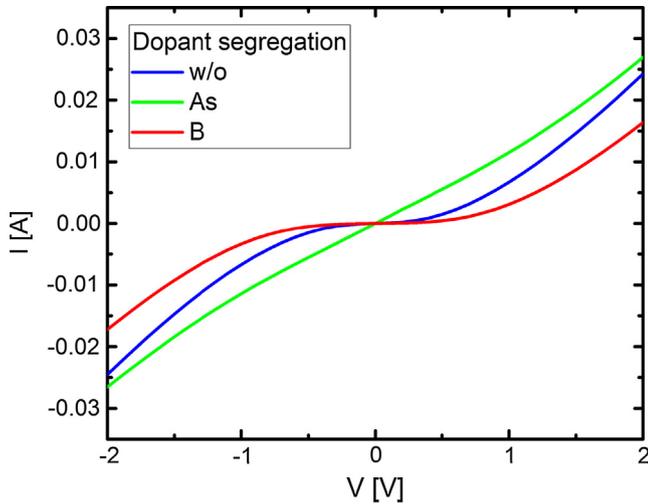


Fig. 6. Impact of DS on n-type $\text{Ge}_{0.875}\text{Sn}_{0.125}:\text{P}$. As the n-type carrier concentration at the NiGeSn/GeSn:P interface increases, the I-V characteristic becomes more Ohmic.

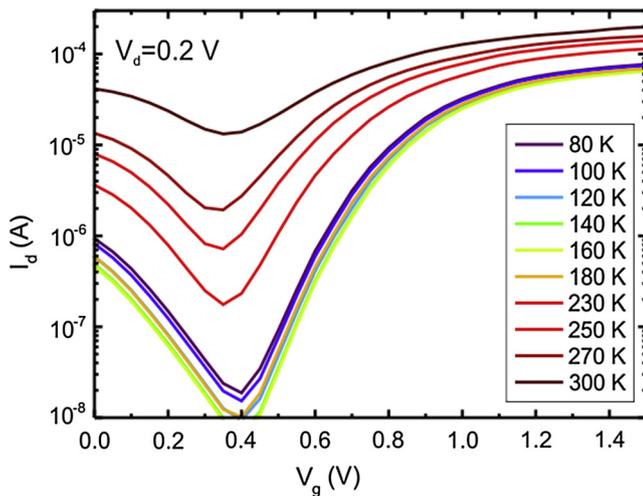


Fig. 7. Transfer characteristics of $\text{Ge}_{0.93}\text{Sn}_{0.07}$ n-FETs at different temperatures.

growth in the S/D region. The in-situ doping is discussed below in terms of tunneling diodes.

As a demonstration of the potential of direct bandgap GeSn for band to band tunneling and the advantage of in-situ doping over ion implantation, we have fabricated GeSn tunneling diodes as an important step towards advanced GeSn based TFETs. We could push the Sn-content up to 13 at.% as a follow up to previous results with a stack of 9 at.% and 11 at.% [1] enabling an even lower bandgap and higher directness of the GeSn. As a proof of band-to-band tunneling, negative differential resistance (NDR) is observed at cryogenic temperatures (Fig. 9), demonstrating a high doping level of both p- and n-type dopants, which is essential for MOSFETs and TFETs. However, due to enhanced diffusion and trap assisted tunneling (TAT) in this low-bandgap semiconductor, the NDR vanishes for temperatures above 100 K. For forward bias > 0.1 V two distinct regions, separated by a kink in the slope of the I-V curve, are visible. While the middle part of the curve $0.1 \text{ V} < V_d < 0.3 \text{ V}$ can be attributed to TAT, the diffusion current dominates for strong forward bias $> 0.3 \text{ V}$. We expect further improvements in the peak to valley current ratio and a move towards room temperature NDR with optimized doping profiles.

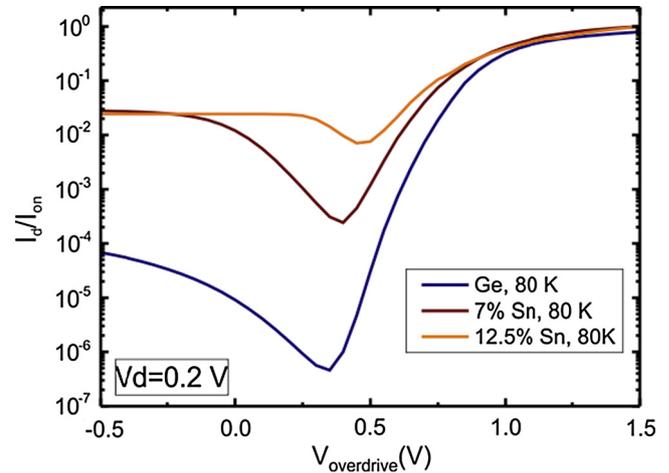


Fig. 8. I_d/I_{on} ratio of GeSn n-FETs at 80 K for several Sn-contents.

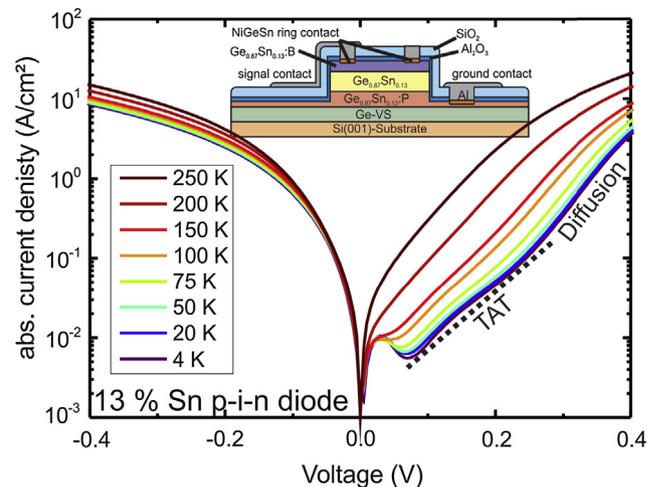


Fig. 9. Temperature dependent I-V measurements of a $\text{Ge}_{0.87}\text{Sn}_{0.13}$ p-i-n diode showing clear NDR at cryogenic temperatures.

3. Conclusion

In this work, process module developments for GeSn FETs were presented and assembled to for the fabrication of GeSn n-FETs. A wide range of Sn-contents was covered, allowing the study of both indirect and direct bandgap GeSn alloys. TiN/ HfO_2 /GeSn MOScaps, showing good C-V characteristics with D_{it} levels of $10^{12} \text{ eV}^{-1} \text{ cm}^{-2}$, have been studied for use as gate stacks. NiGeSn is shown to have low sheet resistances over the entire Sn-content range and very small Schottky barrier heights on p-GeSn. To further optimize the Schottky contacts, dopant segregation with both As and B, was demonstrated for NiGeSn contacts. In the case of n-GeSn, it is shown that As dopant segregation leads to increasingly Ohmic contact behavior. Due to the metastability of GeSn, junctions made by ion implantation have proven to be challenging. A possible solution is in-situ doping which reveals its potential in GeSn-tunnel diodes with 13 at.% Sn where characteristic negative differential resistance is observed at cryogenic temperatures.

Acknowledgments

This research received funding from the EU FP7 project E2SWITCH (619509) and the BMBF project UltraLowPow (16ES0060 K).

References

- [1] Schulte-Braucks C, Stange D, von den Driesch N, Blaeser S, Ikonic Z, Hartmann JM, et al. Negative differential resistance in direct bandgap GeSn p-i-n structures. *Appl Phys Lett* 2015;107:042101-1-4. <http://dx.doi.org/10.1063/1.4927622>.
- [2] Soref RA, Buca D, Yu S-Q. Group IV photonics- driving integrated optoelectronics. *Opt Photonics News* 2016(January):32-9.
- [3] Wirths S, Geiger R, Driesch NV Den, Stange D, Zabel T, Ikonic Z, et al. Direct bandgap GeSn microdisk lasers at 2.5 μm for monolithic integration on Si-platform; 2015, p. 36-9.
- [4] Wirths S, Geiger R, Den Driesch NV, Mussler G, Stoica T, Mantl S, et al. Lasing in direct-bandgap GeSn alloy grown on Si 2015:1-14. <http://dx.doi.org/10.1038/nphoton.2014.321>.
- [5] Stange D, Wirths S, von den Driesch N, Mussler G, Stoica T, Ikonic Z, et al. Optical transitions in direct-bandgap $\text{Ge}_{1-x}\text{Sn}_x$ alloys. *ACS Photon* 2015;2:1539-45. <http://dx.doi.org/10.1021/acsp Photonics.5b00372>.
- [6] Wirths S, Ikonic Z, Tiedemann AT, Holländer B, Stoica T, Mussler G, et al. Tensely strained GeSn alloys as optical gain media. *Appl Phys Lett* 2013;103. <http://dx.doi.org/10.1063/1.4829360> [192110_1-192110_5].
- [7] Wirths S, Stange D, Pampillón M-A, Tiedemann AT, Mussler G, Fox A, et al. High- k gate stacks on low bandgap tensile strained Ge and GeSn alloys for field-effect transistors. *ACS Appl Mater Interfaces* 2015;7:62-7. <http://dx.doi.org/10.1021/am5075248>.
- [8] Gupta S, Vincent B, Yang B, Lin D, Gencarelli F, Lin J-Y, et al. Towards high mobility GeSn channel nMOSFETs: Improved surface passivation using novel ozone oxidation method. In: *IEDM*. 2012. p. 16.2.1-4. <http://dx.doi.org/10.1109/IEDM.2012.6479052>.
- [9] Han G, Su S, Zhan C, Zhou Q, Yang Y, Wang L, et al. High-Mobility Germanium-Tin (GeSn) P-channel MOSFETs Featuring Metallic Source/Drain and Sub-370 C Process Modules; 2011, p. 402-4.
- [10] Yang Y, Su S, Guo P, Wang W, Gong X, Wang L, et al. Towards direct band-to-band tunneling in P-channel tunneling field effect transistor (TFET): technology enablement by Germanium-tin (GeSn). In: *Int Electron Devices Meet*. 2012. p. 16.3.1-4. <http://dx.doi.org/10.1109/IEDM.2012.6479053>.
- [11] von den Driesch N, Stange D, Wirths S, Mussler G, Holländer B, Ikonic Z, et al. Direct bandgap group IV epitaxy on Si for laser applications. *Chem Mater* 2015;27:4693-702. <http://dx.doi.org/10.1021/acs.chemmater.5b01327>.
- [12] Martens K, Chui CO, Brammertz G, De Jaeger B, Kuzum D, Meuris M, et al. On the correct extraction of interface trap density of MOS devices with high-mobility semiconductor substrates. *IEEE Trans Electron Devices* 2008;55:547-56. <http://dx.doi.org/10.1109/TED.2007.912365>.
- [13] Nicollian EH, Brews JR. *MOS (metal oxide semiconductor) physics and technology*. Wiley; 2002.
- [14] Schulte-Braucks C, von den Driesch N, Glass S, Tiedemann AT, Breuer U, Besmehn A, et al. Low temperature deposition of high- k /metal gate stacks on high-Sn content (Si)GeSn-alloys. *ACS Appl Mater Interfaces* 2016. <http://dx.doi.org/10.1021/acsami.6b0242>.
- [15] van der Pauw LJ. A method of measuring the resistivity and hall coefficient of discs of arbitrary shape. *Philips Res Reports* 1958;13:1-9. citeulike-article-id:8438442.
- [16] Wirths S, Troitsch R, Mussler G, Hartmann J-M, Zaumseil P, Schroeder T, et al. Ternary and quaternary Ni(Si)Ge(Sn) contact formation for highly strained Ge p- and n-MOSFETs. *Semicond Sci Technol* 2015;30:055003-1-3-8. <http://dx.doi.org/10.1088/0268-1242/30/5/055003>.
- [17] Tong Y, Han G, Liu B, Yang Y, Wang L, Wang W, et al. Ni ($\text{Ge}_{1-x}\text{Sn}_x$) Ohmic contact formation on N-type $\text{Ge}_{1-x}\text{Sn}_x$ using selenium or sulfur implant and segregation. *IEEE Trans Electron Devices* 2012;60:1-7.
- [18] Zheng J, Wang S, Zhang X, Liu Z, Xue C, Li C. Ni($\text{Ge}_{1-x-y}\text{Si}_x\text{Sn}_y$) Ohmic contact formation on p-type $\text{Ge}_{0.86}\text{Si}_{0.07}\text{Sn}_{0.07}$. *IEEE Electron Device Lett* 2015;36:878-80. <http://dx.doi.org/10.1109/LED.2015.2459062>.
- [19] Mueller M, Zhao QT, Urban C, Sandow C, Buca D, Lenk S, et al. Schottky-barrier height tuning of NiGe/n-Ge contacts using As and P segregation. *Mater Sci Eng, B* 2008;154-155:168-71. <http://dx.doi.org/10.1016/j.mseb.2008.09.037>.



Study of silicon n- and p-FET SOI nanowires concerning analog performance down to 100 K



Bruna Cardoso Paz^{a,*}, Mikaël Cassé^b, Sylvain Barraud^b, Gilles Reimbold^b, Maud Vinet^b, Olivier Faynot^b, Marcelo Antonio Pavanello^a

^a Department of Electrical Engineering, Centro Universitário da FEI, São Bernardo do Campo, Brazil

^b Département des Composants Silicium – SCME/LCTE, CEA-LETI Minattec, Grenoble, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Analog performance
Nanowires
Fin width influence
Temperature influence
Mobility

ABSTRACT

This work presents an analysis of the performance of silicon triple gate SOI nanowires aiming the investigation of analog parameters for both long and short channel n-type and p-MOSFETs. Several nanowires with fin width as narrow as 9.5 nm up to quasi-planar MOSFETs 10 μm -wide are analyzed. The fin width influence on the analog parameters is studied for n-type and p-MOSFETs with channel lengths of 10 μm and 40 nm, at room temperature. The temperature influence is analyzed on the analog performance down to 100 K for long channel n-MOSFETs by comparing the quasi-planar device to the nanowire with fin width of 14.5 nm. The intrinsic voltage gain, transconductance and output conductance are the most important figures of merit in this work. An explicit correlation between these figures of merit and the mobility behavior with temperature is demonstrated.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Scaling plays an important role in the continuity of the CMOS roadmap. The reduction of the transistors dimensions brought new challenges to the semiconductor industry and the scientific community. Problems related to the loss of control of the channel charges appeared due to scaling and revealed to be crucial on determining the performance degradation of the devices. These undesirable effects known as short channel effects motivated the development of new architectures and technologies. Multiple gate field effect transistors have been proposed to overcome such problems and push the limits of scaling beyond those imposed by conventional planar devices [1].

Tridimensional structures characterized by semiconductor stripes surrounded by the gate have shown good electrostatic coupling and strong short channel effects immunity. By reducing the dimension of these stripes to a few nanometers, a new generation of devices called nanowires (NWs) has been evolved. Recently, many works have reported the excellent performance of NWs [2–4], showing their advantages for digital applications, associating them to others technologies such as strain and investigating different fabrication processes to enhance their performance. Moreover, other published works focus on the investigation of the transport

characteristics of NWs and their mobility dependence on temperature and semiconductor materials [5–7]. Concerning analog applications, NWs performance are still nearly unexplored, not much attention has been paid and the first work has been recently published [8]. The importance on studying the analog performance of NWs remains not only on their application in analog circuits, but also on their integration in mixed analog-digital circuits such as comparators and converters.

In this work we present an experimental study of analog properties for silicon Silicon-On-Insulator (SOI) NWs, analyzing both n-type and p-MOSFETs with channel lengths (L) of 10 μm and 40 nm and fin width (W_{FIN}) varying from 9.5 nm to 10 μm , at room temperature. For the n-MOSFETs, the quasi-planar device ($W_{\text{FIN}} = 10 - \mu\text{m}$) is compared to the NW with $W_{\text{FIN}} = 14.5$ nm, from 300 K down to 100 K. An explicit correlation between the analog performance of Si NWs and their mobility behavior is presented for the first time, to the best of our knowledge. Moreover, it is the first time where the analog parameters of such ultimate devices are explored at cryogenic temperatures. Numerical simulations are also presented in order to explain the similarity observed for the intrinsic voltage gain (A_v) for both n-type and p-MOSFETs varying W_{FIN} . Transconductance (g_m), output conductance (g_D), transconductance over drain current ratio (g_m/I_{DS}) and Early voltage (V_{EA}) are important figures of merit studied in this work. Transport characteristics are also investigated as a function of temperature and W_{FIN} to justify the results observed for g_m and g_D .

* Corresponding author.

E-mail address: bcpaz@fei.edu.br (B.C. Paz).

2. Devices and measurements

The measured transistors are silicon [1 1 0]-oriented triple gate NWs fabricated at CEA-Leti using SOI substrate and multi finger structures with 50 fins in parallel. The gate stack is composed by HfSiON/TiN (EOT = 1.4 nm), silicon thickness (H_{FIN}) is 11 nm and buried oxide is 145 nm thick. Details about the fabrication of the transistors can be found in [2,9]. Fig. 1 presents a cross-section TEM image (a) and a schematic (b) of the studied Si SOI NW MOSFET.

The measured curves of drain current (I_{DS}) normalized by effective channel width ($W_{ef} = 2H_{FIN} + W_{FIN}$) as a function of gate voltage (V_{GS}) with drain voltage (V_{DS}) of 40 mV and 0.9 V for long channel n-MOS NW ($W_{FIN} = 14.5$ nm) and quasi-planar devices ($W_{FIN} = 10 \mu\text{m}$) from 300 K down to 100 K are shown in Fig. 2. As indicated in Fig. 3, the devices present the subthreshold slope (S) near the theoretical limit of $k \times T \times \ln(10)/q$ [10] and the threshold voltage variation with temperature ($\Delta V_{TH}/\Delta T$) is equal to -0.74 and -0.56 mV/K, for quasi-planar and NW n-MOS, respectively. The V_{TH} results of devices in this work are similar temperature-dependent comparing to FinFETs in literature [11], where $\Delta V_{TH}/\Delta T = -0.58$ and -0.70 mV/K for $W_{FIN} = 20$ and 370 nm, respectively. The threshold voltage has been extracted by the second derivative method [12] and the subthreshold slope by $\partial V_{GS}/\partial (\log I_{DS})$ versus V_{GS} curves [13].

Fig. 4 shows $|I_{DS}| \times L$ for n-type and p-MOS with L of $10 \mu\text{m}$ and 40 nm, W_{FIN} of 15 nm and 25 nm, at room temperature and $|V_{DS}| = 40$ mV. Devices with $L = 40$ nm and $W_{FIN} = 15$ nm present S slightly degraded around 70 mV/dec for both n- and p-MOS, while for $W_{FIN} = 25$ nm S is 95 mV/dec and 83 mV/dec, for n- and p-MOS, respectively. As W_{FIN} increases, the short channel effects increase as well, due to poor electrostatic control of the channel region. The short channel devices present lower $|I_{DS}| \times L$ than long channel NWs as V_{GS} increases, indicating mobility degradation and stronger influence of series resistance effects.

3. Results and discussion

3.1. Channel width influence – long channel n- and p-MOS

Fig. 5 presents the transconductance normalized by the effective channel width (g_m/W_{ef}) as a function of W_{FIN} , for both n- and p-MOS with $L = 10 \mu\text{m}$, at 300 K. Devices were biased in saturation with $|V_{DS}| = 0.9$ V and gate voltage overdrive ($V_{GT} = V_{GS} - V_{TH}$) of 0 , 200 and 400 mV. The maximum transconductance is also presented in the same figure.

From Fig. 5, it is observed that g_m/W_{ef} decreases (degrades) with W_{FIN} reduction for the n-MOS and increases (improves) for the p-MOS. Besides, the n-type devices show higher g_m/W_{ef} values in comparison to p-MOSFETs. The transconductance behavior is strongly related to the effective carriers' mobility (μ_{eff}), where the plan related to the fin height (1 1 0)/[1 1 0] favors the holes

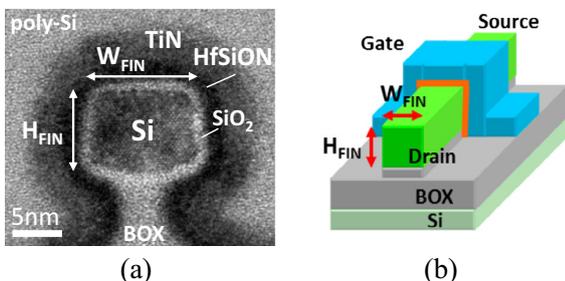


Fig. 1. Si SOI NW MOSFET cross section TEM image (a) and schematic (b).

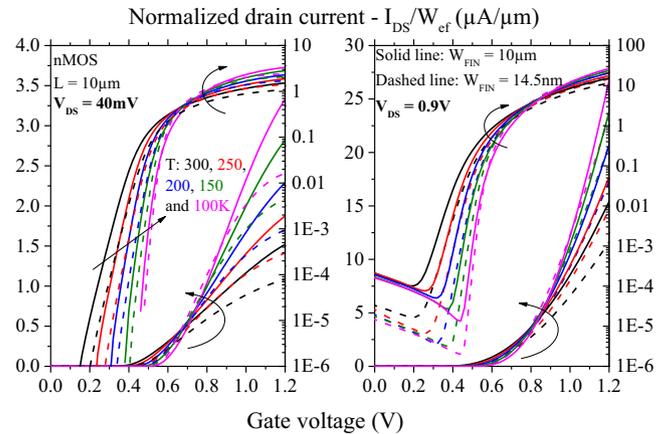


Fig. 2. I_{DS}/W_{ef} as a function of V_{GS} for long channel n-MOS, from 300 K to 100 K, at $V_{DS} = 40$ mV and 0.9 V.

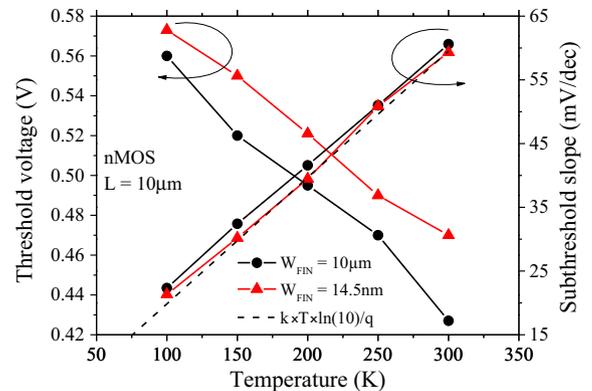


Fig. 3. Extracted threshold voltage and subthreshold slope as a function of T for n-MOS with $L = 10 \mu\text{m}$.

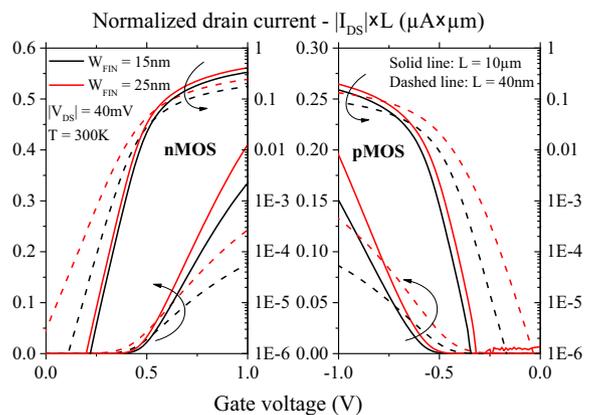


Fig. 4. $|I_{DS}| \times L$ as a function of V_{GS} for n-type and p-MOS at $|V_{DS}| = 40$ mV and 300 K.

mobility and the plan related to the fin width (1 0 0)/[1 1 0] favors the electrons mobility [10]. In Fig. 5, for V_{GT} equal to 400 mV, g_m/W_{ef} degrades 19% comparing the quasi-planar and the narrowest n-MOS, while for p-MOSFETs an improvement of 52% is observed.

Fig. 6 shows μ_{eff} as a function of V_{GT} , extracted through the split C-V method [14], for the n-MOSFETs of Fig. 5. It is indicated the main scattering contributions, Coulomb, phonon and surface roughness and it is observed that Coulomb scattering is dominant

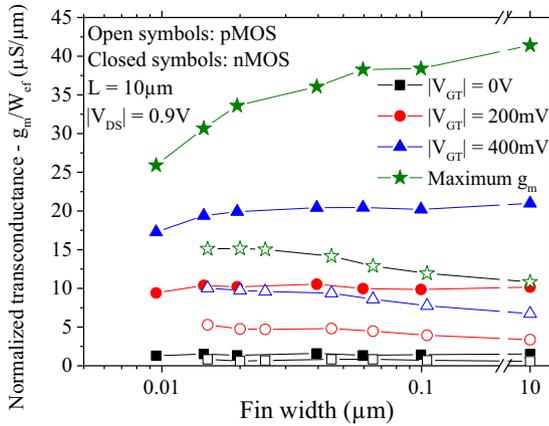


Fig. 5. g_m/W_{ef} as a function of W_{FIN} for n- and p-MOS, $L = 10 \mu\text{m}$, $|V_{DS}| = 0.9 \text{ V}$, $|V_{GT}| = 0, 200$ and 400 mV and maximum g_m .

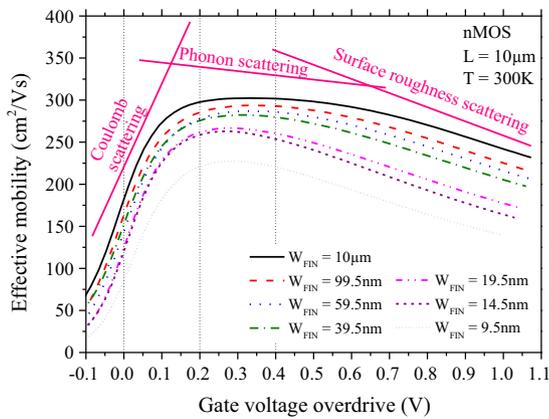


Fig. 6. μ_{eff} as a function of V_{GT} for n-MOS with $L = 10 \mu\text{m}$.

at $V_{GT} = 0 \text{ V}$, while at moderate inversion, V_{GT} equal to 200 mV and 400 mV, phonon scattering determines the effective mobility behavior. Analyzing μ_{eff} as a function of W_{FIN} , extracted for the n-MOSFETs at $V_{GT} = 0, 200$ and 400 mV , in Fig. 7, it is observed that as the sidewall surfaces (1 1 0) contribution becomes more important to the effective mobility, which means reducing W_{FIN} , the μ_{eff} degrades 26% at $V_{GT} = 400 \text{ mV}$, which is close to the g_m/W_{ef} degradation obtained in Fig. 5.

Fig. 8 presents the output conductance normalized by the effective channel width (g_D/W_{ef}) as a function of W_{FIN} , for both n- and p-MOS with $L = 10 \mu\text{m}$, at 300 K, $|V_{DS}| = 0.9 \text{ V}$ and $|V_{GT}| = 0,$

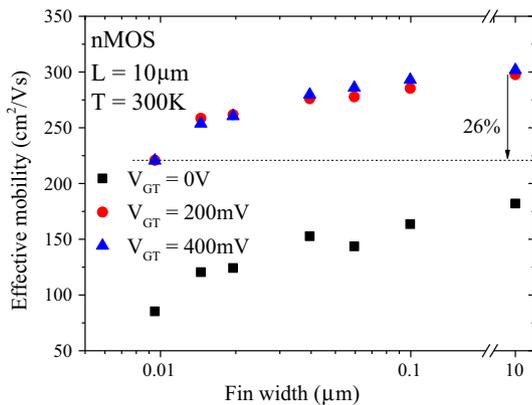


Fig. 7. μ_{eff} as a function of W_{FIN} for n-MOS, $L = 10 \mu\text{m}$, $V_{GT} = 0, 200$ and 400 mV .

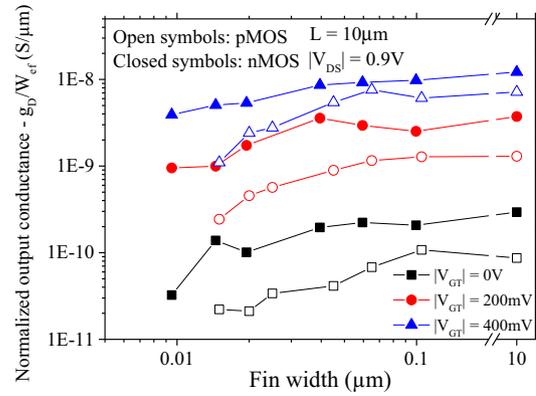


Fig. 8. g_D/W_{ef} as a function of W_{FIN} for n- and p-MOS, $L = 10 \mu\text{m}$, $|V_{DS}| = 0.9 \text{ V}$, $|V_{GT}| = 0, 200$ and 400 mV .

200 and 400 mV. For both n- and p-MOS, g_D/W_{ef} decreases (improves) as W_{FIN} is reduced in almost one order of magnitude, comparing the narrowest and the widest devices. The output conductance relates I_{DS} and V_{DS} ($g_D = \delta I_{DS}/\delta V_{DS}$) and its improvement is expected in NWs due to their better electrostatic control and reduced channel modulation effect (CME) [15].

Fig. 9 shows the results for the calculation of A_V , which is given by the g_m/g_D ratio, as a function of W_{FIN} . For both n- and p-MOS, it is observed that A_V follows the same trend of the inverse of g_D/W_{ef} , which reduction decreasing W_{FIN} is responsible for increasing (improving) significantly the intrinsic voltage gain of NWs in comparison to wide MOSFETs. This strong improvement of A_V indicates an important advantage of such devices concerning their analog performance.

From Fig. 9, it is also observed that the intrinsic voltage gain is higher at lower V_{GT} , which occurs because g_m and g_D have different dependences on V_{GT} . Fig. 10 presents both g_m/W_{ef} and g_D/W_{ef} as a function of V_{GT} for quasi-planar and narrow n-type and p-MOSFETs. It is clear that both g_m and g_D increase with V_{GT} rise. However, it is possible to observe that g_D degradation is stronger than g_m improvement, which causes A_V reduction varying V_{GT} from 0 V to 400 mV.

3.2. Channel width influence – short channel n- and p-MOS

Fig. 11 presents g_m/W_{ef} as a function of W_{FIN} , for both n- and p-MOS with $L = 40 \text{ nm}$, at 300 K, $|V_{DS}| = 0.9 \text{ V}$ and $V_{GT} = 0, 200$ and 400 mV . For short channel devices, the transconductance behavior is not only determined by the effective mobility, once

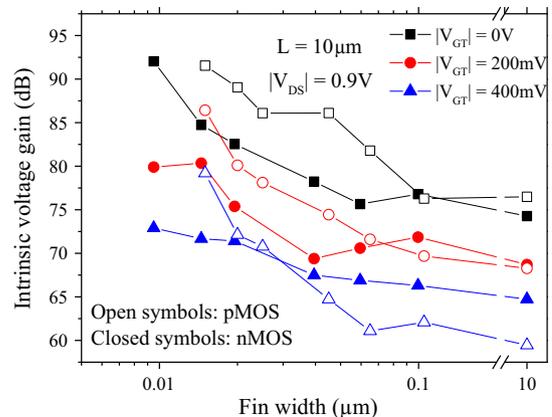


Fig. 9. A_V as a function of W_{FIN} for n- and p-MOS, $L = 10 \mu\text{m}$, $|V_{DS}| = 0.9 \text{ V}$, $|V_{GT}| = 0, 200 \text{ mV}$ and 400 mV .

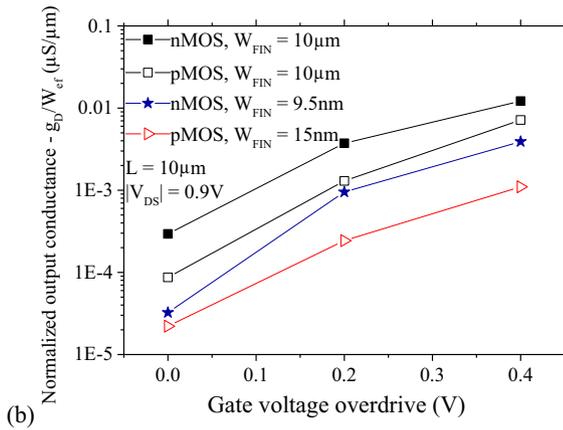
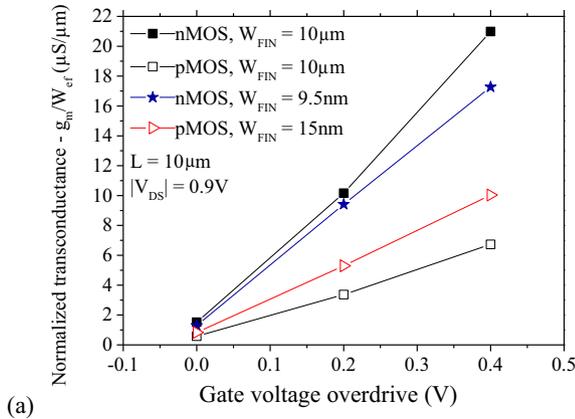


Fig. 10. g_m/W_{ef} (a) and g_D/W_{ef} (b) as a function of V_{GT} for quasi-planar and narrow n- and p-MOS, $L = 10 \mu\text{m}$ and $|V_{DS}| = 0.9 \text{ V}$.

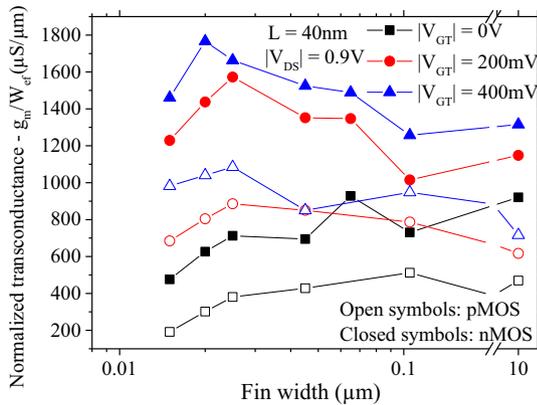


Fig. 11. g_m/W_{ef} as a function of W_{FIN} for n- and p-MOS, $L = 40 \text{ nm}$, $|V_{DS}| = 0.9 \text{ V}$, $|V_{GT}| = 0, 200$ and 400 mV .

short channel effects (SCE) and series resistance play an important role. For example, for the n-MOSFETs, besides μ_{eff} decreases with W_{FIN} reduction, causing a degradation of g_m/W_{ef} , a strong opposite behavior is expected to predominate due to improved SCE in NWs, as demonstrated in [8]. Moreover, as W_{FIN} reduces the series resistance (R_s) effects increase, leading to the g_m/W_{ef} degradation observed for the W_{FIN} narrower than 25 nm.

In order to compare the results of g_m for both long and short channel nanowires, Fig. 12 presents g_m normalized by W_{ef}/L , for both n-type and p-MOSFETs with $L = 10 \mu\text{m}$ and 40 nm and $|V_{GT}| = 400 \text{ mV}$. It is observed that long channel devices present higher

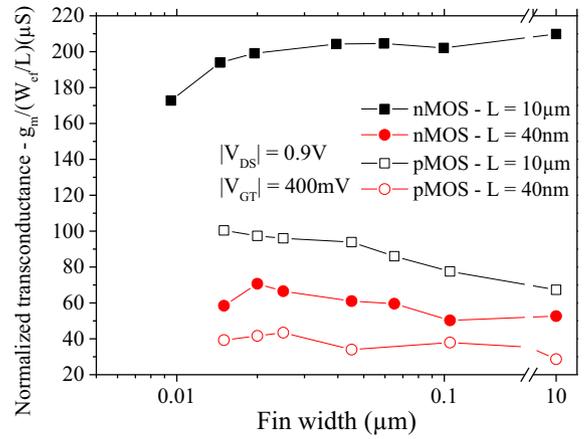


Fig. 12. $g_m/(W_{ef}/L)$ as a function of W_{FIN} for n- and p-MOS, $L = 40 \text{ nm}$ and $10 \mu\text{m}$, $|V_{DS}| = 0.9 \text{ V}$ and $|V_{GT}| = 400 \text{ mV}$.

$g_m/(W_{ef}/L)$ in comparison to short channel devices, because of strong mobility degradation with channel length reduction. From L of $10 \mu\text{m}$ to 40 nm, there is a decrease of 75% and 57%, for quasi-planar n-type and p-MOS, respectively. This percentage remains similar for narrow transistors, for W_{FIN} around 14.5 nm, $g_m/(W_{ef}/L)$ decreases 70% and 60%, for n-type and p-MOS, respectively.

Fig. 13 presents g_D/W_{ef} as a function of W_{FIN} , for both n- and p-MOS with $L = 40 \text{ nm}$, at 300 K, $|V_{DS}| = 0.9 \text{ V}$ and $V_{GT} = 0, 200$ and 400 mV . As observed for the long channel transistors in Fig. 8, g_D/W_{ef} also improves for both n- and p-MOS by decreasing W_{FIN} due to CME reduction. Once the effective channel length reduction in saturation is more significant as the transistor becomes shorter, better CME control is more important in short channel transistors.

From the results of g_m and g_D , Fig. 14 shows A_V as a function of W_{FIN} for devices with $L = 40 \text{ nm}$. The inverse of the output conductance defines the trend of A_V varying W_{FIN} , what leads to the great improve of 20 dB for the narrowest NWs comparing to the quasi-planar MOSFET for both n- and p-MOS. An interesting observation is that both n- and p-MOSFETs show very similar values for A_V and its behavior with respect to W_{FIN} , despite the fact that they present different values of g_m/W_{ef} and g_D/W_{ef} . Higher values for g_m/W_{ef} and g_D/W_{ef} are observed for the n-type devices in comparison to the p-MOSFETs due to higher μ_{eff} . As the effective mobility may affect similarly g_m and g_D , its influence may disappear by taking the g_m/g_D ratio, which emphasizes that channel modulation is the

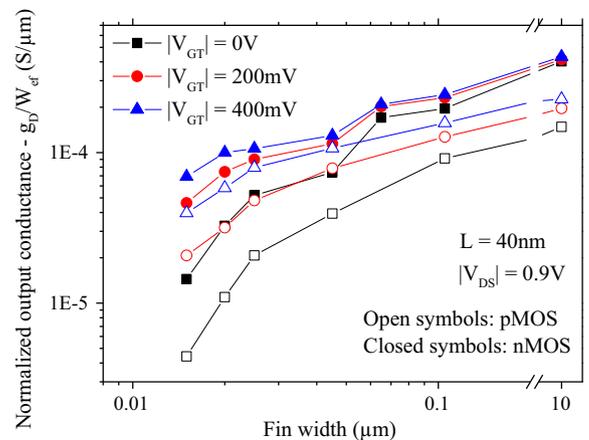


Fig. 13. g_D/W_{ef} as a function of W_{FIN} for n- and p-MOS, $L = 40 \text{ nm}$, $|V_{DS}| = 0.9 \text{ V}$, $|V_{GT}| = 0, 200$ and 400 mV .

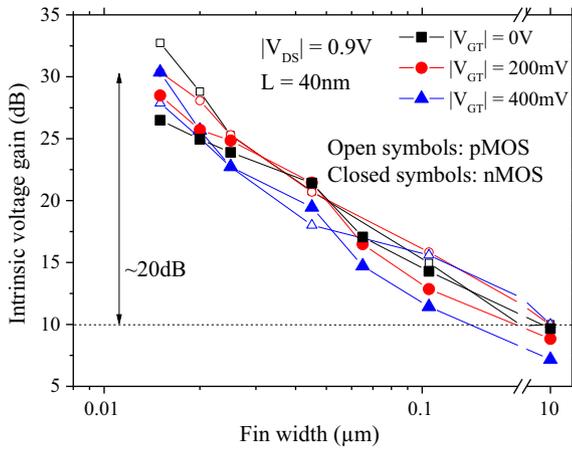
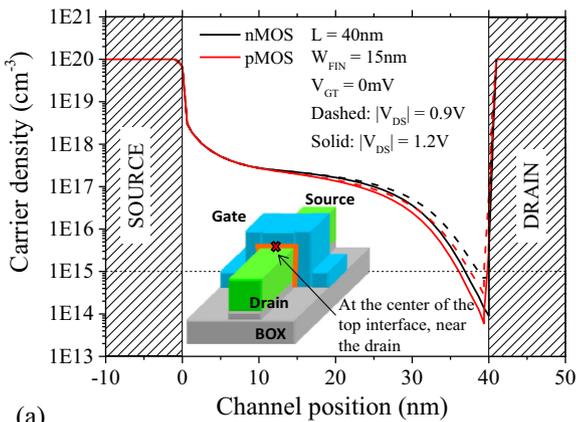


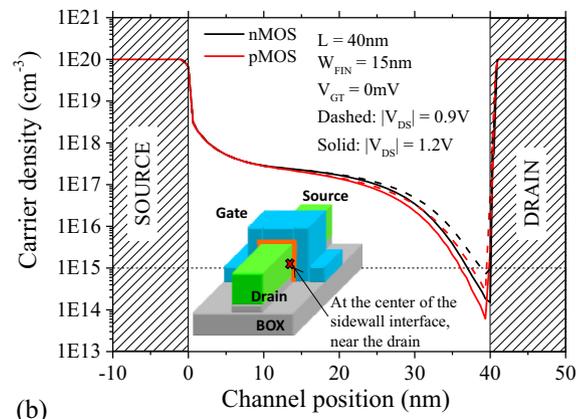
Fig. 14. A_V as a function of W_{FIN} for n- and p-MOS, $L = 40$ nm, $|V_{DS}| = 0.9$ V, $|V_{GT}| = 0$, 200 mV and 400 mV.

main effect on determining A_V , suggesting that both n- and p-MOSFETs may suffer similarly from CME.

To verify if n- and p-type devices present similar CME, tridimensional numerical simulations have been performed with Sentaurus Device Simulator, from Synopsys [16]. Fig. 15 presents simulation results for the electrons and holes density along the channel position for n- and p-MOS at 300 K, $L = 40$ nm and $W_{FIN} = 15$ nm. As indicated in the schematics in Fig. 15, carriers'



(a)



(b)

Fig. 15. Simulation results for the carrier density at the top interface (a) and the side interface (b) along the channel position for n- and p-MOS, $L = 40$ nm, $W_{FIN} = 15$ nm, $|V_{GT}| = 0$ V, $|V_{DS}| = 0.9$ and 1.2 V.

density has been extracted at the top (a) and sidewall (b) SiO_2/Si interfaces, where the conduction happens, near the drain, where depletion is induced by drain voltage for the device operating in saturation. Bias condition is $|V_{GT}| = 0$ V and $|V_{DS}| = 0.9$ and 1.2 V, where the devices operate in saturation regime. A dotted line indicates the channel doping concentration ($1 \times 10^{15} \text{ cm}^{-3}$). When the carrier density is higher than $1 \times 10^{15} \text{ cm}^{-3}$, an inversion layer is observed, otherwise the region is depleted, as it is possible to note near the drain. The CME can be seen in Fig. 15 by the longer depletion region as the $|V_{DS}|$ is increased. For both applied V_{DS} and interfaces analyzed, the depletion region is very similar comparing n- and p-MOSFETs, being the effective channel length of the p-MOS only 1 nm shorter than for the n-MOS. These close results for the carriers' density sustain that the n- and p-MOS suffer from similar CME, which explains the close g_D and, as a consequence, A_V results.

3.3. Temperature influence

Fig. 16 shows g_m/I_{DS} presented as a function of $I_{DS}/(W_{ef}/L)$ for n-MOS with $L = 10 \mu\text{m}$, at $V_{DS} = 0.9$ V. Quasi-planar and NW FETs are compared from 300 K to 100 K. Reducing either T or W_{FIN} , g_m/I_{DS} increases. The maximum value of g_m/I_{DS} is determined in weak inversion by (1), where n is the body factor, q is the electron charge and k is the Boltzmann constant [10]. Although g_m/I_{DS} in weak inversion must be higher for NWs in comparison to quasi-planar MOSFETs, once the better the electrostatic control the closer n is

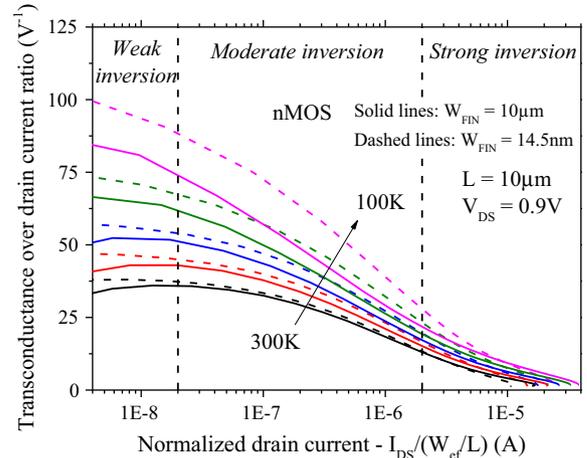


Fig. 16. g_m/I_{DS} as a function of $I_{DS}/(W_{ef}/L)$ for n-MOS, $L = 10 \mu\text{m}$ and $V_{DS} = 0.9$ V.

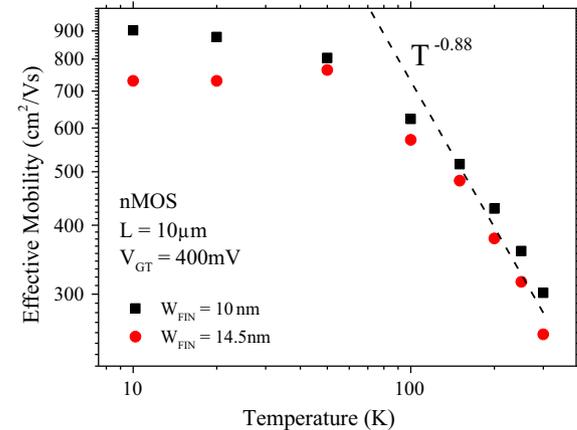


Fig. 17. μ_{eff} as a function of T for n-MOS, $L = 10 \mu\text{m}$, at $V_{GT} = 400$ mV.

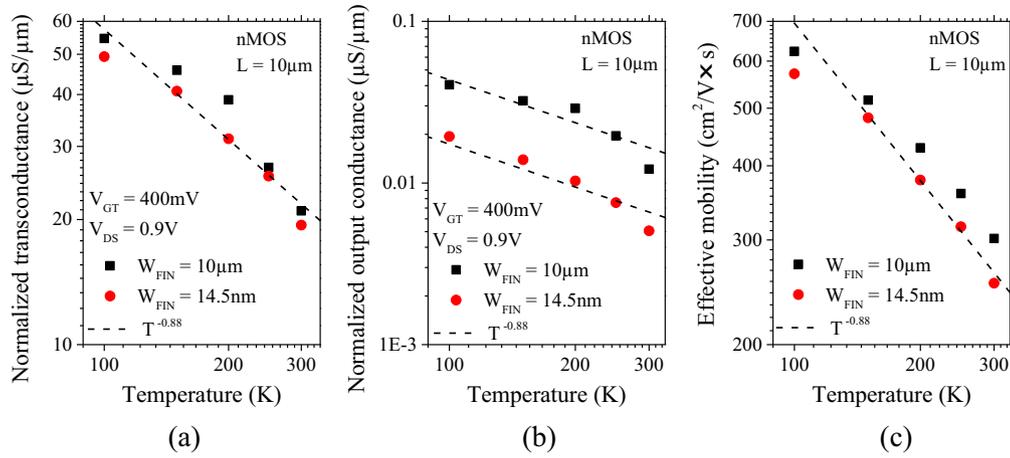


Fig. 18. g_m/W_{ef} (a), g_D/W_{ef} (b) and μ_{eff} (c) as a function of T for n-MOS, $L = 10 \mu\text{m}$, at $V_{DS} = 0.9 \text{ V}$ and $V_{GT} = 400 \text{ mV}$.

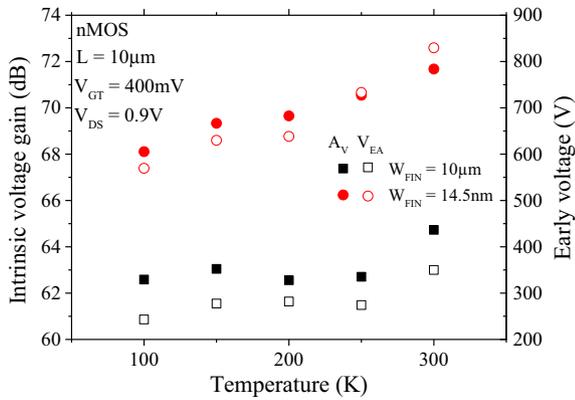


Fig. 19. A_V and V_{EA} as a function of T for n-MOS, $L = 10 \mu\text{m}$, at $V_{DS} = 0.9 \text{ V}$ and $V_{GT} = 400 \text{ mV}$.

to unity, Fig. 16 shows higher differences than expected and strong degradation for $W_{FIN} = 10 \mu\text{m}$, especially at low temperatures. This effect is related to the current originated from tunneling in the drain region under the gate, known as gate-induced-drain-leakage (GIDL) current [17]. As observed in Fig. 2 at $V_{DS} = 0.9 \text{ V}$, GIDL current is one order of magnitude higher for $W_{FIN} = 10 \mu\text{m}$ than 14.5 nm and appears at higher values of V_{GS} decreasing T , which explains the degradations in the g_m/I_{DS} curves. Stronger immunity to GIDL effect allows NWs to present maximum g_m/I_{DS} close to ideal values, even at $T = 100 \text{ K}$.

$$\left(\frac{g_m}{I_{DS}}\right) = \frac{q}{n \times k \times T} \quad (1)$$

According to [18], at saturation, g_m/I_{DS} can be estimated by (2), where R_S , μ_{eff} and W_{ef} are the parameters that depend on W_{FIN} . Extracting R_S according to [19], an increase of a factor close to 10 is obtained from W_{FIN} of $10 \mu\text{m}$ to 14.5 nm , while μ_{eff} decreases of a factor close to 1.5. As W_{ef} reduces almost 3 orders of magnitude, it is found to be the key parameter in (2), where g_m/I_{DS} varies with the inverse of $R_S \times \mu_{eff} \times W_{ef}$, explaining higher g_m/I_{DS} for the NW than the quasi-planar FET.

$$\left(\frac{g_m}{I_{DS}}\right)_{sat} = \frac{2}{V_{GT}} \times \left[\frac{1 + \frac{(R_S \mu_{eff} C_{ox} W_{ef} V_{GT})}{2L}}{1 + \frac{(R_S \mu_{eff} C_{ox} W_{ef} V_{GT})}{L}} \right] \quad (2)$$

Fig. 17 presents μ_{eff} as a function of T , for long channel n-MOS with W_{FIN} of $10 \mu\text{m}$ and 14.5 nm , extracted at $V_{GT} = 400 \text{ mV}$ through split C-V [14]. As previously discussed in Fig. 6, at

$V_{GT} = 400 \text{ mV}$, phonon scattering has the dominant contribution, which can be confirmed by analyzing the temperature influence on the effective mobility. Above 100 K , μ_{eff} shows the phonon mobility dependence with temperature, $\mu_{ph} \propto T^{-\gamma}$ [20], where the temperature dependence coefficient (γ) is found to be close to 0.88 for both NW and quasi-planar FETs. Below 100 K , surface roughness limited contribution becomes dominant, phonon scattering is negligible and, as consequence, μ_{eff} saturates [21].

Fig. 18 presents g_m/W_{ef} (a), g_D/W_{ef} (b) and μ_{eff} (c) as a function of T , for long channel n-MOS. Both g_m/W_{ef} and g_D/W_{ef} have been extracted at $V_{DS} = 0.9 \text{ V}$ and $V_{GT} = 400 \text{ mV}$. The dashed line indicates the same temperature dependence coefficient ($\propto T^{-0.88}$). Reducing T , it is observed that g_m/W_{ef} and g_D/W_{ef} increases following similar dependence in relation to μ_{eff} behavior with temperature. For the transconductance, results show clearly that its variation with temperature is mainly due to mobility. For the output conductance, the rate of increase with temperature reduction is slightly higher than what is observed for μ_{eff} .

Fig. 19 shows A_V and V_{EA} as a function of T . As expected by the results of g_m and g_D , the temperature effect is almost annulled one by each other, leading to a smooth decrease with temperature reduction. Once g_D increases with a higher rate than g_m , decreasing T from 300 K to 100 K , A_V shows a smooth decrease of 3.6 dB and 2.1 dB for W_{FIN} of 14.5 nm and $10 \mu\text{m}$, respectively. As V_{EA} follows the inverse of g_D as well, once $V_{EA} = I_{DS}/g_D$, much higher (more than 2.3 times) values are obtained for the nanowire, in comparison to $W_{FIN} = 10 \mu\text{m}$, due to improved CME control.

4. Conclusions

Results of analog parameters of scaled silicon nanowires SOI MOSFETs have been presented for both long and short channel n- and p-type transistors. Temperature influence has been studied down to 100 K for long channel n-MOSFETs.

An explicit dependence on the effective mobility has been presented for transconductance of long channel devices varying fin width. The reduced output conductance of nanowires due to channel modulation effect control determines strong improvements on the intrinsic voltage gain. Similar channel modulation effects for both short channel n-type and p-MOSFETs has been demonstrated, leading to similar intrinsic voltage gain results varying fin width.

Temperature influence is mainly determined by the phonon mobility dependence and its effect is almost annulled by taking g_m/g_D , giving soft increase of intrinsic voltage gain with temperature increase.

The nanowires have shown great improvements concerning analog performance, due to short channel effects control, reduced channel modulation effect and better body factor. Even n-type MOSFETs presenting lower effective mobility than quasi-planar transistors, both n-type and p-NWs presented 20 dB of increase for intrinsic voltage gain in devices with $L = 40$ nm and $W_{\text{FIN}} = 15$ nm, at room temperature.

Acknowledgements

The authors would like to acknowledge the IBM/STMicroelectronics/Leti Joint Development Alliance, CNPq, CAPES and FAPESP process 2015/10491-7 for the financial support to this work.

References

- [1] Colinge J-P. Multiple-gate SOI MOSFETs. *Solid-State Electron* 2004;48(6):897–905.
- [2] Coquand R, Casse M, Barraud S, Leroux P, Cooper D, Vizios C, et al. Strain-induced performance enhancement of tri-gate and omega-gate nanowire FETs scaled down to 10 nm width. *Symposium on VLSI technology (VLSIT)* 2012;2012:13–4.
- [3] Coquand R, Barraud S, Casse M, Leroux P, Vizios C, Comboroure C, et al. Scaling of high-k/metal-gate Trigate SOI nanowire transistors down to 10 nm width. In: 2012 13th International conference on ultimate integration on silicon (ULIS); 2012. p. 37–40.
- [4] Bangsaruntip S, Cohen GM, Majumdar A, Zhang Y, Engelmann SU, Fuller NCM, et al. High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling. *Electron devices meeting (IEDM), 2009 IEEE International* 2009;1–4.
- [5] Casse M, Barraud S, Le Royer C, Koyama M, Coquand R, Blachier D, et al. Study of piezoresistive properties of advanced CMOS transistors: thin film SOI, SiGe/SOI, unstrained and strained Tri-Gate nanowires. In: *Electron devices meeting (IEDM), 2012 IEEE International*; 2012. p. 28.1.1–4.
- [6] Lai W-T, Wu C-W, Lin C-C, Li P-W. Analysis of carrier transport in trigate Si nanowire MOSFETs. *IEEE Trans Electron Dev* 2011;58(5):1336–43.
- [7] Koyama M, Casse M, Coquand R, Barraud S, Iwai H, Ghibaudo G, et al. Study of carrier transport in strained and unstrained SOI tri-gate and omega-gate Si-nanowire MOSFETs. *Solid-state device research conference (ESSDERC), 2012 Proceedings of the European* 2012:73–6.
- [8] Kilchytska V, Makovejev S, Barraud S, Poiroux T, Raskin J-P, Flandre D. Trigate nanowire MOSFETs analog figures of merit. *Solid-State Electron* 2015;112:78–84.
- [9] Barraud S, Coquand R, Casse M, Koyama M, Hartmann J, Maffini-Alvaro V, et al. Performance of omega-shaped-gate silicon nanowire MOSFET with diameter down to 8 nm. *IEEE Electron Dev Lett* 2012;33(11):1526–8.
- [10] Colinge J-P. *Silicon-on-insulator technology: materials to VLSI: materials to VLSI*. Springer Science & Business Media; 2004.
- [11] Pavanello MA, Martino JA, Simoen E, Claeys C. Cryogenic operation of FinFETs aiming at analog applications. *Cryogenics* 2009;49(11):590–4.
- [12] Ortiz-Conde A, García Sánchez FJ, Liou JJ, Cerdeira A, Estrada M, Yue Y. A review of recent MOSFET threshold voltage extraction methods. *Microelectron Reliab* 2002;42(4–5):583–96.
- [13] Colinge J-P, Colinge CA. *Physics of semiconductor devices*. Springer Science & Business Media; 2005.
- [14] Ohata A, Cassé M, Cristoloveanu S. Front- and back-channel mobility in ultrathin SOI-MOSFETs by front-gate split CV method. *Solid-State Electron* 2007;51(2):245–51.
- [15] Veeraraghavan S, Fossum JG. Short-channel effects in SOI MOSFETs. *IEEE Trans Electron Dev* 1989;36(3):522–8.
- [16] *Sentaurus device user guide, Version C-2009.06*. Synopsys; 2009.
- [17] Galeti M, Rodrigues M, Martino JA, Collaert N, Simoen E, Claeys C. GIDL behavior of p- and n-MuGFET devices with different TiN metal gate thickness and high-k gate dielectrics. *Solid-State Electron* 2012;70:44–9.
- [18] Subramanian V, Mercha A, Parvais B, Loo J, Gustin C, Dehan M, et al. Impact of fin width on digital and analog performances of n-FinFETs. *Solid-State Electron* 2007;51(4):551–9.
- [19] Dixit A, Kottantharayil A, Collaert N, Goodwin M, Jurczak M, De Meyer K. Analysis of the parasitic S/D resistance in multiple-gate FETs. *IEEE Trans Electron Dev* 2005;52(6):1132–40.
- [20] Takagi S, Toriumi A, Iwase M, Tango H. On the universality of inversion layer mobility in Si MOSFETs: Part I-effects of substrate impurity concentration. *IEEE Trans Electron Dev* 1994;41(12):2357–62.
- [21] Tachi K, Casse M, Barraud S, Dupré C, Hubert A, Vulliet N, et al. Experimental study on carrier transport limiting phenomena in 10 nm width nanowire CMOS transistors. In: *Electron devices meeting (IEDM), 2010 IEEE International*; 2010. p. 34.4.1–4.



An in-depth analysis of temperature effect on DIBL in UTBB FD SOI MOSFETs based on experimental data, numerical simulations and analytical models



A.S.N. Pereira^{a,b,*}, G. de Streel^a, N. Planes^c, M. Haond^c, R. Giacomini^b, D. Flandre^a, V. Kilchytska^a

^aICTEAM, Université catholique de Louvain, Louvain-la-Neuve, Belgium

^bDepartment of Electrical Engineering, Centro Universitário da FEI, São Bernardo do Campo, Brazil

^cSTMicroelectronics, Crolles, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Fully-Depleted Silicon-On-Insulator (FD SOI)

UTBB

DIBL

Temperature dependence

ABSTRACT

The Drain Induced Barrier Lowering (DIBL) behavior in Ultra-Thin Body and Buried oxide (UTBB) transistors is investigated in details in the temperature range up to 150 °C, for the first time to the best of our knowledge. The analysis is based on experimental data, physical device simulation, compact model (SPICE) simulation and previously published models. Contrary to MASTAR prediction, experiments reveal DIBL increase with temperature. Physical device simulations of different thin-film fully-depleted (FD) devices outline the generality of such behavior. SPICE simulations, with UTSOI DK2.4 model, only partially adhere to experimental trends. Several analytic models available in the literature are assessed for DIBL vs. temperature prediction. Although being the closest to experiments, Fasarakis' model overestimates DIBL(T) dependence for shortest devices and underestimates it for upsized gate lengths frequently used in ultra-low-voltage (ULV) applications. This model is improved in our work, by introducing a temperature-dependent inversion charge at threshold. The improved model shows very good agreement with experimental data, with high gain in precision for the gate lengths under test.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Driven by the continuous demand for the increase of the operation speed and the integration density of complex digital circuits, semiconductor technology continues its progress in device scaling down to 20 nm and beyond. Drain Induced Barrier Lowering (DIBL) is one of the measures that allows for assessing electrostatic integrity of the device and its capability for scaling down. Strong criterion on DIBL is imposed with typically considered values of DIBL lower than 100 mV/V [1,2]. The Ultra-Thin Body and Buried oxide (BOX) (so-called UTBB) Fully Depleted (FD) Silicon on Insulator (SOI) technology is widely considered as one of the candidates able to satisfy this criterion featuring excellent electrostatic integrity. This is achieved thanks to the thin body and thus good control of the channel from the gate; thin BOX and thus suppression/reduction of fringing electric field lines penetration in BOX; introduction of Ground Plane (GP), or highly doped region just underneath the BOX and thus suppression of the coupling through the substrate

[2–4]. It was demonstrated that the scalability of UTBB devices can reach the 8 nm node if the silicon film and the BOX thicknesses are scaled down to 5 nm and 10 nm, respectively [2]. However, DIBL may vary with temperature and hence devices that fit the requirements at room temperature, may become out of specifications at higher temperatures.

Studying the temperature effects (from room temperature up to 150 °C) is important for two reasons. Firstly, it is, evidently, crucial for high-temperature applications. With deeper electronics penetration in everyday life (automotive, health/medical, smart devices), such usually considered niche-applications start to go out from a niche and get more attention. Secondly, good understanding of high-temperature effects are important even for room-temperature (RT) applications, as these devices can be affected by self-heating with channel temperature reaching ~100 °C under normal operation conditions [5–7].

Some results/data on DIBL variation from room to high temperature available in the literature shows that the DIBL of 0.4–1.5 μm-long bulk Si MOSFETs increase with temperature specially for shorter devices [8,9]. For 28 nm long UTBB devices [10], an increase of DIBL by about 20 mV/V over 100 °C was observed when

* Corresponding author at: Department of Electrical Engineering, Centro Universitário da FEI, São Bernardo do Campo, Brazil.

E-mail address: ariannesoes@fei.edu.br (A.S.N. Pereira).

compared to RT. However, in-depth analysis of DIBL evolution with temperature in such devices was not performed before.

The present work provides a detailed analysis of DIBL behavior at high temperatures (up to 150 °C) in UTBB FDSOI MOSFETs based on experimental data, physical device simulations (Atlas), compact modelling (SPICE) and several previously published analytical models. The final aim is a revision of previously published model (s) to be applicable for DIBL prediction in a wide temperature range. This paper is an extended version of the work/abstract presented at EuroSOI/ULIS conference 2016 [11].

2. Experimental details

The experimental devices were fabricated at ST Microelectronics. More details can be found e.g. in [12]. Devices feature a BOX thickness (t_{BOX}) of 25 nm and silicon body (t_{Si}) of 7 nm. The metal gate stack is composed of 2.3 nm of HfSiON with an equivalent oxide thickness (t_{OX}) of 1.3 nm. The measured devices are n-channel MOSFETs with gate lengths (L) from 34 to 500 nm and channel width (W) of 1 μm . The channel is left undoped. Both standard V_T (STDVT) and low V_T (LVT) devices were characterized up to 150 °C for DIBL evaluation.

Drain current versus gate voltage, I_D - V_G characteristics were measured in linear and saturation regimes. The DIBL values were calculated from $\Delta V_G/\Delta V_D$, where V_G is the gate voltage at the constant current value of 10^{-7} (W/L) for low drain voltage (V_{DL}) of 50 mV or high drain voltage (V_{DH}) of 1 V.

3. Results and discussion

3.1. Experiments

Fig. 1 shows the experimental DIBL behavior as a function of temperature for the devices with different gate lengths. The

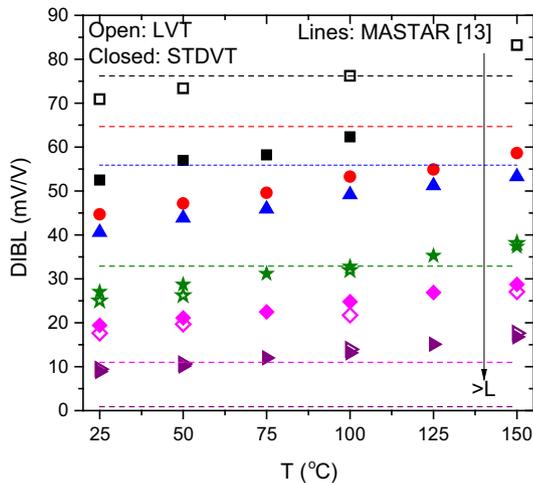


Fig. 1. Measured DIBL as a function of temperature (symbols) and MASTAR modeling (lines) for STDVT and LVT of 38, 42, 60, 90, 120 and 500 nm long devices (from top to bottom).

Table 1

Different FD nMOSFET structures simulated with Atlas.

	UTBB overlap	UTBB underlap	DG	UTB	FDSOI
t_{Si} (nm)	7	7	7	7	30
t_{OX} (nm)	1.3	1.3	1.3	1.3	1.3
t_{BOX} (nm)	25	25	$t_{\text{BOX}} = t_{\text{OX}}$	150	150
L (nm)	50	50	50	50	200
L_S (nm)	0	10	10	0	0

increase of DIBL with temperature can be clearly observed for both LVT and STDVT transistors. Moreover, this increase is stronger in shorter devices. DIBL values calculated using MASTAR software [13] for the above physical device parameters are plotted on the same graph for the sake of comparison. They are seen to be independent of temperature. More details on MASTAR model temperature results will be given below, in Section 3.3.

3.2. Physical and compact model simulation

In order to verify whether some particular process features could generate the observed DIBL behavior we performed physical device simulations using Atlas software [14]. Table 1 presents the various thin-film FD nMOSFETs ($N_A = 10^{15} \text{ cm}^{-3}$) we assessed with the following objectives:

- Comparison between Underlapped and Overlapped structures was done to detect possible effect of effective channel length variation with temperature.
- Comparison between thin and thick BOX devices aimed at distinguishing possible effect of mean channel position variation with temperature.
- An ideal double gate (DG) structure was also studied to remove any possible channel depth position and substrate depletion effects.

Fig. 2 presents the DIBL for different thin-film FD structures simulated with Atlas [14]. For all simulated structures, the trend is the same and is identical to that of experimental results of Fig. 1, highlighting that this DIBL temperature dependence is physical and not specific to the measured UTBB devices or to the experimental setup. The lower DIBL values among single gate devices are obtained for UTBB with underlap, as expected.

Fig. 3 shows the DIBL variation with temperature from circuit-level SPICE simulations using UTSOI DK2.4 model from CEA-Leti

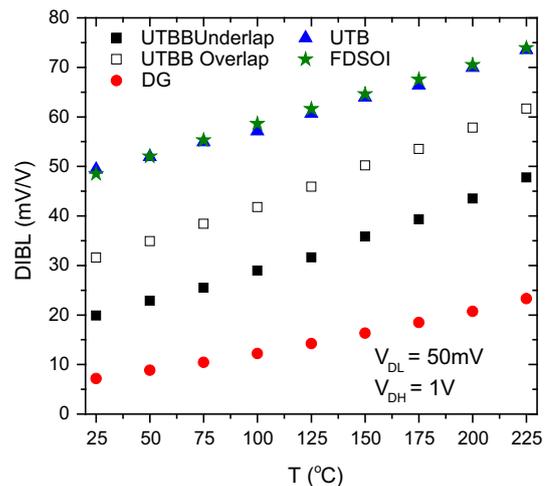


Fig. 2. DIBL as a function of temperature from Atlas simulations of different SOI structures.

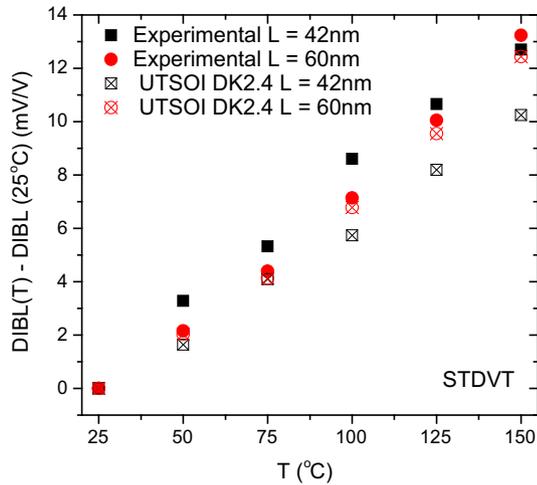


Fig. 3. DIBL increase with temperature for devices with $L = 42$ and 60 nm (STDVT) extracted from measurements and SPICE simulations.

and ST Microelectronics [15] compared to experimental data for standard V_T devices with L of 42 and 60 nm. The UTSOI DK2.4 model qualitatively reproduces the experimental DIBL dependence on temperature. However, the modeled DIBL(T) variation is qualitatively incorrect underestimating DIBL increase with temperature, particularly for shorter-length devices.

3.3. Analytic models analysis

In this section, an analysis of previously published models is performed in a view of their applicability for DIBL prediction in a wide temperature range. We assessed the ability of different analytic models to reproduce DIBL vs. temperature increase in UTBB devices. Three models are considered:

- (1) MASTAR software [13], which is based on Voltage Drop Transformation (VDT) model [16,17].
- (2) Arshad's model [18], an improved version of MASTAR model, which takes into account the effective channel length (L_{EFF}), the mean channel position in the thin film (Y_{MEAN}) and the substrate depletion.

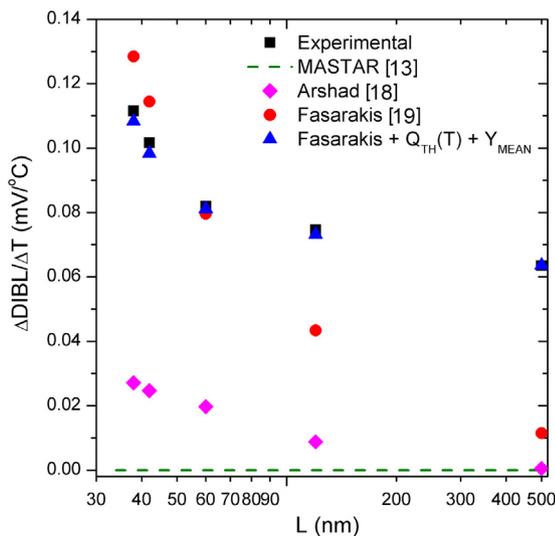


Fig. 4. Experimental and modeled $\Delta DIBL/\Delta T$ up to 150 °C as a function of channel length.

- (3) Fasarakis' model [19] which is a threshold voltage model, defining the DIBL as the difference between the threshold voltage, V_T , for low and high drain voltage. These results have been obtained as follows.

Fig. 4 shows the DIBL variation with temperature, i.e. $dDIBL/dT$ (in a T range up to 150 °C) extracted from the experiments and from different models published in literature as a function of gate length. It can be noticed that both Arshad's and Fasarakis' models describe the trends, featuring enhanced DIBL vs T increase with L shortening.

3.3.1. MASTAR model

Though temperature is an input parameter and MASTAR software [13] can calculate the drain current and SCE (Short-Channel Effects) for different temperatures, the DIBL extracted from MASTAR has no dependence on temperature (Figs. 1 and 4). This happens because the DIBL equation used by MASTAR is based on the VDT model [16,17], which takes into account just dimensional and materials variables. Originally developed for bulk devices, the approach for FDSOI transistors is presented in Eq. (1) [1,13]

$$DIBL = 0.8 \frac{\epsilon_{Si}}{\epsilon_{OX}} \left(1 + \frac{t_{Si}^2}{L^2}\right) \frac{t_{OX}}{L} \left(\frac{t_{Si}}{L} + \frac{\Lambda t_{BOX}}{L}\right) V_D \quad (1)$$

where ϵ_{Si} is the silicon permittivity, ϵ_{OX} is the oxide permittivity and Λ is the factor that represents how deep the gate field in the channel region extends in the buried oxide, its value is calculated by Eq. (2)

$$\Lambda = \frac{3(L - t_{Si})}{(L - t_{Si}) + 3t_{BOX}} \quad (2)$$

3.3.2. Arshad's model

Arshad's model [18] can be adapted for different T by considering $Y_{MEAN}(T)$ and $L_{EFF}(T)$ extracted from simulations. Substrate depletion (considered in Arshad's model) and its variation with temperature can be neglected in our case as GP provides an efficient screening. The Y_{MEAN} is extracted from the electron concentration of the simulated structure file saved at the V_G point of DIBL extraction. The expression to calculate its value is (3)

$$Y_{MEAN} = \frac{\int_0^{t_{Si}} y \cdot eC dy}{\int_0^{t_{Si}} eC dy} \quad (3)$$

where eC is the electron concentration and y is the vertical position in the silicon film. The Y_{MEAN} value is used to correct the t_{Si} and t_{OX} values, which are changed to t_{Siel} and t_{OXel} , and are given by (4) and (5)

$$t_{Siel} = t_{Si} - Y_{MEAN} \quad (4)$$

$$t_{OXel} = t_{OX} + \left(\frac{\epsilon_{Si}}{\epsilon_{OX}}\right) Y_{MEAN} \quad (5)$$

Effective length correction in Arshad's model considers the impact of the underlap gate-source/drain regions. This is made by introducing the underlap (spacer in Arshad's model) length (L_S) (accounting for both gate-to-source and gate-to-drain underlap regions) to calculate effective channel length, as presented in (6)

$$L_{EFF} = L + (2L_S) \quad (6)$$

However, those L_{EFF} from (6) have no dependence with temperature. Hence, we also extracted L_{EFF} values from simulated electron concentration at different V_G points for each temperature as described in [20]. As an example, Fig. 5 shows Y_{MEAN} and L_{EFF} variation with temperature for the devices with the gate length of 38 and 120 nm. The Y_{MEAN} shifts towards bottom interface (with

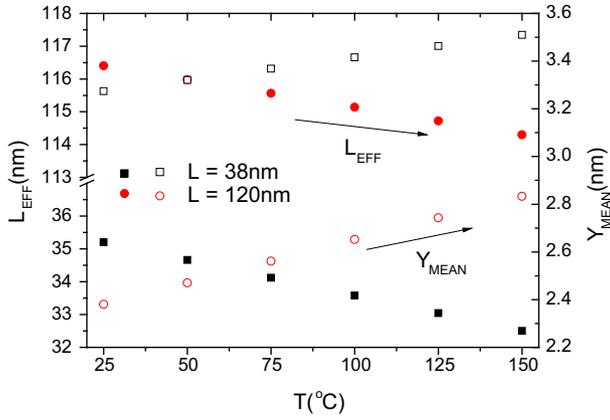


Fig. 5. Simulated Y_{MEAN} (open symbols) and L_{EFF} (closed symbols) at the V_G point of DIBL extraction and $V_D = 1$ V as a function of temperature for Arshad's model (Y_{MEAN} measures from top Si-gate oxide interface).

thicker oxide and hence higher DIBL) and effective length shortens with T increase. Both facts naturally result in DIBL increase at higher T featuring also some enhancement for the shorter lengths, thus confirming our experimental results. However, incorporation of these dependences is not at all sufficient to reproduce experimental DIBL(T) results (from Fig. 1).

3.3.3. Fasarakis' model

In Fasarakis' model [19], the V_T is quantitatively defined as the gate voltage at which the minimum carrier charge sheet density (Q_{INV}) at the effective conductive path reaches a value Q_{TH} adequate to achieve the turn-on condition, such a definition is equivalent to the constant current method of V_T extraction. The effective conductive path is defined according to [21]. The V_T expression for $V_{SUB} = 0$ is given by (7). The third term of original equation in [19] was neglected because of zero V_{SUB} condition.

$$V_T = V_{FB} + A_f V_{th} \ln\left(\frac{Q_{TH} N_A}{n_i^2 t_{Si}}\right) - B_f \left[\phi_d - V_{th} \ln\left(\frac{Q_{TH} N_A}{n_i^2 t_{Si}}\right) \right]^{\frac{1}{2}} \left[\phi_d + V_D - V_{th} \ln\left(\frac{Q_{TH} N_A}{n_i^2 t_{Si}}\right) \right]^{\frac{1}{2}} \quad (7)$$

where V_{FB} is the flat-band voltage, V_{th} is the thermal voltage (kT/q), ϕ_d is the junction built-in voltage, n_i is the intrinsic silicon concentration, A_f and B_f are dependent on dimensional parameters of the transistors. The Q_{TH} value is defined by (8)

$$Q_{TH} = 7 \cdot 10^{10} \left[1 + 8V_D \frac{\lambda}{L} + 4 \frac{t_{OX}}{t_{BOX}} \right]^{-2} \quad (8)$$

where λ is the natural length along the channel. The channel position (Y_{MEAN}) is calculated from equation

$$Y_{MEAN} = A_C \cdot t_{Si} \quad (9)$$

where A_C is a model parameter. For geometrical parameters lying within wide ranges in real device applications A_C value is 0.05 [19]. Considering A_C and t_{Si} values for our experimental devices the calculated Y_{MEAN} is 0.35 nm.

Fasarakis' model shows much better results comparing to other models (Fig. 4). However, it gives stronger dependence than experimentally observed one, particularly overestimating $\Delta DIBL/\Delta T$ values as well as absolute DIBL values (see Fig. 6) in shorter devices and underestimating them in long ones.

3.3.4. Upgrade of Fasarakis' model

We believe that the reason of such quantitative disagreement is that the temperature dependence of the channel position (Y_{MEAN})

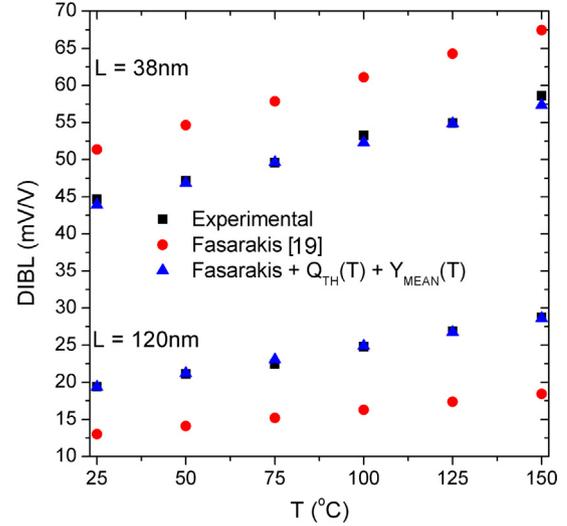


Fig. 6. DIBL as a function of temperature for the devices with $L = 38$ and 120 nm.

and the inversion charge defined at the threshold condition (Q_{TH}) in Fasarakis' model were not taken into account. Our proposal is to extract Y_{MEAN} and Q_{TH} for each temperature from simulated electron concentration at V_G points (low and high V_D) of DIBL extraction. Y_{MEAN} is extracted from Eq. (3) and Q_{TH} values are calculated by multiplying the electron concentration value at Y_{MEAN} position by the Y_{MEAN} value itself. Fig. 7 shows the Q_{TH} values as a function of temperature calculated using Fasarakis original model (Eq. (8)) and extracted from simulations. It can be observed that Fasarakis' model values are constant in all temperature range and the extracted values have temperature dependence.

Therefore, in order to upgrade Fasarakis model we propose to incorporate $Q_{TH}(T)$ dependence. We introduce the Q_{TH} values obtained above (by simulations) in Fasarakis' original model. The results are shown in Figs. 4 and 6, referred as "Fasarakis + $Q_{TH}(T) + Y_{MEAN}$ dependence". It can be observed that the proposed upgraded model reproduces well the experimental data in the whole L range under consideration. Note that the proposed improvement is the most accurate with experimental data, presenting a precision gain of e.g. 35% for $L = 120$ nm. Thus, the upgraded model can be used to estimate/predict the temperature dependence of DIBL in UTBB

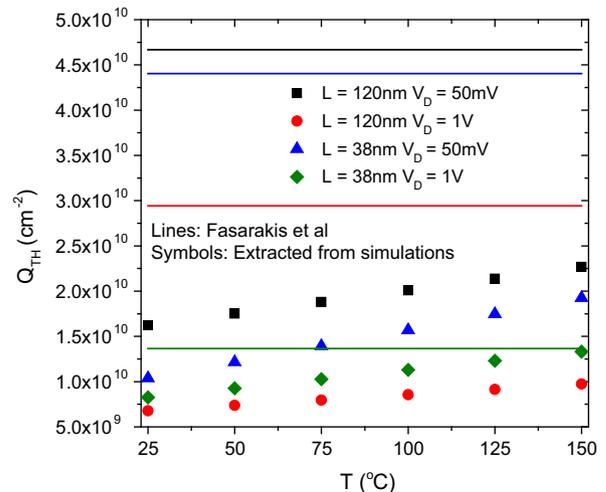


Fig. 7. Q_{TH} as a function of temperature for Fasarakis' model and extracted from simulations for $L = 38$ and 120 nm.

devices with different lengths. It is worth noting a significant improvement in the model accuracy even for longer devices, where the other models get far from experimental results. This is important as in some applications, especially ULV, the use of channel lengths above the minimal length of the technology is common to minimize current leakage [22] and to have better analog performance [23].

4. Conclusions

The DIBL dependence on temperature in UTBB FD SOI MOSFETs was analyzed. The experimental results revealed increase of DIBL with temperature. This trend is confirmed by both physical and SPICE simulations. The fact that physical device simulations of different thin-film architectures provide the same trends emphasizes generality of such DIBL temperature behavior. Existing models, however, do not allow to reproduce DIBL(T) dependence properly. We proposed a way to physically upgrade the threshold voltage model of Fasarakis et al., by including inversion charge (and accounting for channel position) dependence on temperature in order to correctly reproduce/predict DIBL variation with temperature for devices with different lengths. The obtained results show very good agreement with experimental data and significant gain of precision for both shortest devices and for longer ones with channel lengths in the range used for low leakage ULV digital or good output conductance analog applications for this technology.

Acknowledgments

A.S.N. Pereira and R. Giacomini would like to thank FAPESP (process 2014/11627-7) and CNPq for the financial support. The work is partially funded by FNRS (Belgium), Eniac Places2be and Escel WAYTOGO FAST projects.

References

- [1] Skotnicki T, Fenouillet-Beranger C, Gallon C, Boeuf F, Monfray S, Payet F, et al. Innovative materials, devices, and CMOS technologies for low-power mobile multimedia. *IEEE Trans Electron Devices* 2008;55:96–130.
- [2] Faynot O, Andrieu F, Weber O, Fenouillet-Béranger C, Perreau P, Mazurier J, et al. Planar fully depleted SOI technology: a powerful architecture for the 20 nm node and beyond. In: *IEEE international electron devices meeting (IEDM)*; 2010. p. 321–4.
- [3] Grenouillet L, Vinet M, Gimbert J, Giraud B, Noël JP, Liu Q, et al. UTBB FDSOI transistors with dual STI for a multi-Vt strategy at 20 nm node and below. In: *IEEE international electron devices meeting (IEDM)*; 2012. p. 361–4.
- [4] Liu Q, Vinet M, Gimbert J, Loubet N, Wacquez R, Grenouillet L, et al. High performance UTBB FDSOI devices featuring 20 nm gate length for 14 nm node and beyond. In: *IEEE international electron devices meeting (IEDM)*; 2013. p. 921–4.
- [5] Makovejev S, Raskin J-P, Md Arshad MK, Flandre D, Andrieu F, Kilchytska V. Impact of self-heating and substrate effects on small-signal output conductance in UTBB SOI MOSFETs. *Solid State Electron* 2012;71:93–100.
- [6] Karim MA, Chauhan YS, Venugopalan S, Sachid AB, Lu DD, Nguyen BY, et al. Extraction of isothermal condition and thermal network in UTBB SOI MOSFETs. *IEEE Electron Device Lett* 2012;33:1306–8.
- [7] Liu Q, Vinet M, Gimbert J, Loubet N, Wacquez R, Grenouillet L, et al. High performance UTBB FDSOI devices featuring 20 nm gate length for 14 nm node and beyond. In: *IEEE international electron devices meeting (IEDM)*; 2013. p. 921–4.
- [8] El Ghitani H. DIBL coefficient in short-channel NMOS transistors. In: *16th national radio science conference*. p. D4.1–5.
- [9] Jharia B, Sarkar S, Agarwal RP. Controlling factors of drain induced barrier lowering coefficient in short channel MOSFET. *WSEAS Trans Electron* 2004;1:252–5.
- [10] Makovejev S, Kazemi Esfeh B, Raskin J-P, Flandre D, Kilchytska V, Andrieu F. Threshold voltage extraction techniques and temperature effect in context of global variability in UTBB MOSFETs. In: *Proceedings of ESSDERC*. p. 194–7.
- [11] Pereira ASN, de Streeel G, Planes N, Haond M, Giacomini R, Flandre D, et al. Analysis and modelling of temperature effect on DIBL in UTBB FD SOI MOSFETs. In: *2016 joint international EUROSUI workshop and international conference on ultimate integration on silicon (EUROSUI-ULIS)*. p. 116–9.
- [12] Planes N, Weber O, Barral V, Haendler S, Noblet D, Croain D, et al. 28 nm FDSOI technology platform for high-speed low-voltage digital applications. In: *Symposium on VLSI technology (VLSIT)*; 2012. p. 133–4.
- [13] Skotnicki T, et al. *MASTAR 4 User's Guide*, ST Microelectronics; 2011.
- [14] *Atlas User's Manual*. Santa Clara: Silvaco, v. 5.16.3.R; 2010.
- [15] Rozeau O, Jaud M-A, Poiroux T, Benosman M. *UTSOI model 1.1.4 – Model Description*, CEA-Leti; 2012.
- [16] Skotnicki T, Merckel G, Pedron T. The voltage-doping transformation a new approach to the modelling of MOSFET short-channel effects. In: *Proceedings of 17th European Solid State Device Research Conference (ESSDERC)*. p. 543–6.
- [17] Skotnicki T, Merckel G, Pedron T. The voltage-doping transformation: a new approach to the modeling of MOSFET short-channel effects. *IEEE Electron Device Lett* 1988;9:109–12.
- [18] Arshad MKM, Raskin J-P, Kilchytska V, Andrieu F, Scheiblin P, Faynot O, et al. Extended MASTAR modeling of DIBL in UTB and UTBB SOI MOSFETs. *IEEE Trans Electron Devices* 2012;59:247–51.
- [19] Fasarakis N, Karatsori T, Tassis DH, Theodorou CG, Andrieu F, Faynot O, et al. Analytical modeling of threshold voltage and interface ideality factor of nanoscale ultrathin body and buried oxide SOI MOSFETs with back gate control. *IEEE Trans Electron Devices* 2014;61:969–75.
- [20] Narayanan R, Ortiz-Conde A, Liou JJ, Garcia Sanchez FJ, Parthasarathy A. Two-dimensional numerical analysis for extracting the effective channel length of short-channel MOSFETs. *Solid-State Electron* 1995;38:1155–9.
- [21] Tsormpatzoglou A, Dimitriadis CA, Clerc R, Pananakakis G, Ghibaudo G. Threshold voltage model for short-channel undoped symmetrical double-gate MOSFETs. *IEEE Trans Electron Devices* 2008;55:2512–6.
- [22] Bol D, de Streeel G, Flandre D. Can we connect trillions of IoT sensors in a sustainable way? A technology/circuit perspective. In: *IEEE S3S conference*; 2015.
- [23] Kazemi Esfeh B, Kilchytska V, Planes N, Haond M, Flandre D, Raskin J-P. Comparative study of parasitic elements on RF FoM in 28 nm FD SOI and bulk technologies. In: *IEEE S3S conference*; 2015.



Contents lists available at ScienceDirect

Solid-State Electronics

journal homepage: www.elsevier.com/locate/sse

Characterization and modelling of layout effects in SiGe channel pMOSFETs from 14 nm UTBB FDSOI technology



R. Berthelon^{a,b,c,*}, F. Andrieu^b, S. Ortolland^a, R. Nicolas^a, T. Poiroux^b, E. Baylac^a, D. Dutartre^a, E. Josse^a, A. Claverie^c, M. Haond^a

^aSTMicroelectronics, 850 rue Monnet, B.P. 16, F-38926 Crolles, France

^bCEA-LETI, Minatec Campus, 17 rue des Martyrs, 38054 Grenoble, France

^cCEMES-CNRS, 29 Rue Jeanne Marvig, 31055 Toulouse Cedex 4, France

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

CMOS
Ultra-Thin-Body and Buried-Oxide Fully-Depleted-On-Insulator (UTBB FDSOI)
SiGe
Stress
Strain
Layout effects
Length of diffusion (LOD)
SA/SB parameters
Threshold voltage
Mobility

ABSTRACT

The introduction of strained channel is mandatory to achieve high performance in Ultra-Thin-Body and Buried-Oxide Fully-Depleted-Silicon-On-Insulator (UTBB FDSOI) technology. Especially, compressive SiGe channel has been demonstrated to enhance hole mobility and therefore pMOSFETs drive currents. At the same time, the performance gain induced by this mechanical stressor comes along with layout effects. In this study, we characterize experimentally the impact of the active region dimensions and shape on the threshold voltage and linear drain current of SiGe channel pMOSFETs directly on insulator fabricated for the 14 nm technology node. The pMOS threshold voltage increases by 105 mV for a gate-to-STI distance of 80 nm compared to 980 nm while I_{ODLIN} decreases by 51%. An analytical model is proposed to reproduce the layout dependences. The model is based on the stress profile, taking into account both the stress from the SiGe channel and from SiGe source/drain. It reproduces the experimental data with a good accuracy in the cases of symmetric and asymmetric layouts, provided a typical relaxation length of 112 nm is used. Finally, a special attention is paid on multifinger transistors, since they are widely used in standard cells designs.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The 14 nm node UTBB-FDSOI (Ultra-Thin Body and Buried Oxide – Fully Depleted Silicon On Insulator) technology highlights a –34% delay (or +50% frequency) improvement along with a 100 mV supply voltage reduction (0.8 V vs. 0.9 V) over the 28 nm FDSOI technology on a inverter ring oscillator with a fan-out FO = 3 [1]. This performance makes this technology competitive compared to 14 nm finFET and even more flexible for multi-purpose applications, thanks to its high body-bias efficiency [1]. High-speed capability is mainly achieved thanks to the introduction of a high-k metal gate first, a 6 nm thin SiGe channel and SiGe sources/drains in pMOSFETs. Both the SiGe channel and SiGe sources/drains lead to high level of stress in the pMOSFET channel. Intrinsically, these strained channels also exhibit strong layout effects, as evidenced in literature [2–4]. Such effects should be

reproduced into compact models in order to assure the functionality and the performance of the designs. In this study, we characterize and reproduce the layout dependences thanks to an analytical model, which is based on measured strain profiles.

First, a special attention is paid on the level of stress induced by the SiGe channel and SiGe sources/drains. Then, we propose an analytical model based on the physical mechanical behavior to capture the layout dependences. Finally, this model is used to reproduce the threshold voltage and the linear current variations for several layouts. Especially, asymmetric layouts of the active regions, as well as the number of gate fingers have been investigated. In all configurations, the model enables to reproduce the experimental data with a good accuracy.

2. Stress in devices

2.1. Technology summary

Studied devices are from the 14 nm UTBB-FDSOI technology developed at STMicroelectronics [1]. FDSOI devices are built on a 6 nm-thick silicon film on top of a 20 nm Buried-Oxide (BOX)

* Corresponding author at: STMicroelectronics, 850 rue Monnet, B.P. 16, F-38926 Crolles, France.

E-mail addresses: remy.berthelon@st.com (R. Berthelon), francois.andrieu@cea.fr (F. Andrieu).

resulting in excellent electrostatic behaviors. pMOSFETs channel is made of Silicon-Germanium in order to boost the performances over the 28 nm node. Devices feature TiN/HfO₂ gate-first high k metal gate integration and in-situ Boron-doped SiGe source/drain. In this study, we focus on Super-Low Threshold Voltage (SLVT) flavor with gate length of 20 nm, transistor width of 170 nm and a contacted poly pitch (CPP) of 90 nm.

2.2. Channel-induced stress

The integration of SiGe by the Ge-enrichment technique [5] leads to an intrinsically strained SiGe channel directly on insulator (SiGeOI) in pMOSFETs (Fig. 1). The strain comes from the lattice mismatch between Silicon ($a_{Si} = 5.431 \text{ \AA}$) and Germanium ($a_{Ge} = 5.658 \text{ \AA}$). After heteroepitaxy of SiGe on Silicon and condensation (oxidation performed at 1100 °C for 50 s), the lattice parameter of SiGe matches the one of the underneath Silicon leading to the in-plane strain $\epsilon_{cc} = \epsilon_{yy}$ (Fig. 2).

The SiGe lattice parameter a_{SiGe} can be expressed as a function of the Germanium concentration in the SiGe channel x [6]:

$$a_{SiGe} = 5.431 + 0.2x + 0.027x^2$$

The in-plane strain in the SiGe film ϵ_{xx} is then derived from the lattice mismatch:

$$\epsilon_{xx} = \frac{a_{Si} - a_{SiGe}}{a_{SiGe}}$$

With $a_{Si} = 5.431 \text{ \AA}$ is the lattice parameter of Silicon. The relationship between the strain and the biaxial stress is given by Hook's law, using elastic constants of Silicon [7].

$$\begin{pmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{yz} \\ \sigma_{xz} \\ \sigma_{xy} \end{pmatrix} = \begin{pmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{yz} \\ \epsilon_{xz} \\ \epsilon_{xy} \end{pmatrix}$$

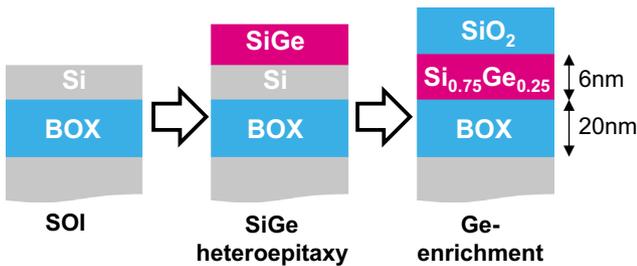


Fig. 1. Scheme of the Ge-enrichment process used to obtain Si_{0.75}Ge_{0.25} film directly on insulator.

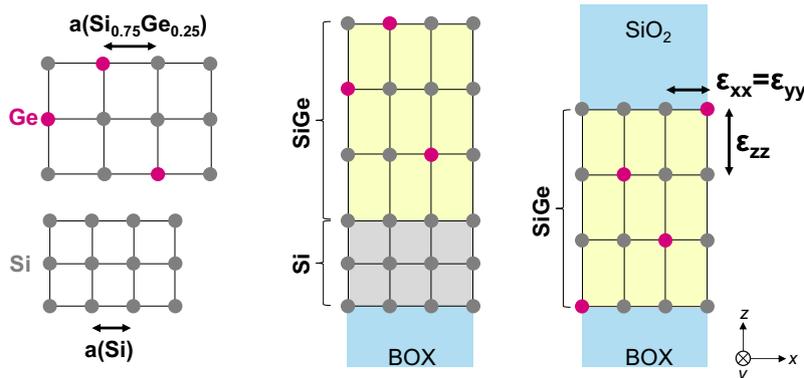


Fig. 2. Strain in SiGe after Ge-enrichment due to the lattice mismatch between SiGe and Silicon.

The elastic constants of SiGe are different compared to the one of pure Silicon and can be determined through Vegard's law:

$$C_{ij}(x) = C_{ij}(Si) + x(C_{ij}(Ge) - C_{ij}(Si))$$

Table 1 gives the elastic constants of Silicon, Germanium and Si_{0.75}Ge_{0.25}. Finally, the biaxial stress as a function of the Germanium concentration is given by:

$$\sigma_{biaxial} = \epsilon(x) \times \left(C_{11}(x) + C_{12}(x) - 2 \frac{C_{12}(x)^2}{C_{11}(x)} \right)$$

and is plotted in Fig. 3. For the Germanium concentration of interest ($x < 50\%$), a linear fit provides a satisfactory approximation.

$$\sigma_{CH} = \gamma \cdot x \text{ with } \gamma = -64.3 \text{ MPa/\%}$$

As a result, the introduction of Si_{0.75}Ge_{0.25} leads to a biaxial compressive stress of $\sigma = -1.6 \text{ GPa}$ in the pMOSFET channel.

2.3. Stress induced by the source and drain

The global stress induced by the SiGe channel integration is not the only source of stress in the device. The heteroepitaxy of

Table 1

Elastic constants of Silicon, Germanium and Si_{0.75}Ge_{0.25}.

Elastic constants	C ₁₁ (GPa)	C ₁₂ (GPa)	C ₄₄ (GPa)
Silicon	165.8	63.9	79.6
Germanium	128.5	48.3	66.8
Si _{0.75} Ge _{0.25}	156.6	60	76.4

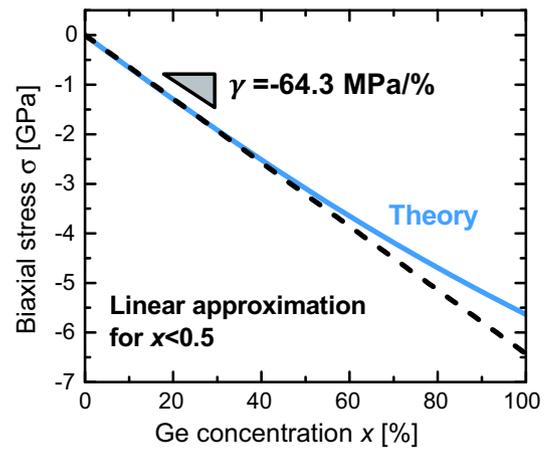


Fig. 3. Biaxial stress in SiGe grown on Silicon according the Germanium concentration x , compared to a linear approximation.

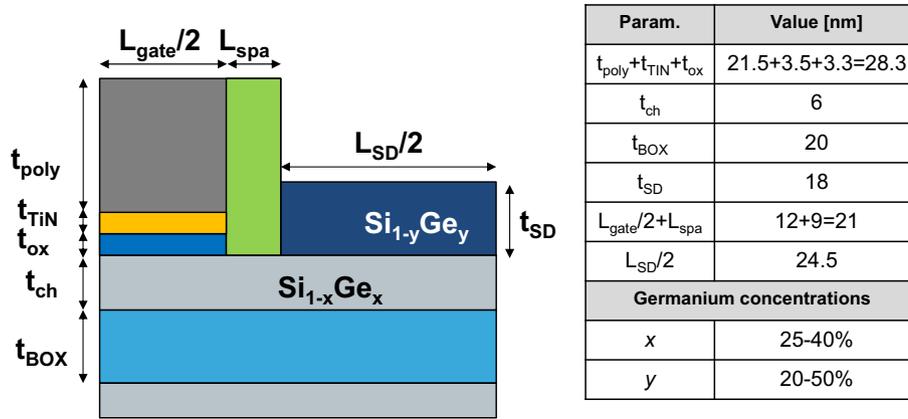


Fig. 4. (Left) Schematics of the simulated structure. (Right) Geometry parameters used in this study.

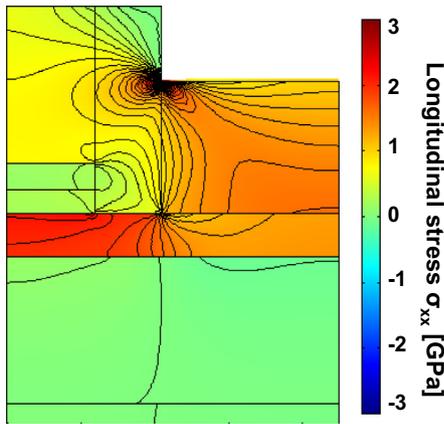


Fig. 5. 2D mapping of the longitudinal stress σ_L from finite element mechanical simulation with $x = 25\%$ and $y = 30\%$.

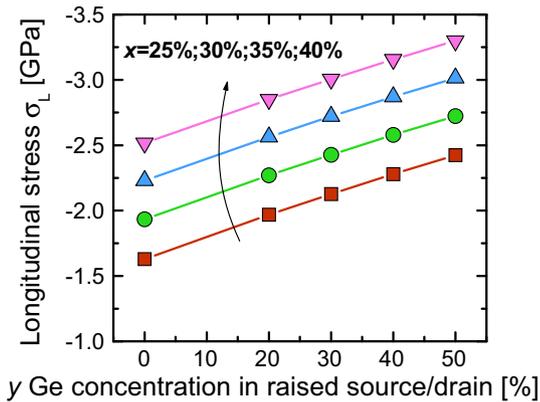


Fig. 6. Total longitudinal stress σ_L in the channel at 1 nm below the gate and in the middle of a fully strained $\text{Si}_{1-x}\text{Ge}_x$ channel for different Ge concentrations y in raised source/drain.

$\text{Si}_{1-y}\text{Ge}_y\text{:B}$ as Sources and Drains (S/D) of the transistor leads to an additional stress in the channel. Using finite element mechanical simulations (with COMSOL software), the longitudinal stress σ_L (i.e. along the $(1\ 1\ 0)$ transport direction) in the channel has been calculated for different Germanium concentrations in the channel (x) and in the Source/Drain (y) while the geometry of the structure remains unchanged (Figs. 4–6). The simulated structure corresponds to a “not isolated device” and thus consists of half a transistor because of symmetry plans. The channel is intrinsically strained

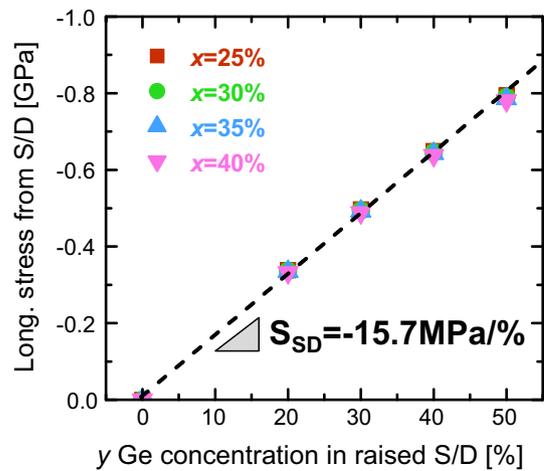


Fig. 7. Additional stress induced only by the source/drain, measured at 1 nm below the gate and in the middle of the gate according to the Germanium concentration in source/drain and for different $\text{Si}_{1-x}\text{Ge}_x$ channel (fully strained). Linear relationship is evidenced with sensibility $S_{\text{SD}} = -15.7\text{ MPa}/\%$.

from $\text{Si}_{1-x}\text{Ge}_x$ integration. The additional stress in the channel induced by source and drain comes from the relaxation of SiGe:B . As the source and drain relax, the $\text{Si}_{1-x}\text{Ge}_x$ film underneath in the source/drain region is under a tensile stress while the channel region is put under compression from both sides. By varying the concentrations x and y , it is found that the additional stress from the $\text{Si}_{1-y}\text{Ge}_y$ source and drain does not depend on the Germanium concentration in the channel x (Fig. 7). This is due to the fact that $\text{Si}_{1-y}\text{Ge}_y$ source and drain are deposited on a $\text{Si}_{1-x}\text{Ge}_x$ channel featuring the lattice parameter of Silicon, whatever its Germanium concentration x . Therefore, the lattice mismatch between the channel and the source and drain only depends on the Germanium concentration in source and drain y . A good linearity between the additional stress and the Germanium concentration in source/drain y is observed with proportionality coefficient $S_{\text{SD}} = -15.7\text{ MPa}/\%$ for the considered geometry.

3. Relaxation model

3.1. Channel relaxation

The SiGe channel enrichment is processed before the patterning of the active area. The etching of the Shallow Trench Isolation (or mesa isolation) introduces a free boundary condition at the edges.

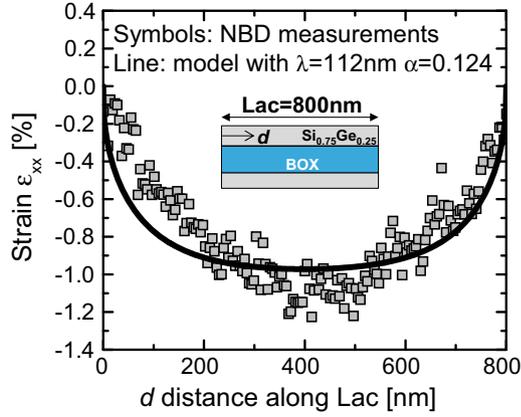


Fig. 8. Strain along a $\text{Si}_{0.75}\text{Ge}_{0.25}$ channel of active length $Lac = 800$ nm measured by Nano-Beam Diffraction (NBD) and compared with the analytical model.

There is thus a stress discontinuity at the edge of the SiGe film. Consequently, the compressive stress in the film tends to relax. The strain in SiGe is measured after etching by Nano-Beam Diffraction along an active length of $Lac = 800$ nm and considering 25% of Germanium (Fig. 8). At the active border, the stress relaxes due to the free boundary condition introduced by etching. Far from the active border, the SiGe channel is fully strained. The stress profile along the active area is modeled thanks to an analytical model [8]:

$$f(d, Lac) = \left[\frac{1}{2} \left(\left(1 - \exp\left(-\frac{d}{\lambda}\right) \right)^\alpha + \left(1 - \exp\left(-\frac{d-Lac}{\lambda}\right) \right)^\alpha \right) \right]^{1/\alpha}$$

With d the position along the length Lac of the active region, λ the typical relaxation length and α a fitting parameter. The typical relaxation length λ is used to reproduce the relaxation that occurs over a certain distance from the edge of the active area. The stress in the SiGe channel is derived from this so-called stress function f , according to the Germanium concentration in the channel x :

$$\sigma_{CH} = \sigma(x) \cdot f = \gamma \cdot x \cdot f$$

When the dimension of the active area is strongly reduced, the edge effects become predominant leading to a partially strained film as evidenced in Fig. 9. The model gives a good qualitative agreement with the measured strain profile with fit parameters $\lambda = 112$ nm and $\alpha = 0.124$.

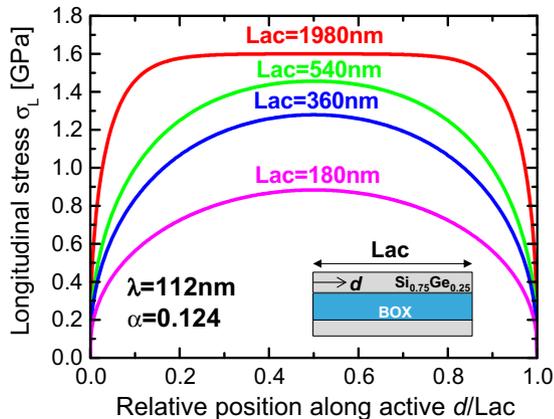


Fig. 9. Longitudinal stress value along the active position d from the analytical model for different active lengths Lac , considering $x = 25\%$.

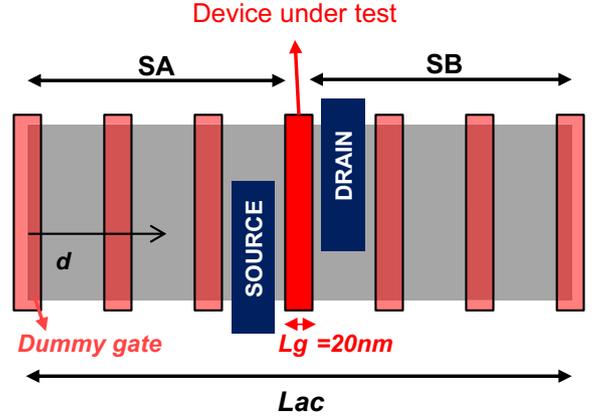


Fig. 10. Typical layout of devices under test. The gate-to-STI distances SA/SB depend on the number of dummy gate fingers.

3.2. The effectiveness of S/D-induced stress on a (partially) relaxed channel

The stress induced in the channel by the raised source and drain depends on the initial level of strain in $\text{Si}_{1-y}\text{Ge}_y:\text{B}$, which is directly linked to the Germanium concentration in the source/drain y . As seen before, because of the patterning, the SiGe channel can be partially relaxed. As a result, the lattice parameter in the channel is no longer the one of Silicon. The source and drain can thus be deposited on a channel with a lattice parameter larger than the one of Silicon. This lattice parameter depends on the level of relaxation of the channel. If the channel is partially relaxed, the intrinsic strain in raised source and drain is reduced since the lattice mismatch between the channel and the source/drain is lower. In order to take this effect into account, it is convenient to introduce the equivalent Germanium concentration y_{eq} . y_{eq} can be defined as the Ge concentration in the source/drain integrated on a fully strained channel $\text{Si}_{1-x}\text{Ge}_x$ channel, for which the stress induced by the source/drain is equivalent to the one induced by $\text{Si}_{1-y}\text{Ge}_y$ on a $\text{Si}_{1-x}\text{Ge}_x$ channel under partial relaxation. This equivalent concentration is derived from the initial Germanium concentrations x and y and using the stress function f :

$$y_{eq} = y - x(1 - f)$$

The additional stress from the source/drain σ_{SD} is then derived using the sensibility S_{SD} previously extracted from simulations performed on “not isolated channel” and assuming the equivalent Germanium concentration in the source/drain y_{eq} :

$$\sigma_{SD} = S_{SD}y_{eq}$$

$$\sigma_{SD} = S_{SD}(y - x(1 - f))$$

If the channel is fully strained (stress function $f = 1$), the stress from the source/drain does not rely on the Germanium concentration in the channel, as previously demonstrated by simulations. In the extreme case of a fully relaxed channel (stress function $f = 0$), the stress from the source/drain depends on the difference of Germanium concentrations in the source/drain and in the channel (i.e. $y - x$). For short active length, the channel can be partially relaxed (see Fig. 9). In this intermediate case, the stress from source/drain is linked to the level of relaxation of the channel.

Finally, the total longitudinal stress in the $\text{Si}_{1-x}\text{Ge}_x$ channel transistor with $\text{Si}_{1-y}\text{Ge}_y$ source/drain at a position d along an active region of a total length Lac is:

$$\sigma_L(d, Lac, x, y) = \gamma x f(d, Lac) + S_{SD}(y - x(1 - f(d, Lac)))$$

For short channels, the longitudinal stress can be considered as the one at the middle of the channel since the channel length is significantly lower than the active length. For instance, the longitudinal stress in the middle of a Si_{0.75}Ge_{0.25} channel with a gate length of 20 nm, located in the middle of an active length $L_{ac} = 0.36 \mu\text{m}$, with Si_{0.7}Ge_{0.3}:B source/drain is $\sigma_L = -1.7 \text{ GPa}$, assuming a typical relaxation length of $\lambda = 112 \text{ nm}$. In this case, 75% of the longitudinal stress comes from the channel (corresponding to σ_{CH} component) and 25% from the source/drain (corresponding to σ_{SD} component).

4. Layout effect and corresponding model

In this section we investigate the layout effects relative to the gate-to-STI distances SA/SB (also called the length of diffusion LOD) on short channel devices ($L_g = 20 \text{ nm}$, see Fig. 10). The stress value of interest is taken at the middle of the channel ($d = SA + L_g/2$), which is a relevant approximation for short channels. The active length is $L_{ac} = SA + SB + L_g$. The value of SA and SB varies according to the number of dummy gates fingers on the active region, with a Contacted-Poly-Pitch (CPP) of 90 nm. The measurement of transistors with various dummy gates thus enables to characterize the layout effects linked to the longitudinal stress, while other layout parameters remain unchanged. Especially, the gate length L_g is kept constant and the source/drain regions are in a similar configuration (i.e. SiGe:B in between two gates) whatever the number of dummy gate fingers of the layout.

In this study, we focus on two relevant electrical parameters impacted by the stress: the threshold voltage V_T , extracted at a constant current, and the I_{ODLIN} drain current in linear regime (measured at a drain voltage of $|V_D| = 50 \text{ mV}$ and at a given overdrive $|V_G - V_T| = 0.5 \text{ V}$). I_{ODLIN} is influenced by the hole mobility in the channel and the series resistance.

In order to understand and interpret our experimental results, we have built an analytical model. In this model, the link between the longitudinal stress and the threshold voltage is established by a first order linear sensibility of $S_{VT} = -113 \text{ mV/GPa}$. This sensibility is consistent with the valence band shift used in Ref. [9] and calculated from Ref. [10], in case of biaxial stress ($\Delta E_V = 103 \text{ mV/GPa}$).

Regarding the mobility, we use the piezoresistive model to predict the mobility variation according to the level of longitudinal stress:

$$\frac{\delta\mu}{\mu} = -\Pi_L \delta\sigma_L$$

With μ being the hole mobility, Π_L the longitudinal piezoresistive coefficient in the $\langle 110 \rangle$ channel orientation and σ_L the longitudinal stress. After integration, the mobility can thus be calculated as:

$$\frac{\mu(\sigma_L)}{\mu(\sigma_L=0)} = \exp\left(\int_0^{\sigma_L} -\Pi_L(s) ds\right)$$

This model requires the knowledge of the evolution of the longitudinal piezoresistive coefficient Π_L with the level of initial stress in the device σ . For the stress range of interest, we use the piezoresistive coefficients extracted on long channels ($L_g = 10 \mu\text{m}$) at $\sigma = 0$ ($\langle 110 \rangle$ Si channel) and at $\sigma = -1.6 \text{ GPa}$ ($\langle 110 \rangle$ SiGe with $x = 25\%$), which are $\Pi_L(0) = 0.56 \text{ GPa}^{-1}$ and $\Pi_L(-1.6) = 0.74 \text{ GPa}^{-1}$ respectively (Fig. 11).

From these measurement, the evolution of Π_L with σ_L is assumed as linear. The relative sensibility can be calculated as $\Lambda = -0.11 \text{ GPa}^{-2}$, and the final model of mobility enhancement given by:

$$\frac{\mu(\sigma_L)}{\mu(\sigma_L=0)} = \exp\left(-\Pi_L(\sigma_L=0) \cdot \sigma_L - \frac{\Lambda}{2} \sigma_L^2\right)$$

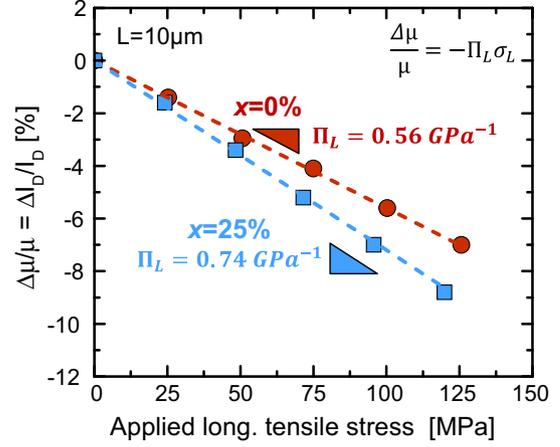


Fig. 11. Piezoresistive coefficients extraction from hole mobility relative variation according to the applied longitudinal tensile stress for two different initial Germanium concentration.

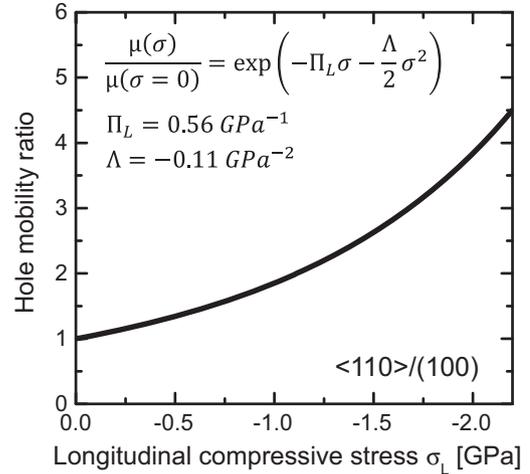


Fig. 12. Model of hole mobility ratio in a SiGe channel vs. its longitudinal compressive stress.

The mobility ratio from this analytical model is represented in Fig. 12.

Finally, the I_{ODLIN} current model and threshold voltage model are given by:

$$I_{ODLIN} = I_{d0} \exp\left(-\Pi_L \sigma_L - \frac{\Lambda}{2} \sigma_L^2\right)$$

$$V_T = V_{T,0} + S_{VT} \sigma_L$$

With I_{d0} and $V_{T,0}$ being respectively the I_{ODLIN} current and threshold voltage of a device which is unstrained in the longitudinal direction.

5. Experimental results

5.1. Symmetric layouts

Layouts are called symmetric if $SA = SB$. In this configuration, the gate-to-STI distance from the source side is equal to the one from the drain side. The nMOS threshold voltage V_T , extracted at a constant current of $300 \text{ nA} \times W/L$ with $V_D = 50 \text{ mV}$, as well as the drain current I_{ODLIN} , measured at a constant overdrive $|V_G - V_T| = 0.5 \text{ V}$ are shown in Fig. 13 for symmetric layouts with

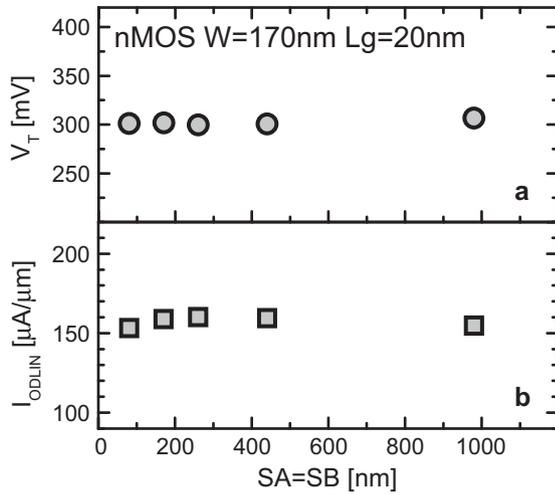


Fig. 13. (a) Threshold voltage V_T and (b) I_{ODLIN} drain current of nMOS for symmetric layouts, as a function of $SA = SB$. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm and V_T flavor is SLVT.

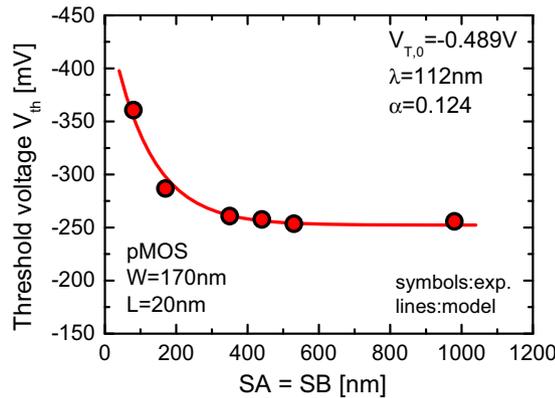


Fig. 14. pMOS threshold voltage V_T for symmetric layouts, as a function of $SA = SB$. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

different gate-to-STI values $SA = SB$. No layout-dependence is observed since V_T and I_{ODLIN} are constant whatever the active length. This result is expected since the nMOS features unstrained silicon channel. It also confirms that the STI does not induce any stress into the device. Figs. 14 and 15 show the pMOS V_T (extracted at a constant current of $70 \text{ nA} \times W/L$ with $V_D = -50 \text{ mV}$) and I_{ODLIN} , respectively, as a function of $SA = SB$. The pMOS threshold voltage increases by 105 mV from $SA = SB = 980 \text{ nm}$ to $SA = SB = 80 \text{ nm}$ while I_{ODLIN} decreases by 51%. The change in threshold voltage and mobility comes from the relaxation of the SiGe stress in the channel that occurs at the active/STI frontiers. For short active regions, the channel is partially relaxed and thus the electrical characteristics are impacted. These layout-dependences are reproduced with a good accuracy by our aforementioned stress-based model with a typical relaxation length λ of 112 nm.

5.2. Asymmetric layouts

The layout effects are also characterized for asymmetric layouts, with different gate-to-STI distances between source and drain sides. In such layouts with $SA \neq SB$, the distance with the active border on drain side, SB , is fixed at either $SB = 980 \text{ nm}$ or $SB = 80 \text{ nm}$. These asymmetric layouts enable to characterize the

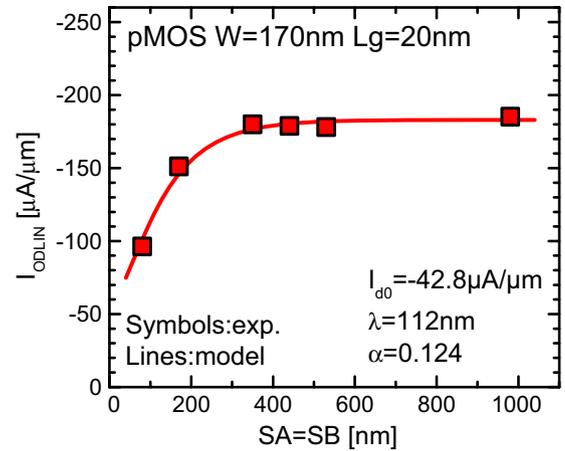


Fig. 15. pMOS I_{ODLIN} drain current for symmetric layouts, as a function of $SA = SB$. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

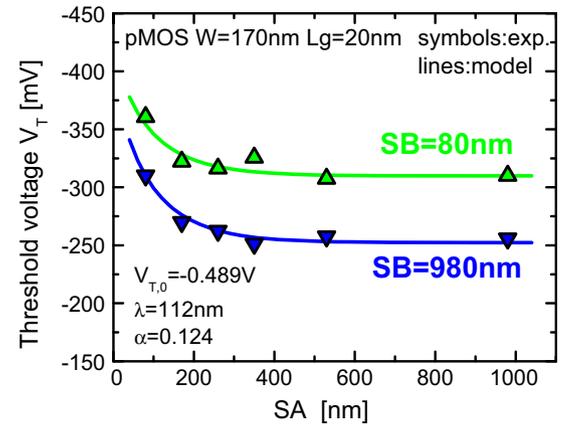


Fig. 16. pMOS threshold voltage V_T for asymmetric layouts as a function of SA and for SB fixed at either $SB = 980 \text{ nm}$ or $SB = 80 \text{ nm}$. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

proximity effect of one active border only. Fig. 16 shows the threshold voltage variations for asymmetric layouts. Similarly to symmetric layouts, the proximity of the active border leads to threshold voltage increase. The threshold voltage is directly limited by the shorter gate-to-STI distance. A V_T shift of 54 mV is found between $SB = 980 \text{ nm}$ and $SB = 80 \text{ nm}$ while $SA = 980 \text{ nm}$. This V_T shift is due to the stress relaxation on the active border of the drain side only and is approximately half the V_T shift induced by the proximity of two edges (54 mV vs. 105 mV). I_{ODLIN} is similarly impacted as demonstrated in Fig. 17. I_{ODLIN} is also strongly limited by the shorter gate-to-STI distance and decreases by 30% between $SB = 980 \text{ nm}$ and $SB = 80 \text{ nm}$ while $SA = 980 \text{ nm}$. Both V_T and I_{ODLIN} variations for asymmetric layouts are well reproduced by our analytical model, provided $\alpha = 0.124$ as a value of the asymmetry parameter.

5.3. Multifinger transistors

In a design, transistors can be built on a same active area. Especially, multifinger MOSFETs consist of different transistors in parallel and on the same active region, sharing source and drain nodes as illustrated in Fig. 18. Such layouts are particularly efficient to enhance the drivability of standard cells. In this configuration, each

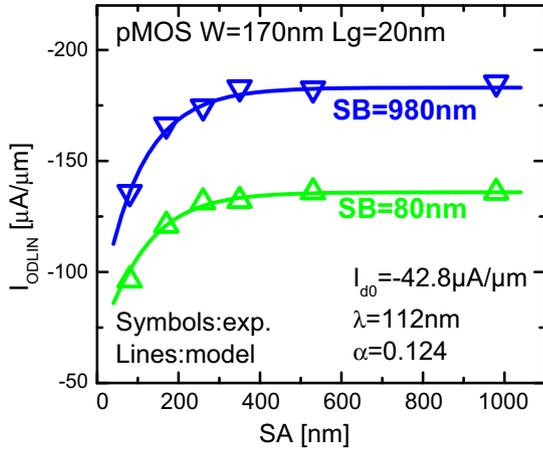


Fig. 17. pMOS I_{ODLIN} drain current for asymmetric layouts as a function of SA and for SB fixed at either SB = 980 nm or SB = 80 nm. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

elementary transistor of multifinger MOSFETs is unique because of a unique set of gate-to-STI distances SA_i/SB_i . Indeed, the stress relaxation occurring at the active borders will affect each finger differently. Therefore, the level of stress in the channel of each finger directly depends on its position.

In order to derive the threshold voltage V_T according to the number of fingers N_f , we consider the N_f transistors built in parallel and for a conduction in the subthreshold regime (no impact of drift current):

$$V_T(N_f) = -SS / \ln(10) \ln \left(\frac{1}{N_f} \sum_{i=1}^{N_f} \exp \left(\frac{-V_T(SA_i, SB_i)}{SS / \ln(10)} \right) \right)$$

with SS the subthreshold swing of the tested devices (considered in the following to be 80 mV/decade). A good agreement between model and experimental data is observed (Fig. 19). Multifinger data are compared to the 1-finger transistor reference (data from Fig. 14), located at the middle of the active area (this time, the rest of the gates are dummies, as previously illustrated in Fig. 10). This result highlights the impact of the fingers close to the active border, slightly increasing the threshold voltage of the whole structure (+17 mV at $N_f = 11$ fingers). However, one can see that the global V_T of multifinger devices can be well approximated by the V_T of the central transistor. This is because this latter is the minimum V_T of all the devices in parallel and the minimum V_T governs the total V_T .

Same methodology is applied to estimate I_{ODLIN} of multifingers transistors. However, in such designs, the different fingers are not at the same gate voltage overdrive $V_G - V_T$ since fingers feature

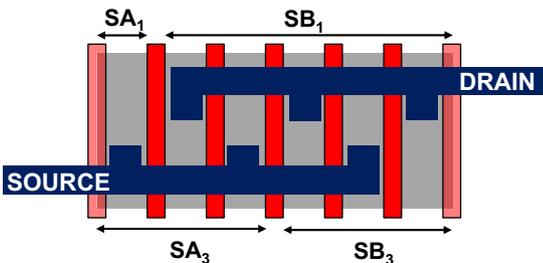


Fig. 18. Illustration of multifinger transistors layout built in parallel (in this example, the number of finger is $N_f = 5$). Each elementary transistor exhibits different gate-to-STI distances SA/SB.

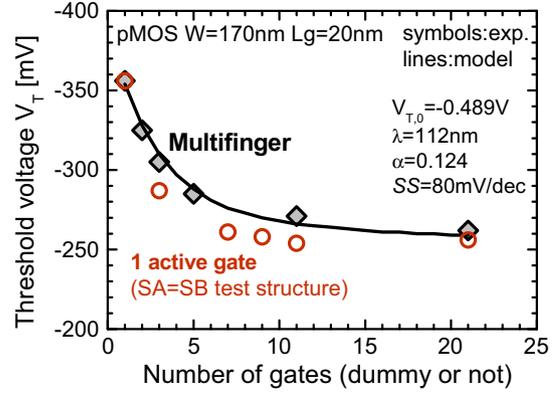


Fig. 19. Threshold voltage of multifinger pMOS as a function of the number of gates. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

different threshold voltages according to their gate-to-STI distances (SA_i/SB_i). A simple model of linear drain current can be used in order to take into account the overdrive shift for each finger:

$$I_D(V_G) = \beta \frac{|V_G - V_T|}{1 + \theta_1 |V_G - V_T|}$$

At $|V_G - V_T| = V_{GT0} = 0.5$ V, $I_D = I_{ODLIN}$. Thus, I_D can be written as a function of I_{ODLIN} :

$$I_D(SA, SB) = I_{ODLIN}(SA, SB) \cdot \frac{V_{GT}}{V_{GT0}} \cdot \left(\frac{1 + \theta_1 |V_{GT0}|}{1 + \theta_1 |V_{GT}|} \right)$$

The parameter $\theta_1 = 0.3$ is used in the following because it reproduces well our experimental $I_D(V_G)$ curves. Finally, I_{ODLIN} of multifingers is measured at $|V_G - V_T(N_f)| = V_{GT0}$, with $V_T(N_f)$ previously described, and can thus be calculated following:

$$I_{ODLIN}(N_f) = \frac{1}{N_f} \sum_{i=1}^{N_f} I_D(SA_i, SB_i)$$

The multifinger pMOS I_{ODLIN} is shown in Fig. 20. Our model of transistors in parallel well reproduces the experimental data, provided an additional series resistance of 10 Ohm considered for $N_f > 10$. This additional resistance may be linked to wires/contact in our structure since it also affects multifinger nMOS I_{ODLIN} (Fig. 21). The multifinger transistor is compared to the 1-finger ref-

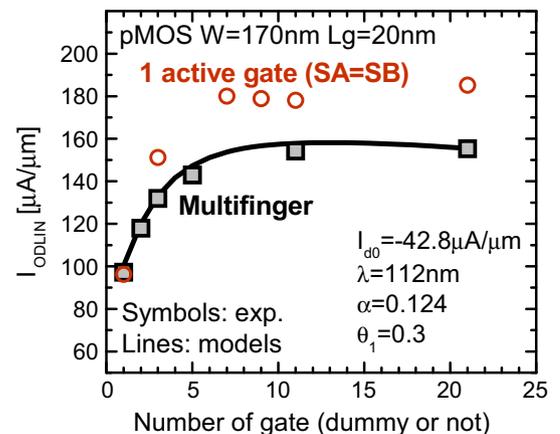


Fig. 20. Multifinger pMOSFETs I_{ODLIN} as a function of the number of fingers, compared to a single-finger pMOS lied on an active regions of the same dimensions. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT and Germanium concentrations are $x = 25\%$ and $y = 30\%$.

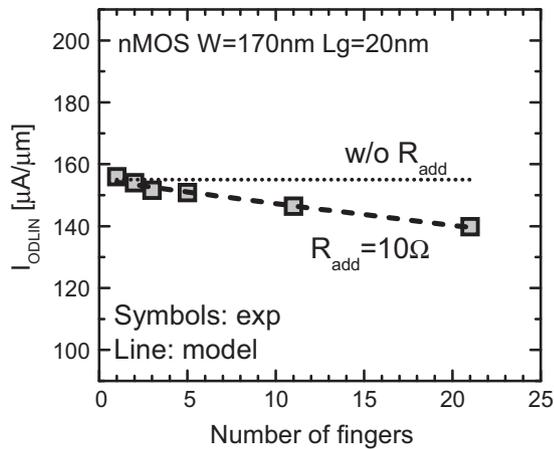


Fig. 21. Multifinger nMOSFETs I_{ODLIN} drain current as a function of the number of fingers with or without an additional series resistance of 10 Ohm. Gate length is $L_g = 20$ nm, channel width is $W = 170$ nm, V_T flavor is SLVT.

erence with same active length ($S_A = S_B$ in Fig. 10, data from Fig. 15). The multifinger design is more sensible to the single-finger design (for a given active dimension), this because the fingers close to the active border lowers I_{ODLIN} in the former case. Especially, I_{ODLIN} is 13% lower for multifinger at $N_f = 11$ than for single-finger layout, this, due to the fingers located close to the active/STI frontier.

6. Conclusion

In this study, we investigate the layout effects induced by the SiGe channel in the pMOSFETs of the 14 nm UTBB-FDSOI technology. The threshold voltage and the linear drain current are characterized vs. the active dimensions (length) and position of the gate. This layout effect is induced by the stress relaxation on the active borders (active/isolation frontier) that occurs during the isolation patterning. An analytical model based on the stress profile is proposed to reproduce the threshold voltage and linear drain current dependences with layouts. This model takes into account both the

intrinsic stress from the SiGe channel and the additional stress from the source/drain. Mechanical simulations as well as physical characterizations enable us to extend this model to a wide range of Germanium content and layout configurations. Finally, a good agreement is found with experimental data for both symmetric and asymmetric layouts. Multifinger transistors (i.e. transistors with multiple gates connected in parallel and put on a same active island) have also been characterized and modeled as they are widely used in standard cells designs, demonstrating the strong impact of gate fingers close to the active borders.

Acknowledgements

This work was partly supported by the Catrene Dynamic-UPL, Places2be KETs and Nano2017 projects, through the French ministry of Industry.

References

- [1] Weber O, Josse E, Mazurier J, Degors N, Chhun S, Maury P, et al. 14 nm FDSOI upgraded device performance for ultra-low voltage operation. In: VLSI tech symp. p. T168–9.
- [2] DeSalvo B, Morin P, Pala M, Ghibaudo G, Rozeau O, Liu Q, et al. A mobility enhancement strategy for sub-14 nm power efficient FDSOI technologies. In: IEEE IEDM. p. 7.2.1–4.
- [3] Cheng K, Khakifirooz A, Loubet N, Luning S, Nagumo T, Vinet M, et al. High performance extremely thin SOI (ETSOI) hybrid CMOS with Si channel NFET and strained SiGe channel PFET. In: IEEE IEDM. p. 419.
- [4] Andrieu F, Cassé M, Baylac E, Perreau P, Nier O, et al. Strain and layout management in dual channel (sSOI substrate, SiGe channel) planar FDSOI MOSFETs. In: Proc of ESSDERC conference. p. 106–9.
- [5] Glowacki F, Le Royer C, Morand Y, Pédini J-M, Denneulin T, Cooper D, et al. Ultrathin (5 nm) SiGe-On-Insulator with high compressive strain (–2 GPa): from fabrication (Ge enrichment process) to in-depth characterizations. Solid-State Electron 2014;97:82–7.
- [6] Dismukes JP, Ekstrom L, Steigmeier EF, Kudman I, Beers DS. Thermal and electrical properties of heavily doped Ge-Si alloys up to 1300 K. J Appl Phys 1964;35:2899.
- [7] Hopcroft MA, Nix WD, Kenny TW. What is the Young's modulus of silicon? J Microelectromech Syst 2010;19(2).
- [8] Poiroux T et al. In: MOS-AK workshop, Grenoble; 2015.
- [9] Rieger MM, Vogl P. Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates. Phys Rev B: Condensed Matter 1993;48(19):14276–87.
- [10] Cassé M, Hutin L, Le Royer C, Cooper D, Andrieu F, Weber O, et al. Experimental investigation of hole transport in strained $Si_{1-x}Ge_x$ /SOI pMOSFETs: part I – scattering mechanisms in long channel devices. IEEE Trans Electron Dev 2011;59(2):316–25.



Back-gated InGaAs-on-insulator lateral N⁺NN⁺ MOSFET: Fabrication and typical conduction mechanisms



H.J. Park^{a,*}, L. Pirro^a, L. Czornomaz^b, I. Ionica^a, M. Bawedin^a, V. Djara^b, V. Deshpande^b, S. Cristoloveanu^a

^aIMEP-LaHC, Minattec, Grenoble INP, Univ. Grenoble Alpes, CNRS, Grenoble, France

^bIBM Research, Ruschlikon, Switzerland

ARTICLE INFO

Article history:

Available online 17 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

III–V
InGaAs
SOI
pseudo-MOSFET
MOSFET
Wafer bonding
Carrier mobility
Strain
Selective regrowth

ABSTRACT

Back-gated InGaAs-on-insulator lateral N⁺NN⁺ MOSFETs are successfully fabricated by direct wafer bonding and selective epitaxial regrowth. These devices were characterized using a revisited pseudo-MOSFET configuration. Two different transport mechanisms are evidenced: volume conduction in the undepleted region of the film and surface conduction at the interface between InGaAs and buried insulator. We propose extraction techniques for the volume mobility and interface mobility. The impact of film thickness, channel width, and length is evaluated. Additional measurements reveal the variation of the transistor parameters at low temperature and under externally applied uniaxial tensile strain.

© 2016 Published by Elsevier Ltd.

1. Introduction

The extension of Moore's law is based on two pillars: (i) electrostatic integrity of small transistors and (ii) improvement of transport properties for high speed. Only fully depleted devices, such as FinFETs, planar SOI or nanowires, benefit from excellent control of the gate on the channel, leading to attenuated short-channel effects and leakage current. Technology-wise, Semiconductor on Insulator (SOI) is the most straightforward, and pragmatic approach. Mobility boosters consist in adding strain, selecting the crystal orientation or, more drastically, replacing silicon with more talented semiconductors [1–5].

According to ITRS predictions [6], high-mobility channel materials such as Ge, SiGe and III–V compounds are promising candidates for the next generations of MOSFETs, able to deliver the necessary power-performance benefits and added functionality for CMOS and System-on-Chip applications [7–9]. Furthermore, devices fabricated on semiconductor-on-insulator substrates have better electrostatic control with decreased short channel effects

and reduced leakage current [10–12]. Merging the merits of III–V compounds and SOI is a recent option explored as starting substrate (III–V on insulator) for device fabrication. In particular, InGaAs-on-insulator structures are attractive [13] to integrate high-performance FinFETs [14], planar MOSFETs on insulator and hybrid InGaAs/SiGe CMOS circuits [15]. This technology is rapidly evolving. Nevertheless, InGaAs-on-insulator films are prone to additional carrier scattering, compared to layers grown on bulk crystalline buffers such as InP [16] or InAlAs [17], owing to the presence of the buried oxide (BOX) and related back interface.

Our work is focused on the technological optimization via detailed investigation of transport properties in simple test devices. In_{0.53}Ga_{0.47}As layers transferred on oxide were characterized using a modified version of the well-known pseudo-MOSFET (Ψ -MOSFET) configuration [18]. Lateral N⁺NN⁺ structures have been fabricated on InGaAs-on-insulator layers to mimic the channel transport of a MOSFET while minimizing the impact of process-induced damages. Section 2 presents the main processing steps [19]. The measurement set-up and typical transistor characteristics are shown in Section 3. A suitable procedure, based on the extension of Y-function technique, is proposed in Section 4 for extracting the electron mobility in the film volume and at the

* Corresponding author.

E-mail address: hyungjin.park@imep.grenoble-inp.fr (H.J. Park).

interface. We discuss the carrier mobility as a function of film thickness, temperature from 77 K to 300 K, and uniaxial tensile strain.

2. Material and device processing

The fabrication of back-gated lateral N^+NN^+ transistors utilizes some of the key steps developed for InGaAs FinFETs [15]. The characteristics are expected to be representative of those in fully-processed devices with back-gate operation. The n-layer is a non-intentionally doped InGaAs-on-insulator (InGaAs-OI) grown in a metal organic vapor phase epitaxial reactor. InGaAs-OI layers on Si which have an initial doping concentration of $N_D \sim 2 \times 10^{17} \text{ cm}^{-3}$ are prepared by direct wafer bonding of nominally-undoped $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ films of varying thickness (25 nm, 50 nm, 100 nm and 200 nm) grown on InP wafers as described in [20]. The carrier concentration measured by Hall Effect on a 200-nm-thick InGaAs on InP layer before bonding is of $2\text{--}3 \times 10^{16} \text{ cm}^{-3}$. The buried oxide (BOX) consists of 25 nm thermal SiO_2 , 10 nm Al_2O_3 and the high-k dielectric stack used in [14], for a resulting C_{BOX} of $0.11 \mu\text{F}/\text{cm}^2$.

The fabrication of lateral N^+NN^+ structures starts with the deposition of a SiO_2 hard-mask to define the length of the n-region (Fig. 1(a)). Sn-doped $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ N^+ regions ($N_D = 5 \times 10^{19} \text{ cm}^{-3}$) are selectively grown (Fig. 1(b)) at low temperature as in [21]. The hard-mask is removed (Fig. 1(c)) and a mesa isolation is performed by wet etching (Fig. 1(d)). The top interface of the n-region is formed with same high-k dielectric as used for the bottom interface. The process is completed by the deposition of a SiO_2 layer (Fig. 1(e)), etching via holes and W metal contacts (Fig. 1(f)). It should be noted that although the bottom and top InGaAs/high-k interfaces are formed with the same deposition process, they differ in the fact that the bottom interface is exposed to the thermal budget of the N^+ selective epitaxial process, contrary to the top interface. This leads to a difference in the interface trap density

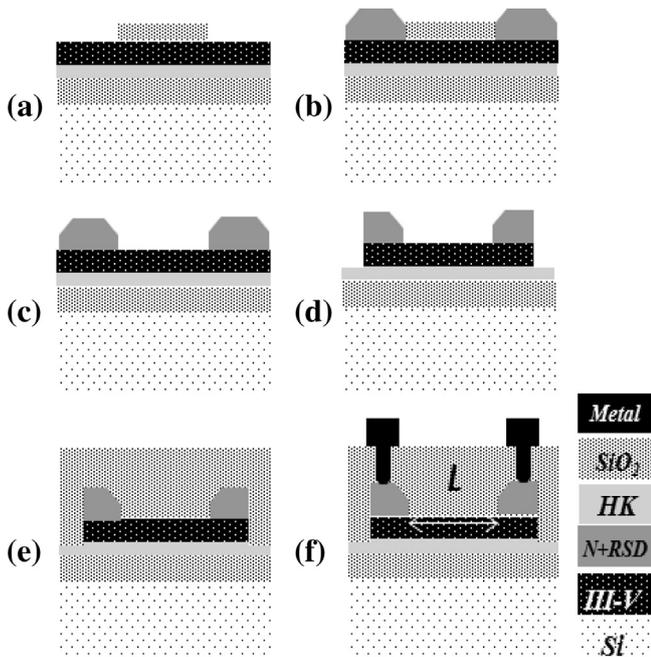


Fig. 1. Schematic processing sequence of lateral N^+NN^+ test structure: (a) SiO_2 hardmask on InGaAs-OI layer with bottom passivation, (b) growth of N^+ InGaAs contacts, (c) hardmask removal, (d) mesa wet etching, (e) top passivation and interlayer dielectric deposition, (f) contact vias and metallization.

comparable that the difference between Gate-first and Replacement-metal-gate devices in [14].

Fig. 2 shows the schematic of the tested structures. The drain current variation with drain bias confirms the ohmic behavior of the contacts.

3. Current and transconductance characteristics

The device characterization was performed by sweeping the gate bias V_G applied on the substrate and measuring the resulting drain current (I_D) between source and drain contacts. This device is similar to Pseudo-MOSFET (Ψ -MOSFET) [22], where the standard pressure-adjustable probes are replaced by implanted N^+ source and drain regions. This configuration is preferable because the impact of series resistance and probe-induced defects on the mobility is reduced. The channel dimensions are also better

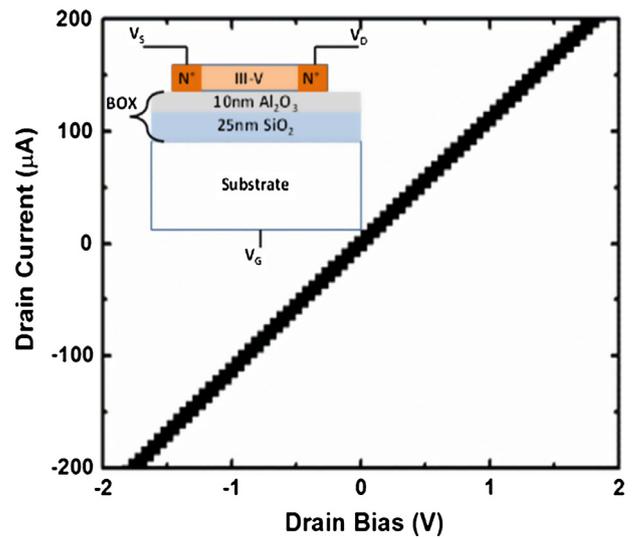


Fig. 2. Measured drain current I_D versus drain bias V_D for gate bias $V_G = 0$, $L = 200 \mu\text{m}$ and $W = 20 \mu\text{m}$.

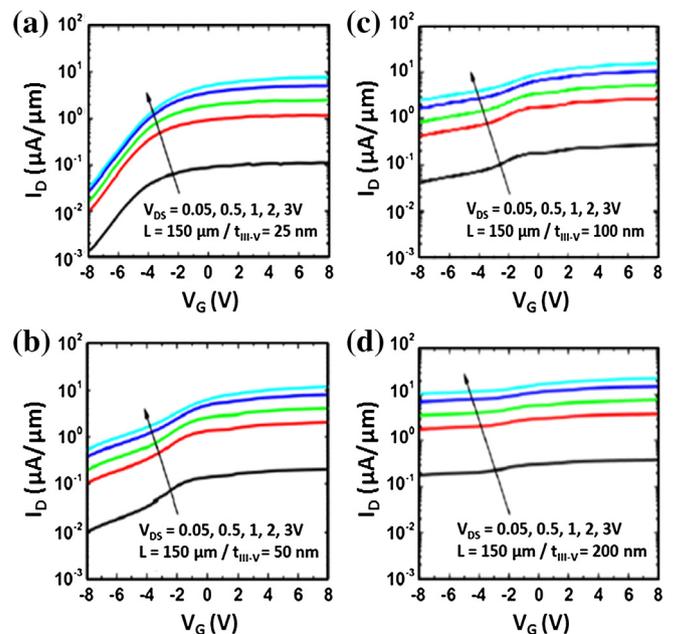


Fig. 3. Back-gate $I_D - V_G$ characteristics of lateral N^+NN^+ InGaAs-OI structures for (a) 25 nm, (b) 50 nm, (c) 100 nm and (d) 200 nm thick InGaAs film. Note the transition from full depletion (a) to partial depletion (d).

defined by the inter-contacts distance, representing the channel length L and width W .

3.1. Drain current

Transfer characteristics are reported in Fig. 3 for transistors with variable InGaAs channel thickness. All devices present a modulation of the drain current with gate voltage, the degree of which depends on thickness. For positive V_G , the field effect induces the formation of an accumulation region at the film-BOX interface, whereas for negative V_G , the film is subject to depletion. The thinner channel (Fig. 3a) shows the typical signature of full-depletion operation: very low leakage current, high ON/OFF current ratio, relatively steep subthreshold slope. These parameters tend to degrade in thicker films (Fig. 3b and c), albeit the ON current increases. The 200 nm thick InGaAs transistor (Fig. 3d) operates at the limit of partial depletion since the drain current saturates and remains high for increasingly negative gate voltage.

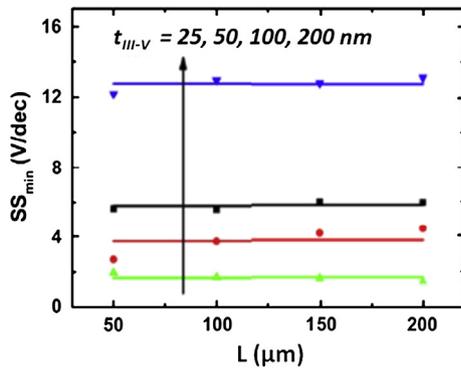


Fig. 4. Minimum subthreshold swing (SS) versus channel length (L) for devices with InGaAs film thickness of 25 nm, 50 nm, 100 nm, and 200 nm.

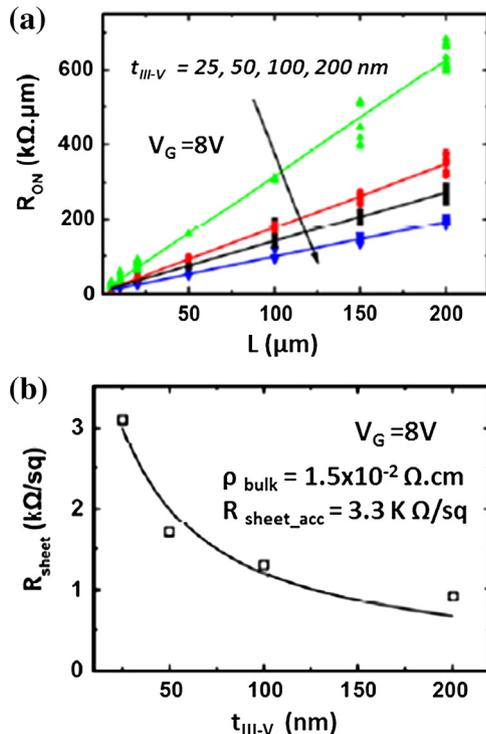


Fig. 5. (a) On-resistance R_{ON} versus channel length L_g extracted at $V_G = 8$ V and $V_{DS} = 50$ mV for four different InGaAs-OI thickness. (b) Sheet resistance R_{sheet} deduced from (a) versus InGaAs film thickness fitted by a two parallel conductance model.

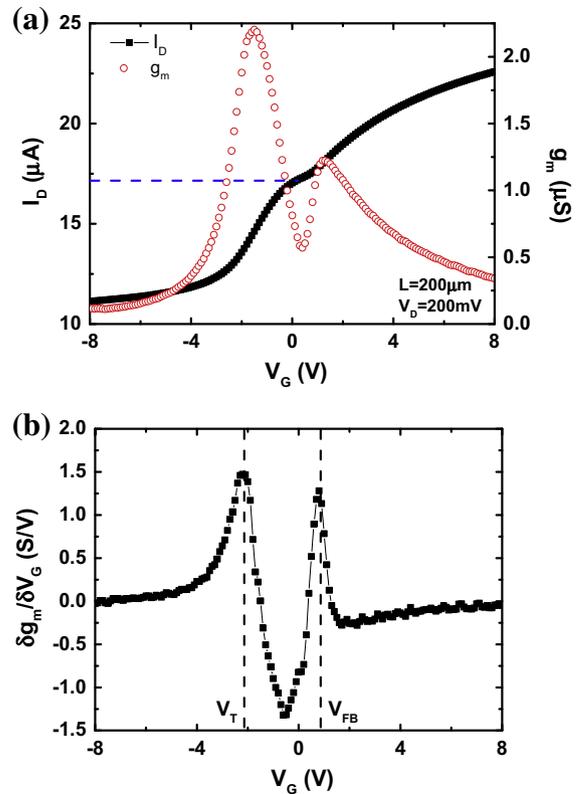


Fig. 6. (a) Drain current I_D (plain symbols) and transconductance g_m (empty symbols) versus gate bias. The dashed line represents the 'volume' current I_{vol} . (b) Corresponding g_m derivative versus V_G , which features a double threshold voltage behavior. $t_{III-V} = 200$ nm and $W = 20$ μ m.

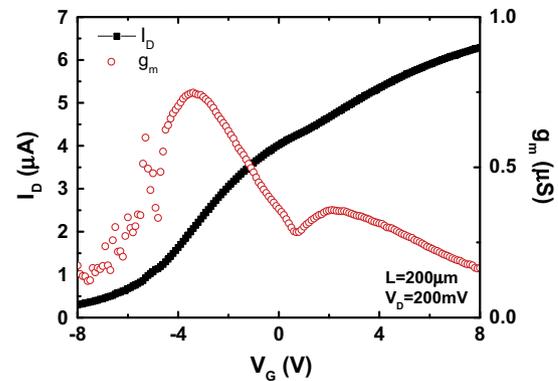


Fig. 7. Drain current I_D (plain symbols) and transconductance g_m (empty symbols) versus gate bias in 25 nm thick InGaAs film with $W = 20$ μ m.

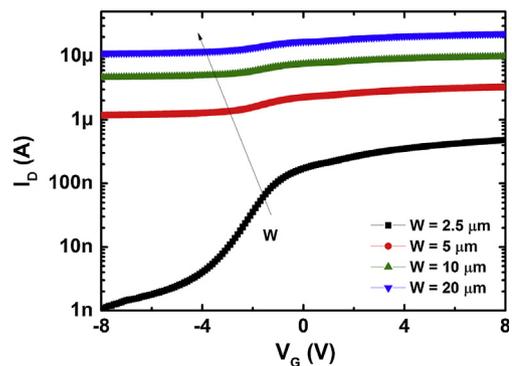


Fig. 8. Drain current I_D versus gate bias V_G in 200 nm thick devices with different widths. $L = 200$ μ m, $V_D = 200$ mV.

The gradual change from full depletion to partial depletion is also reflected by the variation of the subthreshold swing (SS) with thickness: SS increases massively in 200 nm film (Fig. 4).

Fig. 5(a) shows the ON-resistance which varies linearly with the channel length. Extrapolation of these curves to $L = 0$ indicates that all samples have a similar contact resistance (about $5 \text{ k}\Omega \mu\text{m}$), which is rather low for InGaAs-OI samples thanks to the optimized process of the source/drain terminals. The slope of the lines indicates the sheet resistance of the film (measured at $V_G = +8 \text{ V}$), which strongly depends on thickness as illustrated in Fig. 5(b). This curve can be modeled by considering two parallel conduction channels [23].

3.2. Transconductance

Fig. 6(a) shows the drain current I_D (plain symbols) and associated transconductance g_m (empty symbols) versus gate bias in thick InGaAs film (200 nm). A plateau is visible in $I_D(V_G)$ curve for $V_G \approx 0$ and the $g_m(V_G)$ characteristic features two unusual peaks. It is known that in conventional MOSFETs the derivative of the transconductance with respect to gate bias yields one peak, the position of which indicates the threshold voltage [22]. In our samples, there are two clear peaks (see Fig. 6(b)) which give evidence for a double conduction mechanism in the III-V film. Since the films are doped, volume conduction is possible even in absence of V_G . A negative gate bias induces a depletion region at the film-BOX interface, which modulates the thickness of the conduction volume. The left peak in Fig. 6(b), obtained for negative gate bias, indicates the threshold voltage V_T . This means that for $V_G > V_T$, the depletion region starts to shrink. The right peak corresponds to the flat-band voltage V_{FB} that defines the onset of the accumulation channel [12]. Beyond flat-band voltage ($V_G > V_{FB} \approx +1.1 \text{ V}$), an accumulation channel forms at the InGaAs-BOX interface and adds to the volume conduction. We call these mechanisms ‘volume’ conduction and ‘surface’ conduction.

The double-peakshape of the transconductance evolves to a single-peak in very thin films where the contribution of the neutral volume to drain current tends to vanish. For 25 nm thick film, the second peak of transconductance is hardly visible as shown in Fig. 7.

4. Mobility extraction

4.1. Partial and full depletion

Before attempting to extract the mobility, we need to evaluate whether the films are fully depleted or not. Given a doping level

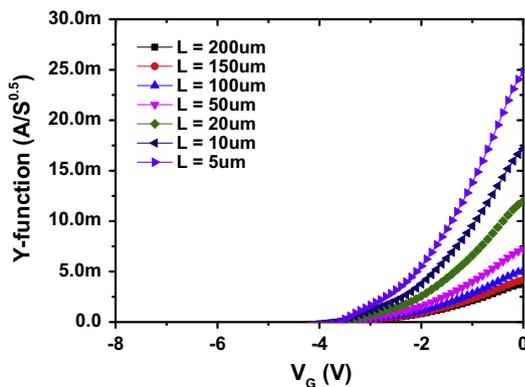


Fig. 9. Y-function in depletion regime $Y_{vol}(V_G)$ for 200 nm thick InGaAs devices with different gate lengths, $W = 10 \mu\text{m}$, $V_D = 200 \text{ mV}$.

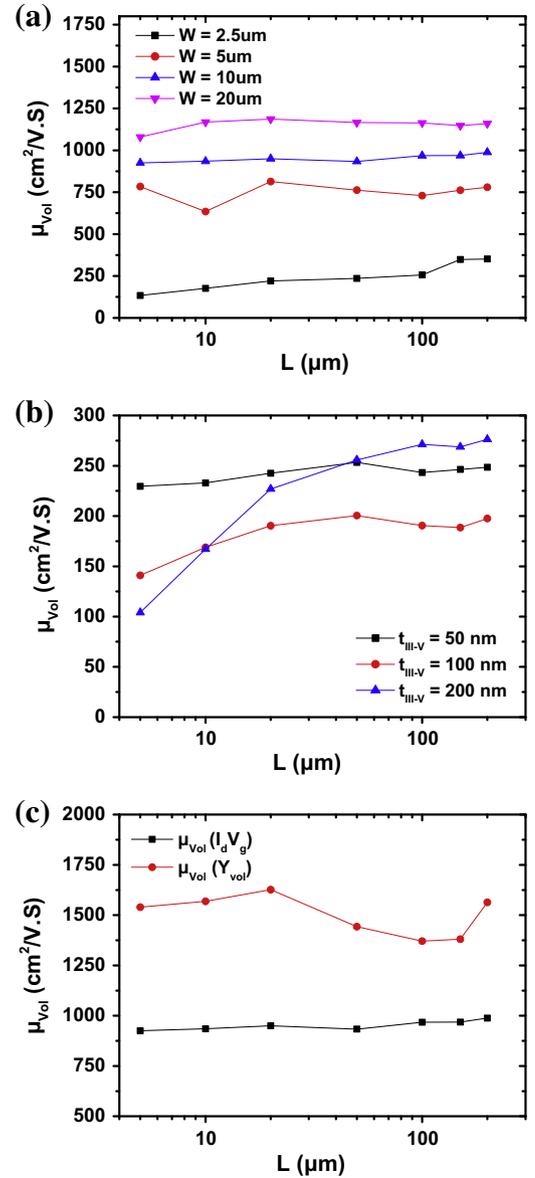


Fig. 10. Extracted volume mobility versus channel lengths in (a) 200 nm thick films with variable width and (b) narrow devices of $2.5 \mu\text{m}$ with different thickness. (c) Comparison of volume mobility values extracted from $I_D(V_G)$ curves with Eq. (2) or by using Y_{vol} -function method (Eq. (8)). $W = 10 \mu\text{m}$, $t_{III-V} = 200 \text{ nm}$, $V_D = 200 \text{ mV}$.

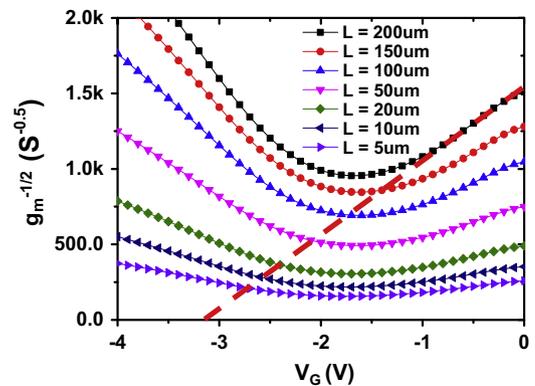


Fig. 11. Plot of $1/\sqrt{g_m}$ versus V_G in volume conduction (on right region). The dotted line shows the linear fit with Eq. (10) in order to determine coefficient θ_1 and the series resistance. $W = 10 \mu\text{m}$, $t_{III-V} = 200 \text{ nm}$.

N_D , it is possible to compute the maximum width W_{d_max} of the depletion region in the InGaAs film [24]:

$$W_{d_max} = \sqrt{\frac{4 \cdot \epsilon_{III-V} \cdot k \cdot T \cdot \ln\left(\frac{N_D}{n_i}\right)}{q^2 \cdot N_D}} \quad (1)$$

where ϵ_{III-V} and n_i are the InGaAs dielectric constant and the intrinsic doping concentration. According to the film quality, ϵ_{III-V} can vary between $13.1 \cdot \epsilon_0$ and $14.1 \cdot \epsilon_0$ [25], where ϵ_0 is the vacuum permittivity.

Considering $n_i = 6.3 \cdot 10^{11} \text{ cm}^{-3}$ [25], the maximum depletion width is between 197 nm and 205 nm, which means that films thinner than 200 nm are definitely fully depleted. In contrast, the 200 nm thick layer is at the boundary between partial and full depletion. Fig. 8 reports the measured $I_D(V_G)$ curves in thick devices with same length ($L = 200 \mu\text{m}$) and different widths. Wide transistors are clearly partially depleted with large I_D current flowing in the undepleted region of the film. But, interestingly, the narrower device ($W = 2.5 \mu\text{m}$) exhibits full depletion, suggesting a 2D mechanism [26]: the lateral depletion, assisted by defects and charges located on the sidewalls, reinforces the vertical depletion induced by the gate.

4.2. Volume mobility

According to the approach developed by Liu et al. [27] for heavily doped Si layers, the drain current in volume conduction regime can be written as:

$$I_D = \frac{W}{L} \cdot C_{BOX} \cdot \mu_{Vol} \cdot V_D \cdot (V_G - V_0) \quad (2)$$

where C_{BOX} is the buried oxide capacitance [19] and V_0 is a characteristic voltage which may enable full depletion of the channel [27]:

$$V_0 = V_{FB} + \frac{q \cdot N_D}{C_{BOX}} \cdot t_{III-V} - V \quad (3)$$

N_d is the active doping concentration. Eq. (2) accounts for the thickness variation of the neutral (undepleted) region with gate bias. In highly doped, partially depleted films, V_0 is hypothetical, as it cannot be reached experimentally; it is extrapolated from the measurements. In principle, the volume mobility μ_{Vol} is independent on gate bias unless the film has to be inhomogeneous with vertical profiles of mobility, defects and dopants. It may also depend on the vertical field induced by surface charges, in particular when the neutral region becomes very thin. The device ON-resistance includes the contribution of the series source/drain resistance R_{SD} :

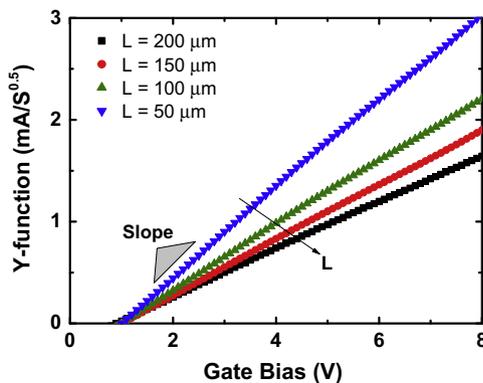


Fig. 12. Y_S -function in accumulation regime $Y_S(V_G)$ for $W = 20 \mu\text{m}$, $t_{III-V} = 200 \text{ nm}$ thick InGaAs devices with different gate lengths.

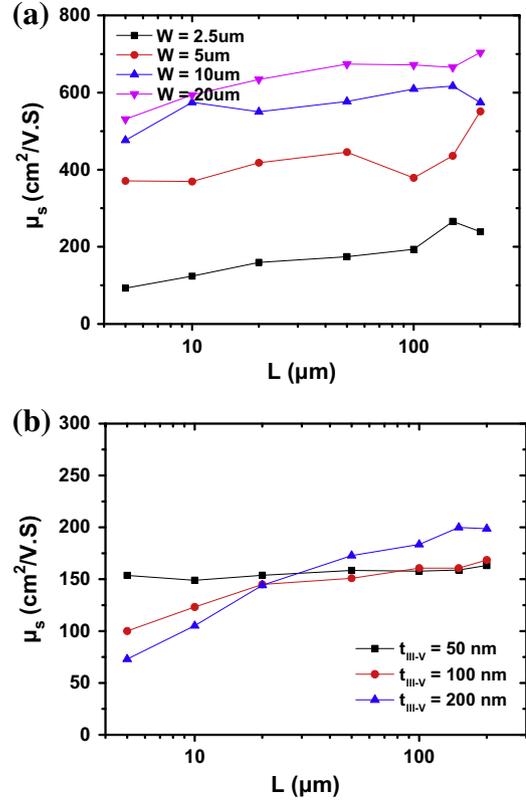


Fig. 13. Extracted surface mobility versus gate length in (a) 200 nm thick devices with different widths, and (b) narrow devices with variable thickness.

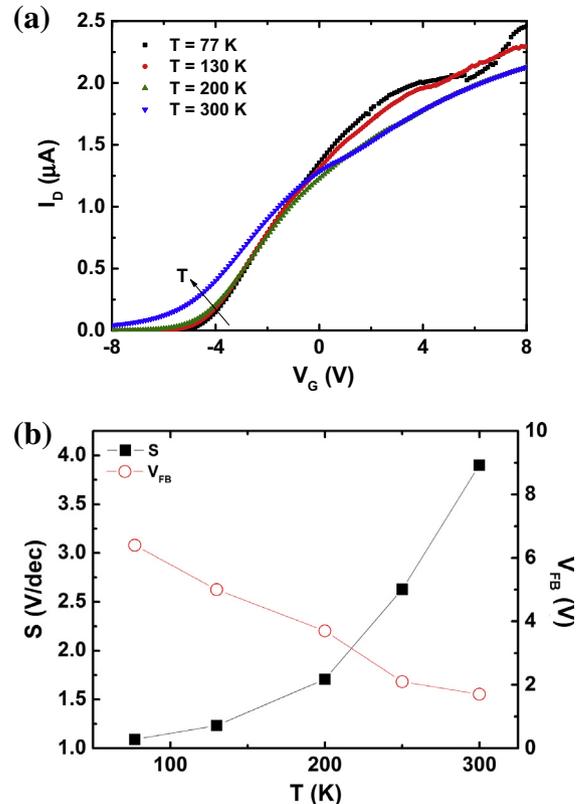


Fig. 14. (a) Drain current as a function of gate bias measured at variable temperature. (b) Associated variations of subthreshold swing S and flat-band voltage V_{FB} with low temperature T . For $t_{III-V} = 25 \text{ nm}$, $W = 5 \mu\text{m}$, and $L = 100 \mu\text{m}$ of InGaAs device.

$$R_{ON} = \frac{V_D}{I_D} = R_{SD} + \frac{L}{W \cdot C_{BOX} \cdot \mu_{Vol} \cdot (V_G - V_0)} \quad (4)$$

Rewriting I_D , we obtain:

$$I_D = \frac{W}{L} \cdot C_{BOX} \cdot \frac{\mu_{Vol}}{1 + \theta_1(V_G - V_0)} \cdot (V_G - V_0) \cdot V_D \quad (5)$$

where

$$\theta_1 = R_{SD} \cdot \mu_{Vol} \cdot C_{BOX} \cdot \frac{W}{L} \quad (6)$$

The transconductance is:

$$g_m = \frac{\partial I_D}{\partial V_G} = \frac{W}{L} \cdot C_{BOX} \cdot \frac{\mu_{Vol}}{[1 + \theta_1(V_G - V_0)]^2} \cdot V_D \quad (7)$$

This equation explains the transconductance variation with gate voltage in volume conduction regime. Since the difference $(V_G - V_0)$ can be very significant, especially for thick films, the transconductance decay with V_G is abrupt, as shown in Fig. 6(a). It follows that the volume mobility extracted from either Eq. (2) or transconductance peak is highly underestimated due to the series resistance effect. This problem can be circumvented by constructing a Y-function:

$$Y_{Vol} = \frac{I_{Vol}}{\sqrt{g_m}} = \sqrt{\frac{W}{L} \cdot C_{BOX} \cdot \mu_{Vol} \cdot V_D \cdot (V_G - V_0)} \quad (8)$$

where the impact of R_{SD} is erased.

Drawing Y_{Vol} versus V_G results in a straight line region, more pronounced in long devices (Fig. 9). The intercept with the horizontal axis yields V_0 , and the slope gives the volume mobility:

$$\mu_{Vol} = \frac{\text{Slope}^2}{\frac{W}{L} \cdot C_{BOX} \cdot V_D} \quad (9)$$

This methodology is similar to the Y-function technique used in inversion-mode MOSFETs [28] except that μ_{Vol} and V_0 have different meaning.

The extracted μ_{Vol} (Fig. 10(a)) is rather constant with the gate length L . The volume mobility reaches $1200 \text{ cm}^2/\text{Vs}$ in thick and wide devices (Fig. 10(a)). Net mobility degradation is noticed in 25–50 nm thick films and also in narrow devices (Fig. 10(a) and (b)), where carrier scattering on the sidewalls comes into play. Fig. 10(c) compares the volume mobility extracted from $I_D(V_G)$ curves with Eq. (2) and from Y_{Vol} -function. The latter method is not affected by source and drain series resistance and provides at least 40% higher mobility. The series resistance is included in coefficient (see Eq. (6)) which can be determined from the slope of $1/\sqrt{g_m}$ versus V_G curve:

$$\frac{1}{\sqrt{g_m}(V_G)} = \frac{1 + \theta_1 \cdot (V_G - V_0)}{\sqrt{C_{BOX} \cdot \frac{W}{L} \cdot \mu_{Vol} \cdot V_D}} \quad (10)$$

Fig. 11 confirms the linear variation of $1/\sqrt{g_m}(V_G)$ curves in volume conduction mode. Using this method (Eq. (10)), the extracted coefficient θ_1 is around 0.5 V^{-1} for all devices, which implies that the series resistance increases in longer channels.

4.3. Surface mobility

Surface conduction occurs for $V_G > V_{FB}$, when the accumulation channel starts forming at the film-BOX interface and adding to the volume current flowing in the whole neutral film. The surface current is given by the difference between the total current I_D and the maximum volume current $I_{Vol,m}$ measured at $V_G = V_{FB}$ (see dashed line in Fig. 6(a)). The surface mobility μ_S at the film-BOX interface is obtained from another version of the Y-function (Y_S) adapted to the accumulation current [28]:

$$Y_S = \frac{I_D - I_{Vol}}{\sqrt{g_m}} = \sqrt{\frac{W}{L} \cdot C_{BOX} \cdot \mu_S \cdot V_D \cdot (V_G - V_{FB})} \quad (11)$$

The curves $Y_S(V_G)$ are linear as shown in Fig. 12. The flat-band voltage V_{FB} is determined from the intercept with the gate voltage axis and the surface mobility in the accumulation channel from the slope, as indicated in Eq. (11).

Fig. 13 shows the surface mobility μ_S versus gate length. The surface mobility is systematically lower than the volume mobility μ_{Vol} by a factor of two due to stronger field-effect and additional carrier scattering at the interface film-BOX [7]. μ_S is superior in thick films (100–200 nm) where the coupling between top surface defects and bottom channel is attenuated. The quality of the InGaAs-BOX interface is assumed constant given the stability of the fabrication process. The lowest mobility is measured for 25 nm thick devices and is attributed to the charged defects exist in the top surface. The difference between the front-surface and back-surface potentials induces a vertical field, $E_{int} = (\Psi_{S1} - \Psi_{S2})/t_{III-V}$, that obviously increases in thinner films [29]. Even if the influence of the gate-induced field is removed in Y-function, the mobility is still affected by the intrinsic field. In other words, the genuine low-field mobility is not accessible.

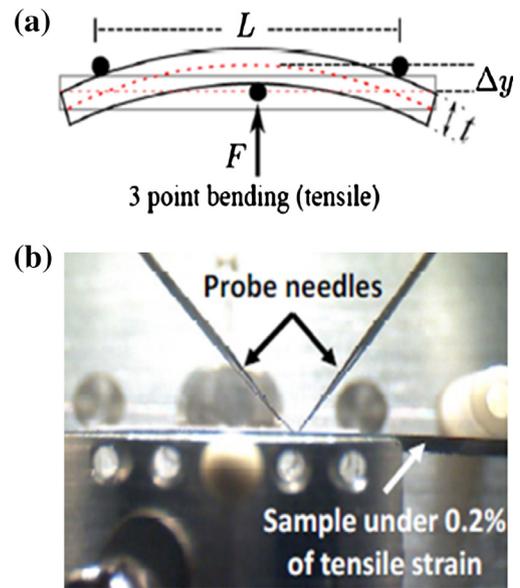


Fig. 15. (a) Schematic of 3-point bending configuration used to apply uniaxial tensile strain on the sample. (b) Picture from the side of the 3-point bending setup showing the semiconductor layer under tensile strain and the contact probes.

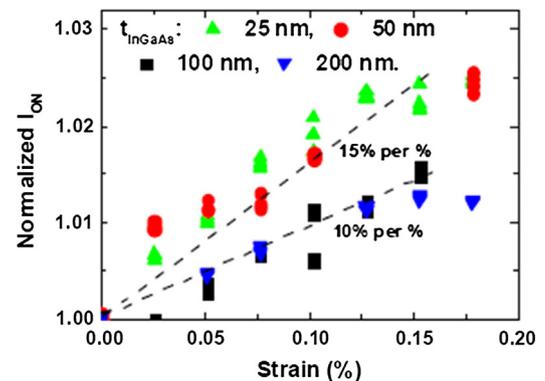


Fig. 16. Improvement rate of I_{ON} (at $V_g = 8 \text{ V}$) versus uniaxial tensile strain for different InGaAs-OI thicknesses.

4.4. Low-temperature measurements

Fig. 14(a) shows the drain current versus gate bias characteristics measured at different temperatures T . As reported for junctionless devices [30], the drain current in accumulation regime increases at lower temperature primarily because the mobility is improved (reduced phonon scattering). At the same time, the flat-band voltage (empty symbols in Fig. 14(b)) increases which delays the formation of the accumulation channel. Fig. 14(b) also presents the subthreshold swing S (plain symbols) versus T . A linear decrease of S with temperature is found between 200 K and 300 K, already documented for undoped MOSFETs [5]. The swing tends to saturate below 150 K, presumably due to the increase of the effective density of traps.

4.5. Strain effect

Samples are diced in 2 cm \times 3 mm pieces and mounted in a 3-point bending setup (Fig. 15) where uniaxial tensile strain can be applied up to 0.2% without breaking the samples. The ON-current I_{ON} benefits from tensile strain applied along the $\langle 110 \rangle$ transport direction and increases at a rate of 10–15% per percent of strain (Fig. 16).

5. Conclusions

We have fabricated and characterized back-gated accumulation layer MOSFETs on InGaAs-on-Insulator wafers. These simple test structures are intended to provide a quick feedback for InGaAs technology optimization. Detailed measurements point out the coexistence of two conduction mechanisms (volume and interface accumulation). The plateau present in $I_D(V_G)$ curves and the double-peak transconductance reveal the transition from volume conduction to surface-dominated conduction. A methodology was proposed for the extraction of carrier mobility in the neutral volume and at the film-BOX interface. The volume mobility is always higher than the surface mobility and exceeds 1100 cm²/Vs in 100–200 nm thick InGaAs films. The lowest mobility was measured on very thin (25 nm) layers which suffer from strong coupling between top interface defects and conduction channel. The InGaAs material properties were also investigated as a function of aspect ratio and low-temperature operation. Finally, uniaxial tensile strain in the $\langle 110 \rangle$ transport direction highlighted the improvement the on-current with a rate of 10–15% per percent of strain.

Acknowledgements

This work is supported by European Union through the projects COMPOSE³ (FP7-ICT-11-619325) and III-V-MOS (FP7-ICT-11-619326).

References

- [1] del Alamo JA. Nanometre-scale electronics with III-V compound semiconductors. *Nature* 2011;479(7373):317–23.
- [2] Egard M, Ohlsson L, Borg BM, Lenrick F, Wallenberg R, Wernersson L-E, et al. High transconductance self-aligned gate-last surface channel In_{0.53}Ga_{0.47}As MOSFET. In: *Electron Devices Meeting (IEDM), IEEE International. IEEE International*; 2011. p. 13.2.1–4.
- [3] Gu JJ, Wang XW, Wu H, Shao J, Neal AT, Manfra MJ, et al. 20–80 nm Channel length InGaAs gate-all-around nanowire MOSFETs with EOT = 1.2 nm and lowest SS = 63 mV/dec. In: *Electron Devices Meeting (IEDM), IEEE International. IEEE International*; 2012. p. 27.6.1–4.
- [4] Banna SR, Mizuno T, Takagi S, Sugiyama N, Satake H, Kurobe A. Electron and hole mobility enhancement in strained-Si MOSFETs on SiGe-on-insulator substrates fabricated by SIMOX technology. *A Toriumi Electron Dev Lett, IEEE* 2000;21(5):230–2.
- [5] del Alamo JA, Antoniadis DA, Lin J, Lu W, Vardi A, Zhao X. III-V MOSFETs for future CMOS. In: *Compound Semiconductor Integrated Circuit Symposium (CSICS). IEEE*; 2015. p. 1–4.
- [6] ITRS 2013. *International Technology Roadmap for Semiconductors, 2013 Update Overview*.
- [7] Daix N, Uccelli E, Czornomaz L, Caimi D, Rossel C, Sousa M, et al. Towards large size substrates for III-V co-integration made by direct wafer bonding on Si. *APL Mater* 2014;2(8):086104.
- [8] Yokoyama M, Yasuda T, Takagi H, Miyata N, Urabe Y, Ishii H, et al. III-V-semiconductor-on-insulator n-channel metal-insulator-semiconductor field-effect transistors with buried Al₂O₃ layers and sulfur passivation: reduction in carrier scattering at the bottom interface. *Appl. Phys. Lett.* 2010;96(14):142106.
- [9] Radosavljevic M, Dewey G, Basu D, Boardman J, Chu-Kung B, Fastenau JM, et al. Electrostatics improvement in 3-D tri-gate over ultra-thin body planar InGaAs quantum well field effect transistors with high-K gate dielectric and scaled gate-to-drain/gate-to-source separation. In: *Electron Devices Meeting (IEDM), IEEE International*. p. 33.1.1–4.
- [10] Fiorenza JG, Braithwaite G, Leitz CW, Currie MT, Yap J, Singaporewala F, et al. Film thickness constraints for manufacturable strained silicon CMOS. *Semicond Sci Technol* 2004;19(1):L4–8.
- [11] Banna SR, Chan PCH, Chan M, Fung SKH, Ko PK. Fully depleted CMOS/SOI device design guidelines for low-power applications. *IEEE Trans Electron Dev* 1999;46(4):754–61.
- [12] Cristoloveanu S, Li S. *Electrical characterization of silicon-on-insulator materials and devices*. Springer Science & Business Media; 1995.
- [13] Czornomaz L, Daix N, Caimi D, Sousa M, Erni R, Rossel M, et al. An integration path for gate-first UTB III-V-on-insulator MOSFETs with silicon, using direct wafer bonding and donor wafer recycling. In: *Electron Devices Meeting (IEDM), 2012 IEEE International*. p. 23.4.1.
- [14] Djara V, Deshpande V, Uccelli E, Daix N, Caimi D, Rossel C, et al. An InGaAs on Si-platform for CMOS with 200 mm InGaAs-OI substrate, gate-first, replacement gate planar and FinFETs down to 120 nm contact pitch. In: *VLSI technology (VLSI Technology), 2015 symposium on*. p. T176–7.
- [15] Czornomaz L, Daix N, Cheng K, Caimi D, Rossel C, Lister K, et al. Cointegration of InGaAs n- and SiGe p-MOSFETs into digital CMOS circuits using hybrid dual-channel ETXOI substrates. In: *Electron Devices Meeting (IEDM), 2013 IEEE International*. p. 2.8.1–4.
- [16] Waldron N, Merckling C, Guo W, Ong P, Teugels L, Ansar S, et al. An InGaAs/InP quantum well finfet using the replacement fin process integrated in an RMG flow on 300 mm Si substrates. In: *VLSI technology (VLSI-Technology): digest of technical papers*. p. 1–2.
- [17] Lee S, Chobpattana V, Huang C-Y, Thibeault BJ, Mitchell W, Stemmer S, et al. Record ION (0.5 mA/μm at VDD = 0.5 V and I_{off} = 100 nA/μm) 25 nm-gate-length ZrO₂/InAs/InAlAs MOSFETs. In: *VLSI technology (VLSI technology), 2014 symposium on*. p. 1–2.
- [18] Cristoloveanu S, Williams S. Point-contact pseudo-MOSFET for in-situ characterization of as-grown silicon-on-insulator wafers. *IEEE Electron Dev Lett* 1992;13(2):102–4.
- [19] Czornomaz L, Djara V, Deshpande V, Caimi D, Pirro L, Cristoloveanu S, et al. *IEEE conf. EuroSOI&ULIS*; 2016. p. 104–7.
- [20] Czornomaz L, Daix N, Kerber P, Lister K, Caimi D, Rossel C, et al. Scalability of ultra-thin-body and BOX InGaAs MOSFETs on silicon. In: *Solid-state device research conference (ESSDERC), 2013 proceedings of the European*. p. 143–6.
- [21] Deshpande V, Djara V, O'Connor E, Hashemi P, Balakrishnan K, Sousa M, et al. Advanced 3D monolithic hybrid CMOS with sub-50 nm gate inverters featuring replacement metal gate (RMG)-InGaAs nFETs on SiGe-OI Fin pFETs. In: *Electron devices meeting (IEDM), 2015 IEEE International*. p. 8.8.1–4.
- [22] Cristoloveanu S, Bawedin M, Ionica I. A review of electrical characterization techniques for ultrathin FDSOI materials and devices. *Solid-State Electron* 2016;117:10–36.
- [23] Pirro L, Czornomaz L, Park HJ, Ionica I, Bawedin M, Djara V, et al. Volume and interface conduction in InGaAs junctionless transistors. In: *Ultimate integration on silicon (EUROSOL-ULIS), 2016 Joint international EUROSOL workshop and international conference*.
- [24] Sze SM, Ng KK. *Physics of semiconductor devices*. John Wiley & Sons; 2006.
- [25] Schubert EF. *Doping in III-V semiconductors*. Cambridge University Press; 2005.
- [26] Allibert F, Pretet J, Pananakakis G, Cristoloveanu S. Transition from partial to full depletion in silicon-on-insulator transistors: impact of channel length. *Appl Phys Lett* 2004;84(7):1192–4.
- [27] Liu FY, Diab A, Ionica I, Akarvardar K, Hobbs C, Ouisse T, et al. Characterization of heavily doped SOI wafers under pseudo-MOSFET configuration. *Solid-State Electron* 2013;90:65–72.
- [28] Ghibaudo G. New method for the extraction of MOSFET parameters. *Electron Lett* 1988;24:543–5.
- [29] Hamaide G, Allibert F, Andrieu F, Romanjek K, Cristoloveanu S. Mobility in ultrathin SOI MOSFET and pseudo-MOSFET: impact of the potential at both interfaces. *Solid-State Electron* 2011;57(1):83–6.
- [30] Filanovsky IM, Allam A. Mutual compensation of mobility and threshold voltage temperature effects with applications in CMOS circuits. *IEEE Trans Circuits Syst Fundam Theory Appl* 2001;48(7):876–84.



DC and RF characterization of InGaAs replacement metal gate (RMG) nFETs on SiGe-OI FinFETs fabricated by 3D monolithic integration



V. Deshpande^{a,*}, V. Djara^a, E. O'Connor^a, P. Hashemi^b, K. Balakrishnan^b, D. Caimi^a, M. Sousa^a, L. Czornomaz^a, J. Fompeyrine^a

^a IBM Zurich Research Laboratory, Säumerstrasse 4, CH-8803 Rüschlikon, Switzerland

^b IBM T.J. Watson Research Center, 1101 Kitchawan Rd., Route 134, Yorktown Heights, NY, USA

ARTICLE INFO

Article history:

Available online 25 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

3D monolithic

InGaAs

RMG

High-frequency

FinFET

ABSTRACT

We report the first RF characterization of short-channel replacement metal gate (RMG) InGaAs-OI nFETs built in a 3D monolithic (3DM) CMOS process. This process features RMG InGaAs-OI nFET top layer and SiGe-OI fin pFET bottom layer. We demonstrate state-of-the-art device integration on both levels. The bottom layer SiGe-OI pFETs are fabricated with a Gate-First (GF) process with fins and featuring epitaxial raised source drain (RSD) as well as silicide contact layer. The top layer InGaAs nFETs are fabricated with a RMG process featuring a self-aligned epitaxial raised source drain (RSD). We show that the 3D monolithic integration scheme does not degrade the performance of the bottom SiGe-OI pFETs owing to an optimized thermal budget for the top InGaAs nFETs. From the RF characterizations performed (post-3D monolithic process) on multifinger-gate InGaAs-OI nFETs, we extract a cut-off frequency (f_t) of 16.4 GHz at a gate-length (L_g) of 120 nm. Measurements on various gate lengths shows increasing cut-off frequency with decreasing gate-length.

© 2016 Published by Elsevier Ltd.

1. Introduction

3D monolithic (3DM) integration is attracting much attention owing to density scaling benefits and the potential to stack independently optimized multifunctional layers at transistor level [1]. However, due to the inherently high thermal budget of Si MOSFET process, Si(Ge)-on-Si 3DM integration scheme faces major challenges in top layer optimization without degrading bottom layer performance. This necessitates development of low temperature top layer Si/SiGe process which presents further challenges to obtaining high-performance MOSFETs on top layer. As the InGaAs MOSFET processing thermal budget is significantly lower, it is well-suited to be used as the top layer channel material. Moreover, InGaAs also has higher mobility which enables high performance at lower voltages [2,3] and is an excellent channel material for high-frequency devices enabling very high cut-off frequency [4,5]. This aspect is of great interest as it can enable a truly multifunctional 3D monolithic integration scheme. On one hand, InGaAs nFETs on top of Si/SiGe FETs can allow higher performance hybrid CMOS [6] and on the other, high frequency InGaAs RF-FETs can benefit from closely integrated CMOS circuits [7]. As a step towards

such a multi-functional 3D monolithic integration, here, we show DC and RF characteristics of InGaAs nFETs fabricated with RMG process on top of SiGe-OI finFETs. We perform RF characterization of InGaAs-OI nFETs of various gate-lengths designed with optimized ‘multi-finger gate’ layout to enable efficient characterization. We demonstrate a cut-off frequency of 16.4 GHz for gate length (L_g) of 120 nm. Also, cut-off frequency is shown to increase with decreasing gate-lengths. We also demonstrate that the impact of 3DM integration on the bottom pFET performance is negligible, despite the top nFET RMG process featuring a self-aligned raised source drain epitaxy (with relatively high thermal budget). Thus we demonstrate a robust InGaAs-on-SiGe 3DM integration scheme that can enable multifunctional circuits involving dense CMOS and high-frequency InGaAs nFETs.

2. Device fabrication

The steps involved in the InGaAs-on-SiGe 3D monolithic process flow are shown in Fig. 1.

Firstly, bottom layer SiGe-OI fin pFETs are fabricated with a gate-first (GF) process as described in [8,9]. The process begins with thinning of silicon layer of an 8 in. SOI wafer followed by Ge condensation to obtain SiGe-OI layer (with 25% Ge content). Then active pFET areas are patterned. This is then followed by gate

* Corresponding author.

E-mail address: VEE@zurich.ibm.com (V. Deshpande).

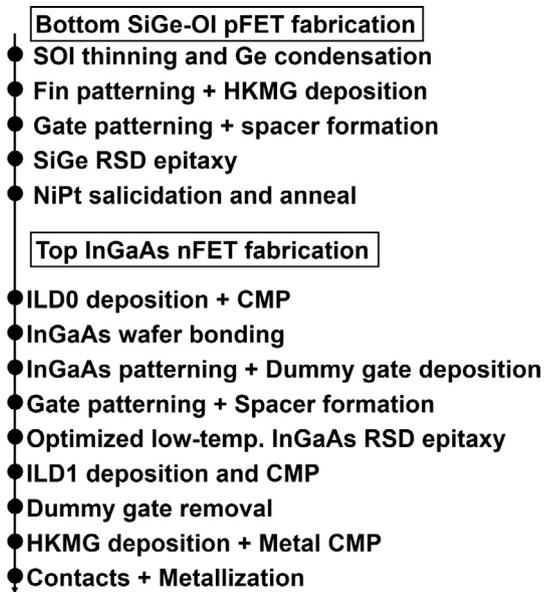


Fig. 1. InGaAs-on-SiGe 3D monolithic integration process flow.

stack deposition as typical of a gate-first process flow. Thereafter, the gate lithography and gate etching is performed. After gate patterning, spacer deposition is carried out. Then the spacers are formed with anisotropic dry etching. After this step, in situ 'p type' doped SiGe epitaxy is carried out to form self-aligned raised source drain (RSD) regions. Then NiPt salicidation (self-aligned silicidation) is performed to obtain low contact resistivity on the pFETs. This silicide sets the thermal budget limit for the top nFET processing as high temperature can lead to degradation of silicide resistivity impacting the bottom FET performance [10]. The top layer nFET fabrication starts after this silicidation step of SiGe-OI finFET process. It begins with InGaAs layer transfer on top of processed pFETs through direct wafer bonding [11]. To achieve this, first an inter-layer oxide (ILD0 in Fig. 2) is deposited and chemical-mechanical-polish (CMP) planarization is carried out. The InGaAs layer is transferred on to this oxide with direct wafer bonding from 2 in. InP donor wafers [11]. The thickness of the transferred InGaAs layer is 20 nm. Owing to direct wafer bonding technique, a planar continuous InGaAs layer is obtained on top of the ILD0. Therefore, this layer can be patterned in any desired structures with same alignment accuracy of the lithographic process of the bottom pFET layer. After transferring the InGaAs layer, the top InGaAs nFET fabrication is performed with a replacement metal gate (RMG) process [12,13]. The process starts with the patterning of InGaAs layer to form active transistor mesa regions either as planar areas or

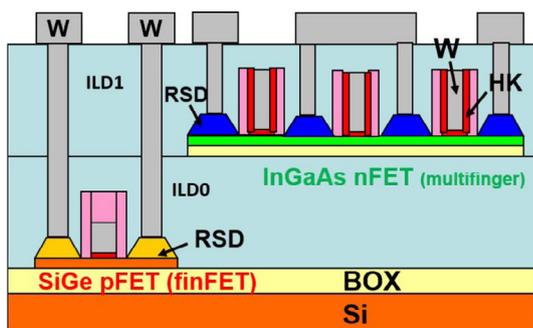


Fig. 2. Schematic of the InGaAs-on-SiGe 3DM stack showing multi-finger gate nFETs on top layer.

wide-fins (40–100 nm width). After forming the active areas, a dummy gate stack deposition is done. The dummy gate is then patterned. Spacers are formed on either side of the dummy gate in a similar way as in the bottom pFET process. It is then followed by self-aligned in situ 'n type' doped InGaAs epitaxy to form raised source-drain (RSD) regions. In general, this step is a relatively high thermal budget process and can be detrimental to bottom pFET performance. Therefore, it has been optimized to minimize the process temperature while obtaining high doping in the layer [6]. After the RSD process, the dummy gate replacement process steps are carried out. An encapsulation oxide layer (ILD1) is first deposited and planarized with CMP process to expose the top of dummy gate. Then the dummy gate stack is selectively etched out exposing the InGaAs channel in the region where the dummy gate was present. An optimized high- k /metal gate (HKMG) stack featuring a scaled $\text{Al}_2\text{O}_3/\text{HfO}_2$ dielectric [14] is then deposited. This stack forms the actual gate oxide and gate metal on top of the InGaAs channel. The HKMG stack is immediately capped with W metal. After this deposition, the gate metal (W) is planarized again with CMP process to the same level as the encapsulation oxide. Finally, another oxide encapsulation is deposited and contact holes are opened to both top nFET and bottom pFET source/drain and gate regions. Metallization is completed to create contact pads for both layers. After completion of the contacts and metallization, the devices are annealed in H_2/Ar ambient for optimizing the gate stack in terms of interface state density (D_{it}) [15]. The schematic of the so completed 3D monolithic stack is shown in Fig. 2 and a cross section STEM image of the 3D monolithic stack taken along the gate is shown in Fig. 3. The InGaAs fins with RMG can be seen on the top layer and SiGe-OI fins are visible on the bottom layer.

3. Electrical characterization

In this section, first the DC characteristics of the InGaAs nFETs is discussed, followed by the RF characteristics. The contact pads are

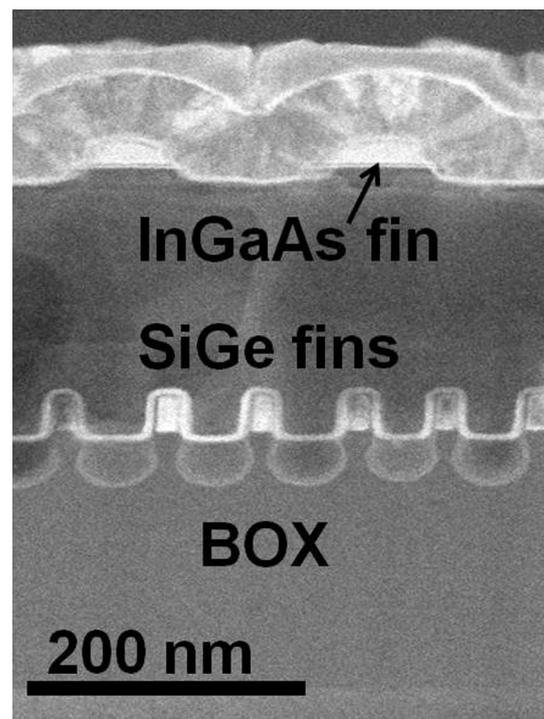


Fig. 3. STEM image of the InGaAs-on-SiGe 3DM stack showing nFETs on top of SiGe finFETs.

designed in coplanar waveguide (CPW) structure and can be measured with standard Ground-Signal-Ground (GSG) probes. The pad layout and the transistor layout featuring multi-finger gates is shown in Fig. 4.

The multi-finger gate design allows lowering the gate-resistance owing to multiple parallel gates. Furthermore, the nFETs characterized in this work have dense features with gate-to-contact spacing of 100 nm. The DC $I_d - V_g$ characteristics of an InGaAs planar nFET with $L_g = 120$ nm are shown in Fig. 5.

This device features 10 finger gates in parallel. DC characteristics show competitive electrostatic control with drain-induced-barrier-lowering (DIBL) of 104 mV/V and $SS_{sat} = 100$ mV/dec due a scaled high- k gate stack with a CET of 1.6 nm. Fig. 6 shows the DC $G_m - V_g$ characteristics for the same device. A peak G_m of about 1.7 mS is obtained at $V_d = 0.5$ V and $V_g = 0.65$ V. This value of V_g is used for the device for subsequent characterization to estimate cut-off frequency. Fig. 7 shows the $I_d - V_d$ characteristics for the same device. Fig. 8 shows the DC $G_{ds} - V_g$ characteristics. The drain-induced-barrier-lowering (DIBL) vs. L_g and saturation subthreshold slope (SS_{sat}) vs. L_g curves are shown in Figs. 9 and 10 respectively. Owing to low CET, competitive sub-threshold characteristics are observed on the nFETs. However, devices on the same chip with ‘single gate finger’ layout presented in Ref. [6] show relatively better electrostatics. This could be due to smaller spacer width in multi-finger gate devices, as the fabrication process is not fully optimized for dense multi-finger layout. A comparison of the $I_d - V_g$ characteristics bottom layer SiGe-OI finFET before and after top nFET fabrication is shown in Fig. 11. The figure shows the $I_d - V_g$ characteristics of a SiGe-OI pFET with $L_g = 36$ nm and fin width 15 nm, before (dash) and after (solid) top nFET fabrication. Due to the lower thermal budget of the top layer InGaAs process, very minimal impact is observed on the bottom pFET even for a scaled gate length. Nearly the same drain current (I_d) is maintained at both linear and saturation regime in the pFET indicating no degradation of the bottom silicide. Therefore, it is seen that our integration scheme allows obtaining competitive DC performance for top InGaAs nFETs without degrading the performance of the bottom pFETs.

Now, RF characterization of the top nFET is discussed. The RF characterization on nFETs is performed on devices with 10 parallel finger gates, each with a width of 2 μm (= total device width of 20 μm), and having a ground-signal-ground (GSG) pad configura-

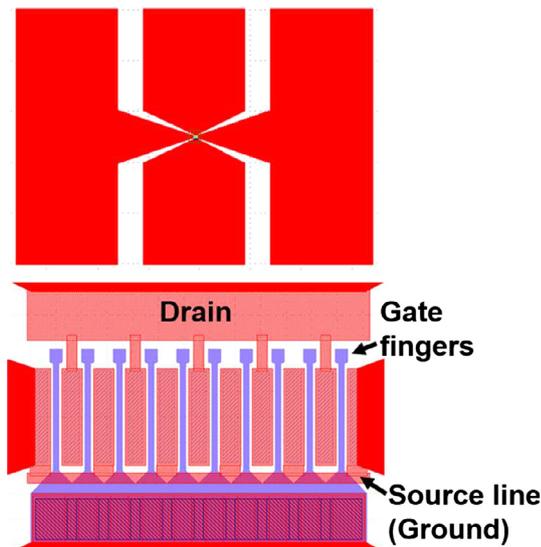


Fig. 4. CPW pad structure (top) and multi-finger gate layout (bottom) of the InGaAs nFET used for RF characterization.

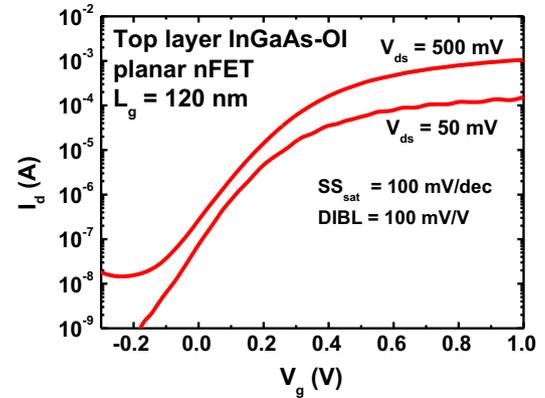


Fig. 5. DC $I_d - V_g$ characteristics of top layer InGaAs-OI planar nFET with $L_g = 120$ nm.

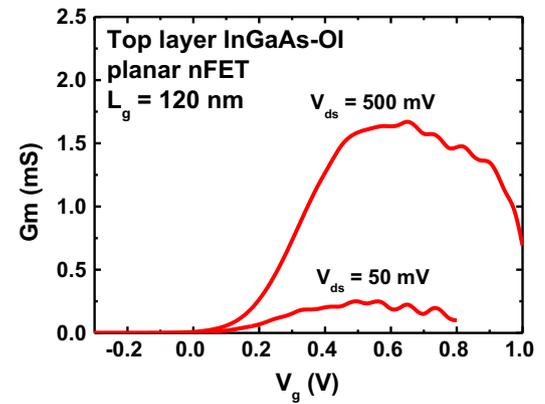


Fig. 6. DC $G_m - V_g$ characteristics of top layer InGaAs-OI planar nFET with $L_g = 120$ nm.

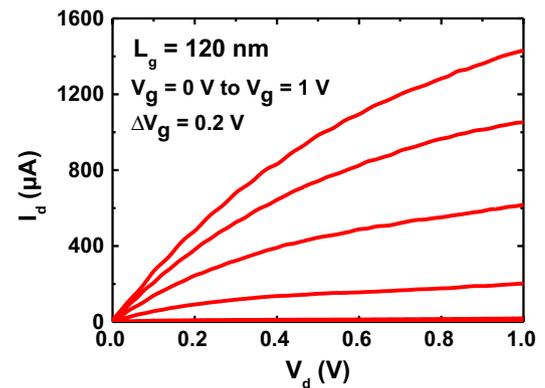


Fig. 7. DC $I_d - V_d$ characteristics of InGaAs-OI top layer planar nFET with $L_g = 120$ nm.

tion. A LRRM calibration with a vector-network-analyzer (VNA) is first carried out on a commercial standard reference calibration substrate, to move the reference plane to probe tips. Dedicated on-chip ‘open’ pad structures are used to de-embed the device. The ‘open’ pad structures are similar to the pad layout shown in Fig. 4, without the MOSFET in the center. S-parameters are measured from 45 MHz to 40 GHz. First, S parameters of the device are measured at V_g which corresponds to the peak in DC transconductance of the MOSFET. Then S parameters are converted to Y parameters (denoted Y_{TOTAL}). Then the S parameters of the ‘open’

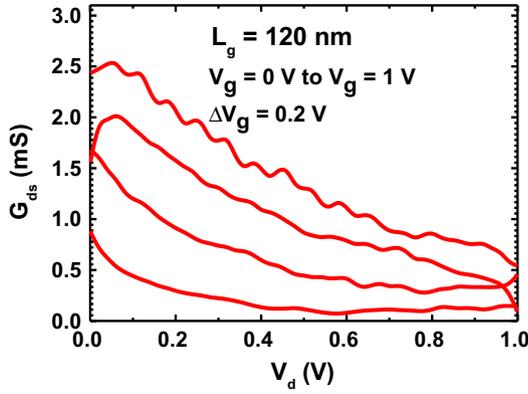


Fig. 8. DC $G_{ds} - V_d$ characteristics of InGaAs-OI top layer planar nFET with $L_g = 120$ nm.

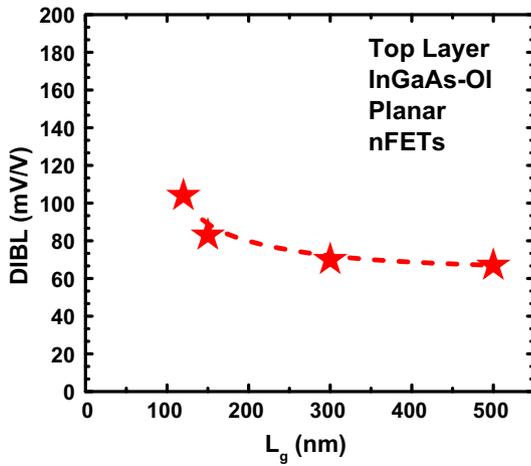


Fig. 9. $DIBL - L_g$ characteristics of InGaAs-OI top layer planar nFETs.

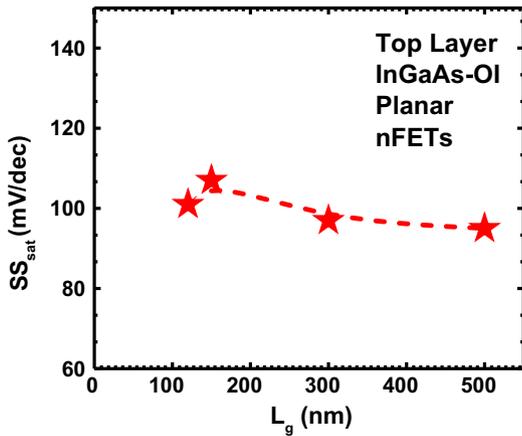


Fig. 10. $SS_{sat} - L_g$ characteristics of InGaAs-OI top layer planar nFETs.

pad structure are measured and converted to Y parameters (denoted Y_{OPEN}). For de-embedding, the Y parameters of 'open' structure are subtracted from the Y parameters of the MOSFET. Then the current gain ($|h_{21}|$) is calculated with equation below:

$$|h_{21}| = \frac{|Y_{21_{MOSFET}}|}{|Y_{11_{MOSFET}}|} \quad (1)$$

where

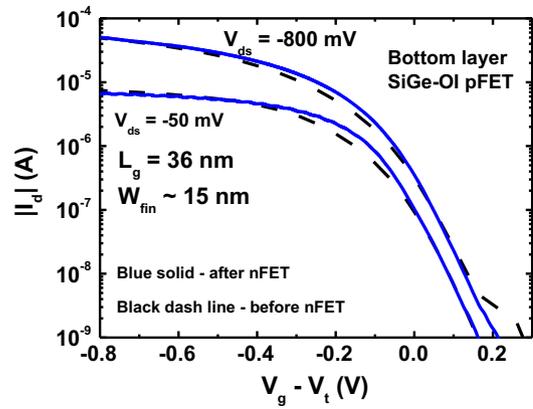


Fig. 11. Comparison of $I_d - V_g$ characteristics of bottom layer SiGe-OI pFET before and after top nFET fabrication.

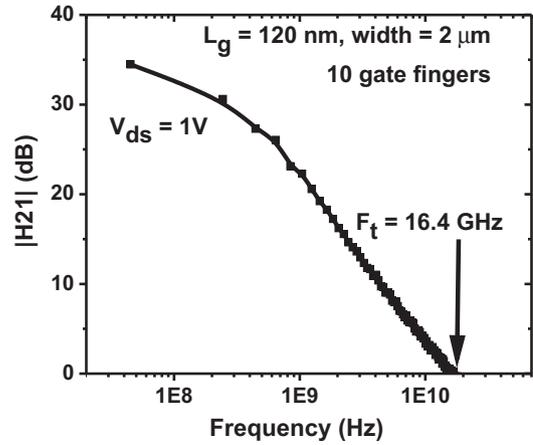


Fig. 12. Measured current gain ($|h_{21}|$) vs. frequency for an InGaAs nFET (top layer) with $L_g = 120$ nm and 10 parallel gate fingers. Cut-off frequency (F_t) of 16.4 GHz is obtained for $V_{ds} = 1$ V.

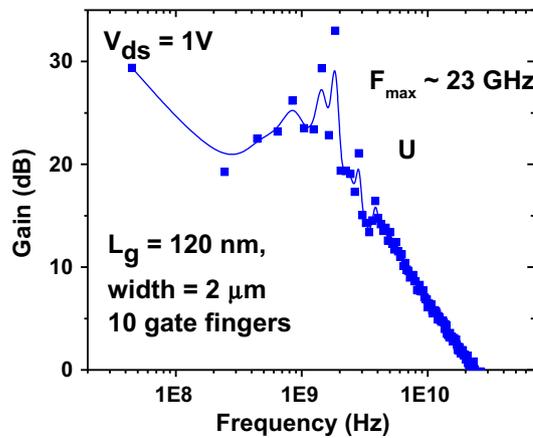


Fig. 13. Measured unilateral gain (U) vs. frequency for an InGaAs nFET (top layer) with $L_g = 120$ nm and 10 parallel gate fingers. F_{max} of 23 GHz is obtained for $V_{ds} = 1$ V.

$$Y_{MOSFET} = Y_{TOTAL} - Y_{OPEN} \quad (2)$$

The current gain $|h_{21}|$ (dB) vs. frequency is shown in Fig. 12, for a device with $L_g = 120$ nm. A cut-off frequency (F_t) of 16.4 GHz is obtained for $V_{ds} = 1$ V. Lower F_t value is probably due to higher

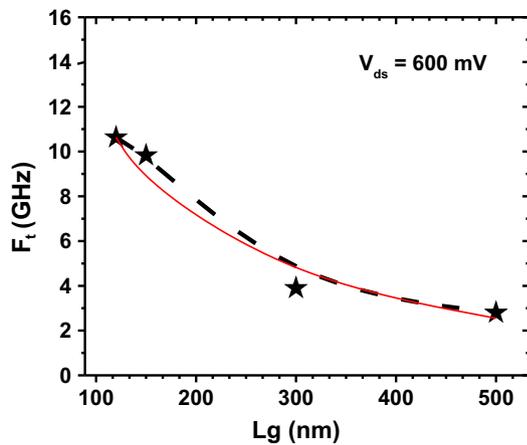


Fig. 14. Cut-off frequency vs. gate length (L_g) for top InGaAs nFETs for $V_{ds} = 600$ mV.

access resistance in the device which degrades the transconductance and higher gate-source capacitance resulting from the overlap of the source line with gate line. Fig. 13 shows the unilateral gain (U) vs. frequency. From the graph, maximum oscillation frequency (F_{max}) is estimated to be 23 GHz. Cut-off frequency vs. L_g plotted in Fig. 14 shows and an increase in cut-off frequencies with decreasing L_g . The solid line shows $1/L_g$ trend line indicating that F_t is inversely proportional to L_g . Hence, scaling down L_g further, along with improving access resistance and a relaxed gate-contact spacing could provide a way to significantly increase the cut-off frequency.

4. Conclusion

We show, for the first time, RF characterization of InGaAs RMG nFETs fabricated on top of SiGe-OI finFETs in 3D monolithic integration. A cut-off frequency of 16.4 GHz is obtained for $L_g = 120$ nm nFET with negligible impact on the bottom pFET performance. The InGaAs nFETs also feature a scaled gate stack and tight pitch design (gate-contact spacing = 100 nm). Thus we demonstrate the benefit of InGaAs-on-SiGe 3D monolithic integration, showing that independently optimized multi-functional layers can be fabricated exploiting the advantages of both device layers.

Acknowledgment

Funding from the EU is acknowledged under the following EU and Marie-Curie projects: ICT-2013-11 COMPOSE3, ICT-2013-11

IIVMOS, H2020-ICT-2015-688784-INSIGHT and PEOPLE-2013-IEF FACIT. Authors also acknowledge the IBM MRL and BRNC staff as well as management support at both sites.

References

- [1] Batude P, Vinet M, Pouydebasque A, Le Royer C, Previtali B, Tabone C, et al. GeOI and SOI 3D monolithic cell integrations for high density applications symposium on VLSI technology, 2009 [p. 166–7].
- [2] Lee S, Chobpattana V, Huang CY, Thibeault BJ, Mitchell W, Stemmer S, et al. Record ion (0.50 mA/m at $V_{DD} = 0.5$ V and $I_{off} = 100$ nA/m) 25 nm-gate-length $ZrO_2/InAs/InAlAs$ MOSFETs. In: 2014 Symposium on VLSI technology (VLSI-technology): digest of technical papers. p. 1–2.
- [3] Sun Y, Majumdar A, Cheng CW, Kim YH, Rana U, Martin RM, et al. Self-aligned III-V MOSFETs: towards a CMOS compatible and manufacturable technology solution. In: 2013 IEEE international electron devices meeting. p. 2.7.1–4.
- [4] del Alamo JA, Antoniadis D, Guo A, Kim DH, Kim TW, Lin J, et al. InGaAs MOSFETs for CMOS: recent advances in process technology. In: 2013 IEEE international electron devices meeting. p. 2.1.1–4.
- [5] Zota CB, Roll G, Wernersson LE, Lind E. Radio-frequency characterization of selectively regrown InGaAs lateral nanowire MOSFETs. IEEE Trans Electron Dev 2014;61(12):4078–83.
- [6] Deshpande V, Djara V, O'Connor E, Hashemi P, Balakrishnan K, Sousa M, et al. Advanced 3D monolithic hybrid CMOS with sub-50 nm gate inverters featuring replacement metal gate (RMG)-InGaAs nFETs on SiGe-OI Fin pFETs. In: IEEE international electron devices meeting (IEDM). p. 8.8.1–4.
- [7] Ge Tong, Guo Linfei, He Huiqiao, Yang Kang, Jia Yu, Chang Joseph. Envelope tracking RF power amplifiers: fundamentals, design challenges, and unique opportunities offered by LEES-SMART InGaAs-on-CMOS process. Proc Eng 2016;141:94–7.
- [8] Hashemi P, Kobayashi M, Majumdar A, Yang LA, Baraskar A, Balakrishnan K, et al. High-performance Si1-xGex channel on insulator trigate PFETs featuring an implant-free process and aggressively-scaled fin and gate dimensions. In: Symposium on VLSI technology (VLSIT). p. T18–9.
- [9] Hashemi P, Balakrishnan K, Majumdar A, Khakifirooz A, Kim W, Baraskar A, et al. Strained Si1-xGex-on-insulator PMOS FinFETs with excellent sub-threshold leakage, extremely-high short-channel performance and source injection velocity for 10nm node and beyond. In: VLSI technology (VLSI-technology): digest of technical papers. p. 1–2.
- [10] Batude P, Fenouillet-Beranger C, Pasini L, Lu V, Deprat F, Brunet L, et al. 3DVLSI with CoolCube process: an alternative path to scaling. In: 2015 Symposium on VLSI technology (VLSI-technology): digest of technical papers. p. T48–9.
- [11] Czornomaz L, Daix N, Caimi D, Sousa M, Erni R, Rossell MD, et al. 3DVLSI with CoolCube process: an alternative path to scaling. In: 2012 IEEE international electron devices meeting (IEDM). p. 23.4.1–4.
- [12] Djara V, Deshpande V, Uccelli E, Daix N, Caimi D, Rossel C, et al. An InGaAs on Si-platform for CMOS with 200 mm InGaAs-OI substrate, gate-first, replacement gate planar and FinFETs down to 120 nm contact pitch. In: 2015 Symposium on VLSI technology (VLSI technology). p. T176–7.
- [13] Djara V, Deshpande V, Sousa M, Caimi D, Czornomaz L, Fompeyrine J. CMOS-compatible replacement metal gate InGaAs-OI FinFET With $I_{ON} = 56\mu$ textA μ textm at $V_{DD} = 0.5$ V and $I_{OFF} = 100$ nA/ μ textm. IEEE Electron Dev Lett 2016;37(2):169–72.
- [14] Djara Vladimir, Sousa Marilyne, Dordevic Nikola, Czornomaz Lukas, Deshpande Veeresh, Marchiori Chiara, et al. Low Dit $HfO_2/Al_2O_3/In_{0.53}Ga_{0.47}As$ gate stack achieved with plasma-enhanced atomic layer deposition. Microelectron Eng 2015;147:231–4.
- [15] Djara Vladimir, Czornomaz Lukas, Deshpande Veeresh, Daix Nicolas, Uccelli Emanuele, Caimi Daniele, et al. Tri-gate InGaAs-OI junctionless FETs with PE-ALD Al_2O_3 gate dielectric and H_2/Ar anneal. Solid-State Electron 2016;115(B):103–8.



Elimination of the channel current effect on the characterization of MOSFET threshold voltage using junction capacitance measurements



Daniel Tomaszewski^{a,*}, Grzegorz Głuszko^a, Lidia Łukasiak^b, Krzysztof Kucharski^a, Jolanta Malesińska^a

^aDivision of Silicon Microsystem and Nanostructure Technology, Instytut Technologii Elektronowej (ITE), ul. Okulickiego 5E, 05-500 Piaseczno, Poland

^bInstitute of Microelectronics and Optoelectronics, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

MOSFET
CMOS
Threshold voltage
Junction capacitance
Parameter extraction

ABSTRACT

An alternative method for an extraction of the MOSFET threshold voltage has been proposed. It is based on an analysis of the MOSFET source-bulk junction capacitance behavior as a function of the gate-source voltage. The effect of the channel current on the threshold voltage extraction is fully eliminated. For the threshold voltage and junction capacitance model parameters non-iterative methods have been used. The proposed method has been demonstrated using a series of MOS transistors manufactured using a standard CMOS technology.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The threshold voltage is one of the first-order specifications of CMOS technologies. It is used for the process development (e.g. setting doping profiles, if applicable), for the characterization of the process including its variability, and for the development of process design kits (PDK's). Therefore, the MOSFET threshold voltage extraction is a mandatory step in the evaluation of CMOS processes, and in the sequences of parameter extractions leading to complete device-level compact models necessary for electrical simulation, thus for a design of integrated circuits (ICs).

The standard definition of the large area MOSFET threshold voltage (V_{th}) results from the 1D electrostatic analysis, in which the channel current flow is neglected [1,2]. On the other hand the threshold voltage extraction methods, which are in common use, are based on the measured I–V characteristics [3,4]. Obviously, the theory and experiment are inconsistent in this case. Mobility degradation related to the electric field and a voltage drop across the source and drain resistances in real devices affect V_{th} extraction. Methods aimed at minimization of these effects have been proposed, e.g. in [5,6]. However they do not eliminate fully the channel current effect on V_{th} extraction. Extensive research on

the identification of the current flow effect (via mobility and lateral electric field) on the extracted V_{th} was presented in [7,8]. However this approach is still based on I–V characteristics and requires the calculation of higher order derivatives.

The V_{th} extraction using the current-based methods is disadvantageous also with respect to the characterization of the process variability, which becomes one of the main limitations of the MOSFET scaling. The unambiguous V_{th} extraction is clearly needed to distinguish the device behavior variability linked to V_{th} itself, from other variability sources such as mobility, interface roughness, or series resistance, which lead to inaccurate V_{th} extraction when using strong inversion-based current-voltage methods and hence to unwanted artifacts and cross-correlations in process or variability analyses [9].

The same phenomena, which define the carrier concentration distribution in the channel and the MOSFET operation in steady-state conditions are responsible for the device small signal behavior. The MOSFET capacitances consist of intrinsic and extrinsic parts [1,2]. The intrinsic components are related to the area between the source and drain, whereas the extrinsic ones are related to the device peripheries. While the modeling of the most of the intrinsic capacitances is difficult because of problems with the channel charge sharing between the source and drain terminals (unless no current flow conditions are considered), the modeling of the parasitic extrinsic capacitances is much more straightforward in terms of their interpretation. By definition the phenomena in the areas associated with the extrinsic capacitances are of electro-

* Corresponding author.

E-mail addresses: dtomasz@ite.waw.pl (D. Tomaszewski), ggluszko@ite.waw.pl (G. Głuszko), l.lukasiak@imio.pw.edu.pl (L. Łukasiak), kucharsk@ite.waw.pl (K. Kucharski), jmales@ite.waw.pl (J. Malesińska).

static nature. This has encouraged us to apply the MOSFET capacitance measurements for the threshold voltage extraction. In this work we have explored first of all a method based on the measurements of the MOSFET source-bulk (SB) junction capacitance C_{bs} vs the V_{GS} and V_{BS} voltages. The drain terminal remains open, therefore the channel current effect is efficiently eliminated. It may be expected that such an approach not only allows ambiguities in V_{th} extraction to be avoided, but can be also helpful in eliminating other variability sources affecting the threshold voltage via the channel current flow (unavoidable in the I-V based methods) from the V_{th} variability. These are e.g. mobility variations and line edge/width roughness effects.

The paper is arranged as follows. In Section 2 the proposed method has been described and illustrated using numerical simulation results. In Section 3 the experiment details are discussed. Test devices, electrical measurement conditions and results are described. In Section 4 the parameter extraction algorithms based on the measurement data are presented. They allow for the efficient extraction of the threshold voltage, body effect factor and Fermi level, as well as for extraction of the junction capacitance, its grading coefficient and built-in voltage. In Section 5 the considerations are summarized.

2. Description of the method

As a starting point let us begin with a short discussion of the p-channel MOSFET numerical simulation results obtained using Silvaco Atlas, and shown in Fig. 1. It is visible, that the potential distribution in the space charge area of the SB junction below the gate behaves differently than in the area outside the gate. As the gate-source voltage V_{GS} changes the potential distribution below the gate changes too, while that below the junction remains constant. In accumulation the potential distribution below the gate is rather insensitive to the gate bias, therefore it is expected that the C_{bs} capacitance does not change with V_{GS} . However at the onset of strong inversion a channel is induced. A capacitor is formed, which plates are the inversion layer connected with the source and the bulk area below the gate. Between these plates there is a gate-induced depletion area. This mechanism should be reflected by a steep increase of the C_{bs} capacitance. This observation is the basis of the proposed method.

The simulated hole distributions in the pMOSFET calculated at the accumulation and inversion below the gate are shown in Fig. 2. Two simple capacitor networks represent the $C_{bs}(V_{BS}, V_{GS})$ characteristics behavior for these conditions. $C_{bs}(V_{BS})$ capacitance in Fig. 2a and b denotes the capacitance of the junction including its side-wall areas, whereas $C_{bs,c}(V_{BS}, V_{GS})$ in Fig. 2b denotes the capacitance component induced by the inversion channel buildup. In Fig. 3 it has been shown that an increase of the C_{bs} capacitance appears at the onset of the strong inversion below the gate, i.e. at the threshold. As expected, below the threshold the C_{bs} capacitance remains almost constant relative to the gate voltage. A small variation of the C_{bs} with V_{GS} noticeable for $V_{BS} = 0$ V is due to a decrease of the side-wall component, which occurs if the gate bias approaches the threshold. This is because under such conditions the depletion area width at the surface increases. This effect is clearly visible in Fig. 1a and b. For completeness it should be added, that at this gate bias range the hole concentration beneath the gate increases exponentially with V_{GS} voltage.

The inversion onset is visible also in other AC characteristics of the MOSFETs. In Fig. 4 the $C_{bs}(V_{GS})$ and $C_{gs}(V_{GS})$ data sets are compared. In the latter characteristics even a more pronounced increase at the threshold is visible. However, it may be seen, that in the considered case the C_{gs} capacitance is V_{GS} -dependent both above and below the threshold, where the effect is much stronger

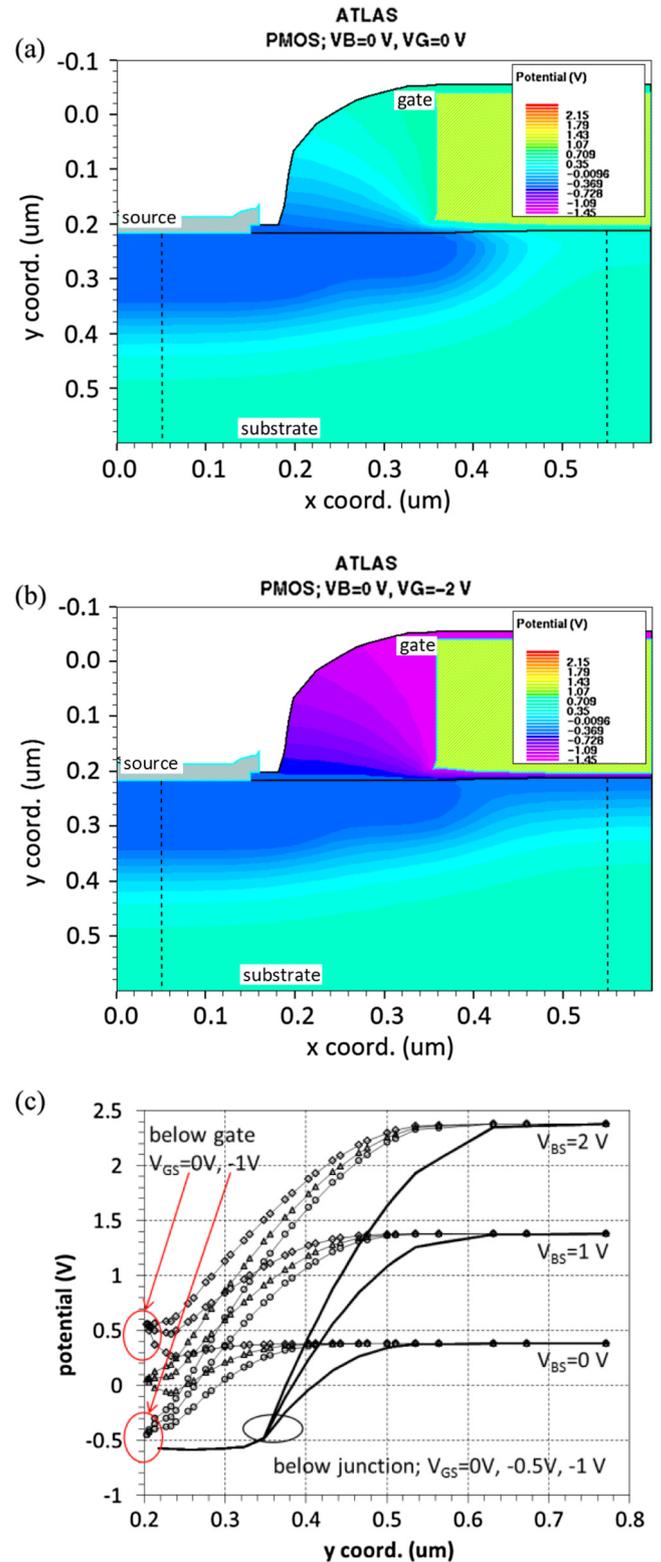


Fig. 1. Potential distribution in p-channel MOSFET for (a) $V_{BS} = 0$ V, $V_{GS} = 0$ V, (b) $V_{BS} = 0$ V, $V_{GS} = -2$ V; (c) Potential distributions along vertical dashed lines drawn in (a), (b) for $V_{BS} = 0, 1, 2$ V, $V_{GS} = 0, -0.5, -1$ V.

than for the C_{bs} capacitance. Such behavior is induced by the variation of the so-called internal fringing component of the C_{gs} capacitance predominant below the threshold and its smooth transition into the gate-channel capacitance C_{gc} prevalent above the threshold. This is why the threshold voltage cannot be unambiguously

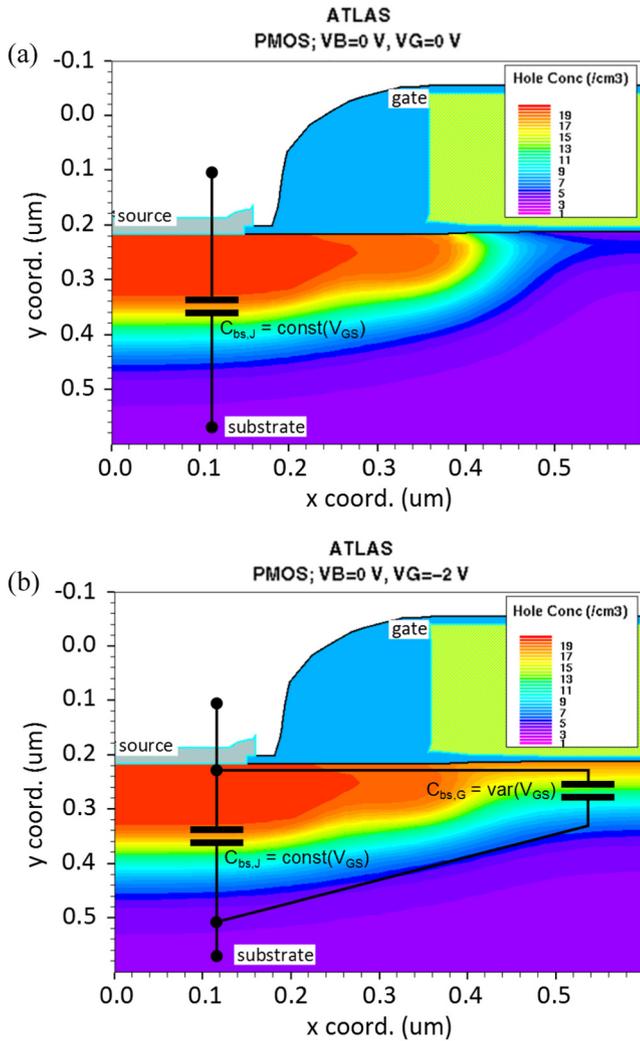


Fig. 2. Hole distribution in the p-channel MOSFET for $V_{BS}=0\text{ V}$ and for (a) depletion, (b) inversion below the gate; components of the C_{bs} capacitance are shown in order to illustrate the effect of the inversion layer formation.

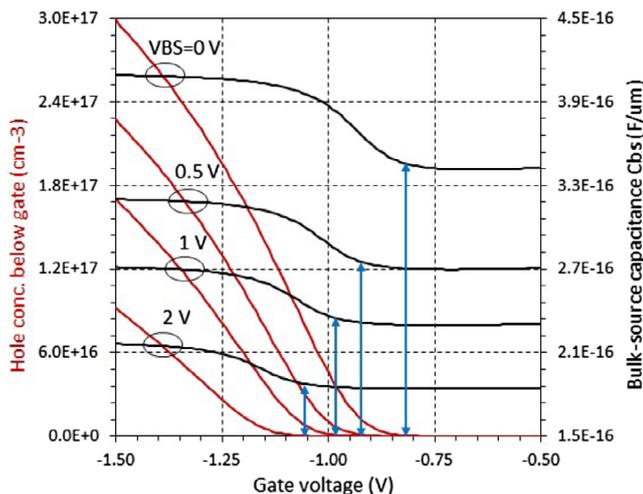


Fig. 3. $C_{bs}(V_{GS})$ characteristics (in black) in comparison with the curves illustrating hole concentration below the gate vs. gate voltage (in red) for different substrate bias values; blue lines indicate the adequacy between points of C_{bs} increase and inversion onset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

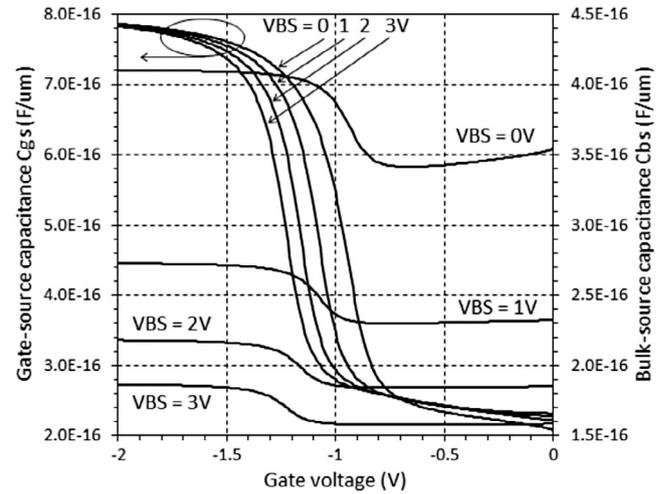


Fig. 4. A comparison of $C_{bs}(V_{GS})$ and $C_{gs}(V_{GS})$ characteristics showing a behavior of both characteristics families at the threshold.

extracted from the $C_{gs}(V_{GS})$ characteristics as the gate voltage, at which C_{gs} starts to increase. Using the described approach is easier if the threshold is determined from the $C_{bs}(V_{GS})$ characteristics. On the other hand we must acknowledge that a reliable method for V_{th} extraction based on the $C_{gs}(V_{GS})$ data has been proposed in [10]. In this method V_{th} corresponds to the maximum of dC_{gs}/dV_{GS} , which can be easily determined. It should be added, though that in theory the method is equivalent to the I-V based method presented in [6] which uses the maximum of $d(g_m/I_D)/dV_{GS}$. Nevertheless, in this paper we propose the threshold voltage extraction procedure based on the less “dynamic” $C_{bs}(V_{GS})$ characteristics. It will be shown that our method not only allows the threshold voltage and its parameters to be estimated, but also directly enables junction capacitance modeling and provides an additional method for the Fermi voltage extraction.

The proposed method consists in the measurement of the $C_{bs}(V_{GS})$ characteristics for a series of V_{BS} voltages. It is expected, that on the $C_{bs}(V_{GS})$ curves at the given V_{BS} voltages and at the accumulation/depletion conditions below the gate the C_{bs} capacitance is constant regardless of V_{GS} variation. It corresponds to the junction capacitance $C_{bs,j}(V_{BS})$ (see Fig. 2) including the area and side-wall components. The clearly visible transition points between the accumulation/depletion and inversion ranges define the $V_{th}(V_{BS})$ characteristics.

3. Experimental results

For the experimental validation of the method described in the previous section, we have used the n- and p-channel MOSFETs fabricated in ITE using a single p-well, polysilicon gate CMOS process with the following characteristics: $3\ \mu\text{m}$ polysilicon line width, $65\ \text{nm}$ gate oxide thickness, $1.2 \cdot 10^{16}$ and $6 \cdot 10^{15}\ \text{cm}^{-3}$ substrate doping concentrations in the n- and pMOSFETs accordingly. Besides the standard devices of rectangular shape we have also used finger-type wide ones, originally designed for the characterization of the gate overlap capacitances. In such devices significantly larger junction capacitances are expected. Layouts of two types of the pMOSFETs are shown in Fig. 5.

The MOSFET C-V characteristics have been measured using a Keithley 4200-SCS parametric analyzer equipped with a 4200-CVU unit. Its force and sense terminals have been connected with the MOSFET source and bulk electrodes respectively. In this way the DC bias of the SB junction has been provided by the CVU unit. The gate has been biased by a source-measure unit (SMU) included

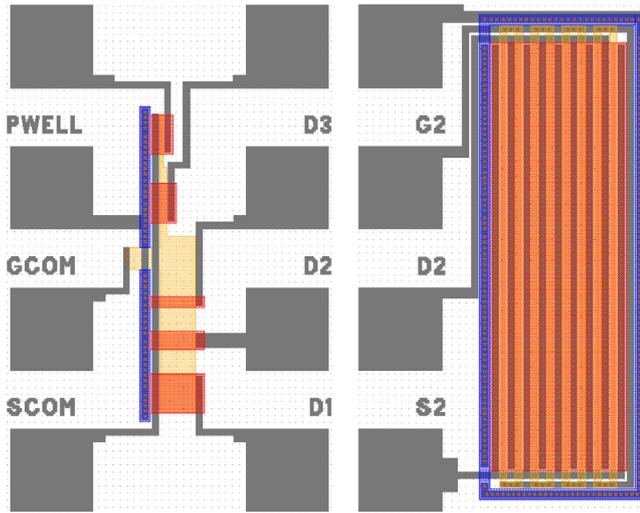


Fig. 5. Layout of p-channel MOSFETs under test: left - a set of standard devices with the $W(\mu\text{m})/L(\mu\text{m})$ ratio 50/50, 20/50, 10/50, 50/10, 50/5, right - a finger-type MOSFET with $W/L = 4800/10$.

in the analyzer. We have used a test signal of 100 kHz frequency and 30 mV rms level. The measurements have been done at the 10 pF range, for which the accuracy is approximately 0.02 pF. In order to minimize the effect of stray capacitances and parasitic resistances open and short corrections have been done.

Since the measurement set-up makes it easier to measure the C_{bs} capacitance as a function of V_{BS} than V_{GS} , the proposed method consists in the measurement of the $C_{bs}(V_{BS})$ characteristics for a series of V_{GS} voltages. Next, $C_{bs}(V_{GS})$ characteristics with V_{BS} as a parameter are constructed. The accuracy of the method is limited by the gate voltage step size.

The $C_{bs}(V_{BS})$ characteristics of a p-channel MOSFET with $W = 50 \mu\text{m}$ and $L = 50 \mu\text{m}$ measured at different V_{GS} voltages are shown in Fig. 6a. As expected, a strong influence of the gate voltage is clearly visible. Enlarged bottom parts of these curves are shown in Fig. 6b. Though in the full-scale plot these parts appear to be constant, a noticeable curvature is revealed in the enlarged plot. An “envelope” of the whole family corresponds to the accumulation/depletion conditions below the gate ($-0.5 \text{ V} \leq V_{GS} \leq 0 \text{ V}$). Each point of the “envelope” has been calculated as a mean value of C_{bs} capacitances measured for a series of four V_{GS} voltages. The “envelope” corresponds to the $C_{bs}(V_{BS})$ characteristics of the SB junction including its edge.

The $C_{bs}(V_{GS})$ characteristics are shown in Fig. 7a. As expected, three regions may be distinguished in these curves. In the first region the C_{bs} capacitance is the lowest and is gate voltage independent. In this gate bias range the area below the gate is accumulated or depleted and the C_{bs} capacitance is determined only by the space charge of the SB junction. If the gate voltage becomes more negative and exceeds the threshold voltage, a steep increase of the C_{bs} capacitance is observed. This effect is caused by the formation of a highly conductive inversion layer. Similarly to the real source area, below the gate-induced virtual source there is also a depletion region which separates the source plate from the bottom quasi-neutral substrate. In order to illustrate this mechanism four $C_{bs}(V_{GS})$ curves are plotted in Fig. 7b after subtracting the “envelope” value taken from Fig. 6. Points of the capacitance sharp increase are clearly visible.

The $C_{bs}(V_{GS})$ characteristics allow the $V_{th}(V_{BS})$ characteristics of the MOSFETs to be derived. In [10] it has been shown, that the threshold voltage corresponds to the maximum of dC_{gs}/dV_{GS} , thus to the maximum of dC_{gs}/dV_{GS} . There is no similar formal proof concerning the maximum of dC_{bs}/dV_{GS} . However, based on the inter-

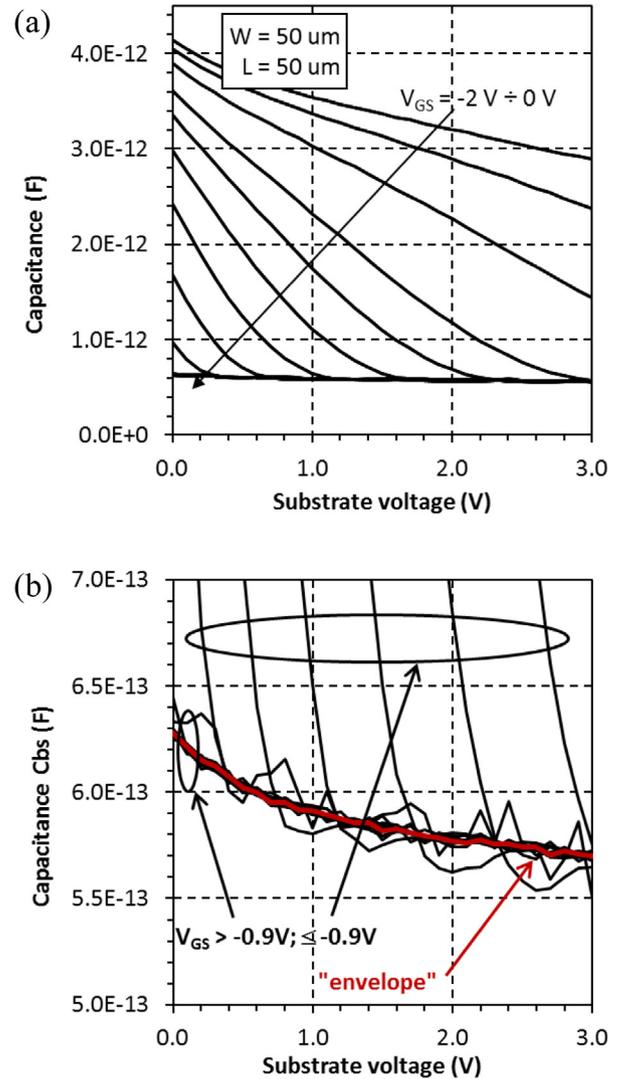


Fig. 6. (a) $C_{bs}(V_{BS})$ characteristics of a p-channel MOSFET, $W = 50 \mu\text{m}$, $L = 50 \mu\text{m}$; $V_{GS} = -2.0 \text{ V}$; (b) the envelope of the $C_{bs}(V_{BS})$ characteristics family.

pretation of the voltage effect on C_{bs} capacitance (see Section 2) we dare to say that the largest variation of C_{bs} corresponds to the onset of the inversion. Moreover, based on the numerical simulations shown in Fig. 4 the dC_{gs}/dV_{GS} and dC_{bs}/dV_{GS} curves may be constructed (Fig. 8). It may be clearly noticed that the maxima of the corresponding derivative characteristics perfectly coincide. E. g. at $V_{BS} = 0 \text{ V}$ their maximum is at $V_{GS} = -0.93 \text{ V}$, and at $V_{BS} = 3 \text{ V}$ the maximum is at $V_{GS} = -1.23 \text{ V}$. Following this consideration, on each $C_{bs}(V_{GS})$ curve the threshold voltage is determined as the V_{GS} value, at which the dC_{bs}/dV_{GS} reaches maximum.

In Fig. 9a the $V_{th}(V_{BS})$ data obtained using the proposed method are set together with the $V_{th}(V_{BS})$ curves obtained using the standard current-based methods, namely a linear extrapolation of I_D-V_{GS} characteristics at the transconductance maximum, and a linear extrapolation of $I_D/g_m^{0.5}-V_{GS}$ characteristics [5]. The proposed C-V based approach gives results close to the method based on the linear extrapolation of I_D-V_{GS} characteristics (Fig. 9b).

In Fig. 9a several points deviate slightly from the $V_{th}(V_{BS})$ curves corresponding to the C-V based method of V_{th} extraction. They are due of the fact that the V_{GS} step is too large. This effect is noticeable mainly for the relatively small values of the measured capacitance. However it makes further data processing more difficult and confirms the need for the direct measurement of the $C_{bs}(V_{GS})$ characteristics with V_{BS} voltage stepped.

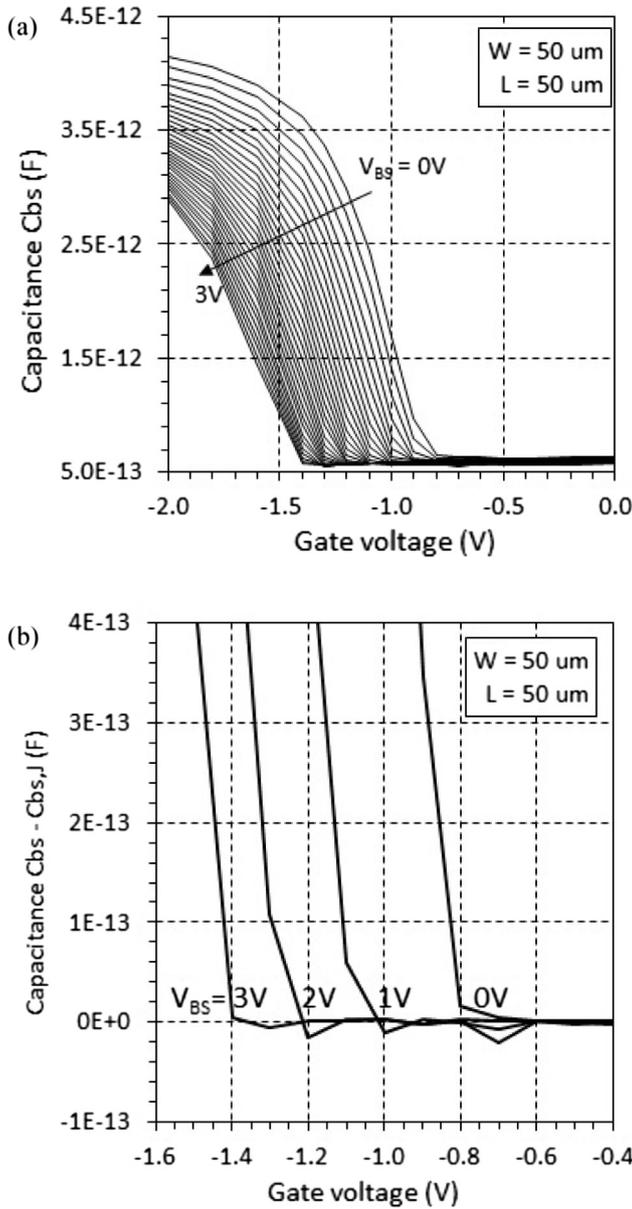


Fig. 7. (a) $C_{bs}(V_{GS})$ characteristics for p-channel MOSFET obtained from the data in Fig. 6; (b) $C_{bs}(V_{GS})$ data after subtraction of the “envelope”; points of the strong inversion onset are clearly visible.

The proposed method for the threshold voltage characterization has been also applied to the wide, finger-type MOSFETs (Fig. 5). In these devices much larger junction capacitances are expected. In Figs. 10 and 11 the appropriate C-V and the resulting $V_{th}(V_{BS})$ data for the p- and n-channel MOSFETs, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$ are shown. In the $C_{bs}(V_{BS})$ characteristics of the large area devices two “envelopes” are visible. The bottom one represents the $C_{bs,J}$ component of the total C_{bs} capacitance, whereas the top one represents the sum of $C_{bs,J}$ and $C_{bs,G}$ components. The $C_{bs,G}$ at the top C_{bs} “envelope” corresponds to the maximum depletion width below the gate, which appears at the strong inversion conditions. Accordingly, in the $C_{bs}(V_{GS})$ characteristics at any V_{BS} voltage two flat regions are visible. They are more pronounced than in the case of the device with $W = L = 50 \mu\text{m}$ considered earlier and are suitable for the parameter extraction. It should be added that the threshold voltages extracted using the C_{bs} - V_{GS} -based method are lower than those determined using the I-V-based methods. This remains in agreement with the results presented in [6].

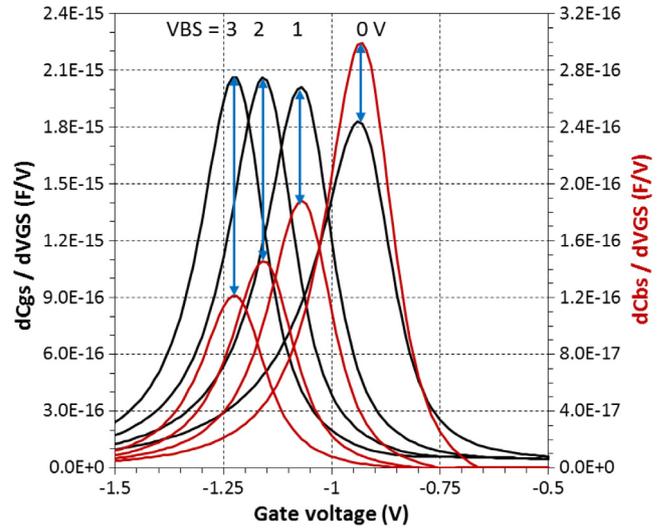


Fig. 8. Derivatives of the $C_{gs}(V_{GS})$ (in black) and $C_{bs}(V_{GS})$ (in red) characteristics for p-channel MOSFET simulated using Silvaco ATLAS and shown in Fig. 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

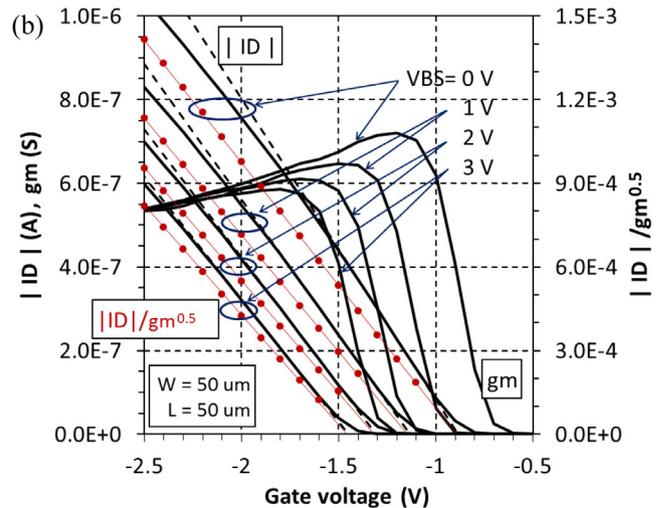
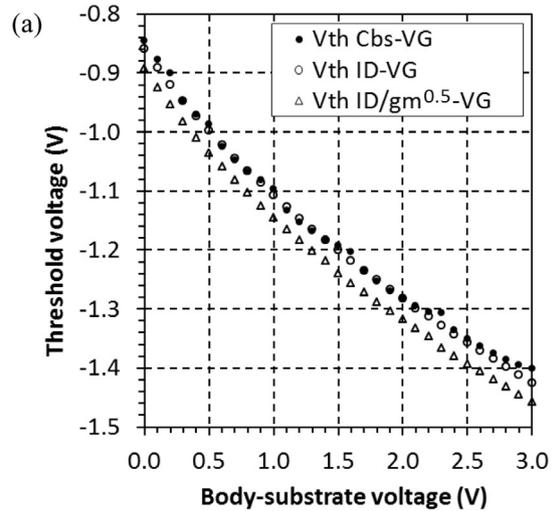


Fig. 9. Threshold voltage extracted with three methods: (a) dependence on V_{BS} voltage, (b) extraction of V_{th} based on I-V characteristics.

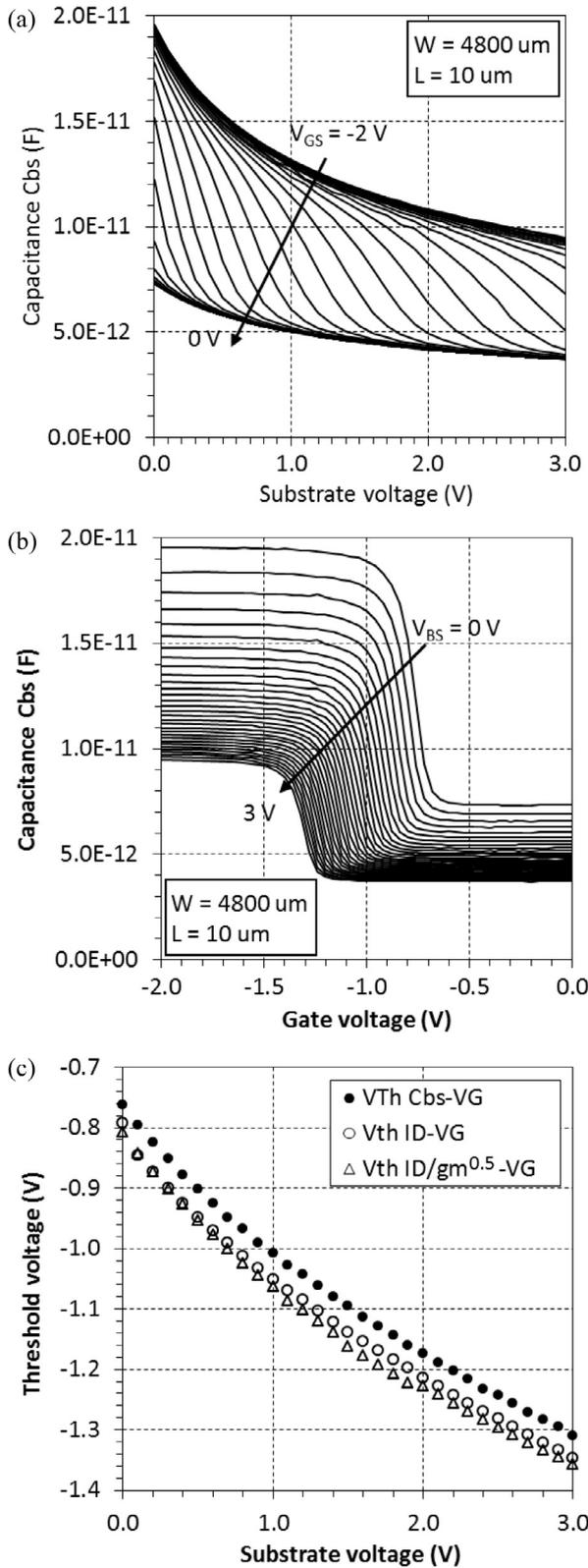


Fig. 10. Extraction of the threshold voltage for p-channel MOSFET, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$: (a) $C_{bs}(V_{BS})$ data, (b) $C_{bs}(V_{GS})$ data obtained from $C_{bs}(V_{BS})$ data, (c) $V_{th}(V_{BS})$ curves determined with C-V- and I-V-based methods.

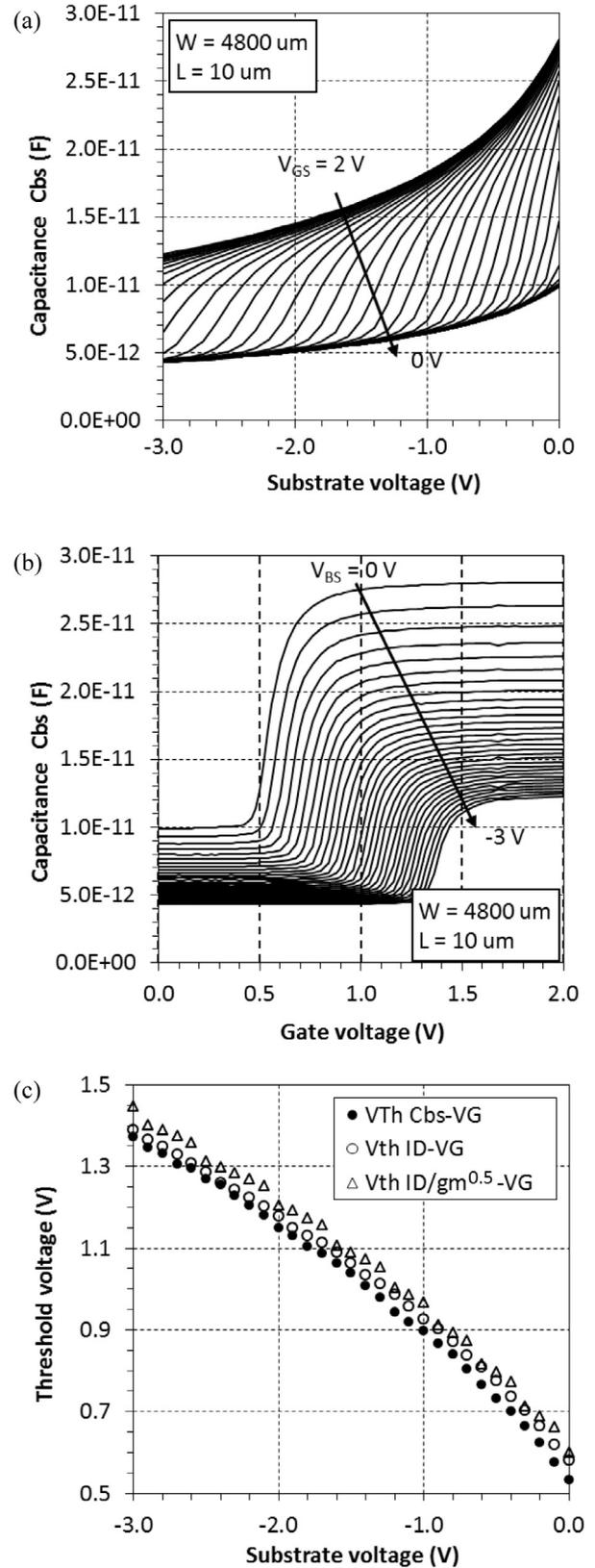


Fig. 11. Extraction of the threshold voltage for n-channel MOSFET, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$: (a) $C_{bs}(V_{BS})$ data, (b) $C_{bs}(V_{GS})$ data, (c) $V_{th}(V_{BS})$ curves determined with C-V- and I-V-based methods.

4. Parameter extraction

4.1. Threshold voltage

In the case of the long-channel enhancement-mode MOSFETs the threshold voltage V_{th} is given by a well-known formula (1).

$$V_{th} = V_{th,0} + t \cdot \gamma \cdot \left(\sqrt{2 \cdot \phi_F - t \cdot V_{BS}} - \sqrt{2 \cdot \phi_F} \right) \quad (1)$$

where $V_{th,0}$ – threshold voltage at $V_{BS} = 0$ V, γ – body factor, $2\phi_F$ – double Fermi voltage. The variable t is equal to +1 for the n-channel MOSFETs and to –1 for the p-channel MOSFETs. Typically γ and $2\phi_F$ parameters in (1) are extracted iteratively starting from a given approximation of $2\phi_F$. However, using a non-iterative method described in [11,12] for $V_{th}(V_{BS})$ data, the parameters of V_{th} in (1) may be determined with (2a) and (2b) using a linear regression between terms $V_{th} - V_{th,0}$ and $(dV_{th}/dV_{BS})^{-1}$.

$$V_{th,0} = V_{th}(0) \quad (2a)$$

$$V_{th} - V_{th,0} = -0.5 \cdot t \cdot \gamma^2 \cdot \frac{1}{\frac{dV_{th}}{dV_{BS}}} - t \cdot \gamma \cdot \sqrt{2 \cdot \phi_F} \quad (2b)$$

The regressions for the $V_{th}(V_{BS})$ data extracted using the I-V based and C_{bs} - V_{GS} methods are illustrated in Fig. 12. In each chart two types of fitting are visible. The four regressions based on the $V_{th}(V_{BS})$ data obtained directly from the measurement data are plotted in black. A large spread of the points subject to regression is visible, in particular in the case of the n-channel MOSFET (Fig. 12b). Other four regressions based on the $V_{th}(V_{BS})$ data filtered using Savitzky-Golay method are plotted in red¹.

The expected values of the body factor are approximately $1.2 V^{0.5}$ (nMOSFET), and $0.7 V^{0.5}$ (pMOSFET). The expected Fermi voltage values are 0.75 V (nMOSFET) and 0.65 V (pMOSFET). The extracted threshold voltage parameter values are listed in Table 1. The difference between the γ values for the p-channel and n-channel MOSFETs results from the fact, that CMOS technology with a single p-well has been used. The γ values extracted based on the raw data are far below the expected ones. The same situation is with the $2\phi_F$ values. This effect may be attributed to the noisy data being subject to the linear fit. A significant spread of the data points in Fig. 12 strongly affects the slope and the intersection of the regression line with the horizontal axis, via the $(dV_{th}/dV_{BS})^{-1}$ term used in (2b). The slope determines the body factor value, whereas the intersection point determines the Fermi voltage. The filtering procedure has improved the extraction of γ and $2\phi_F$ in the case of the nMOSFET. In the case of the pMOSFET the filtering has not helped much, because the $V_{th}(V_{BS})$ data quality is better than in the case of the n-channel device. Following the above discussion it may be stated, that retrieving the smooth $V_{th}(V_{BS})$ characteristics both from the I-V and C-V measurement data is relevant for the reliable extraction. In practice it may be a non-trivial task, but in such a case data filtering may be helpful.

4.2. Junction capacitance

The proposed characterization method allows for a deeper analysis of the C_{bs} capacitance in MOSFETs. The bottom “envelope” of the $C_{bs}(V_{BS})$ characteristics family measured for different V_{GS} voltages (Figs. 6b, 9a, 10a) represents the total C_{bs} capacitance of the MOSFET SB junction including its planar and sidewall components. Modeling of such a characteristics is based on the well-known formula (3).

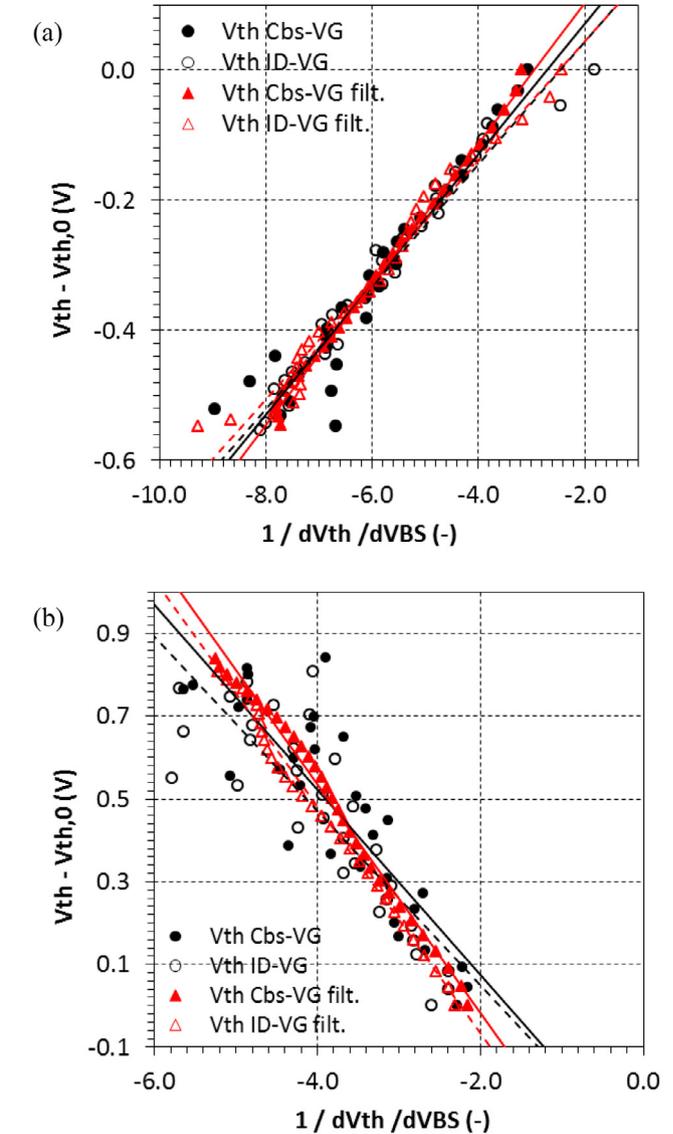


Fig. 12. Linear fits (2) corresponding to the (a) p- and (b) n-channel MOSFET $V_{th}(V_{BS})$ data extracted using I-V and C_{bs} - V_{GS} based methods; each method has been used for the non-filtered and filtered data; solid lines represent regression based on C_{bs} - V_{GS} data; dashed lines represent regression based on I-V data.

$$C_{bs,J} = \frac{C_J}{\left(1 - \frac{t \cdot V_{BS}}{V_{bi}}\right)^{M_J}} \quad (3)$$

where C_J is the capacitance at $V_{BS} = 0$ V, V_{bi} is the built-in voltage, and M_J is a grading coefficient, which is theoretically 1/2 for an ideal abrupt junction and 1/3 for an ideal linearly graded junction. The C_J parameter is easily determined as $C_{bs}(V_{BS} = 0$ V). The remaining two parameters are typically extracted using an iterative procedure. Here, we propose a non-iterative approach, which is based on (4).

$$\begin{aligned} \frac{dC_{bs,J}}{dV_{BS}} &= C_J \cdot t \cdot \frac{M_J}{V_{bi}} \cdot \left(1 - \frac{t \cdot V_{BS}}{V_{bi}}\right)^{-M_J-1} \\ &= C_{bs,J} \cdot t \cdot \frac{M_J}{V_{bi}} \cdot \left(1 - \frac{t \cdot V_{BS}}{V_{bi}}\right)^{-1} \end{aligned} \quad (4)$$

After a simple transformation of (4) the method (5) for the junction capacitance parameter extraction may be formulated. (5b) is subject to The linear regression (5b) between V_{BS} and $C_{bs,J}/(dC_{bs,J}/dV_{BS})$ terms is used for extraction of M_J and V_{bi} .

¹ For interpretation of color in Fig. 12, the reader is referred to the web version of this article.

Table 1Threshold voltage parameters of $4800 \times 10 \mu\text{m}$ MOSFETs extracted from $V_{th}(V_{BS})$ characteristics obtained from the non-filtered and filtered measurements.

Param	NMOS		PMOS	
	Method		Method	
	I_D-V_G	$C_{bs}-V_G$	I_D-V_G	$C_{bs}-V_G$
$V_{th,0}$ (V)	0.582	0.532	-0.792	-0.763
γ ($V^{0.5}$)	0.649	0.669	0.434	0.447
$2\phi_F$ (V)	0.33	0.31	0.29	0.37
After filtering of $V_{th}(V_{BS})$ data obtained from I_D-V_G $C_{bs}-V_G$ meas				
γ ($V^{0.5}$)	0.740	0.743	0.428	0.465
$2\phi_F$ (V)	0.69	0.59	0.29	0.48

$$C_J = C_{bs,J}(0) \quad (5a)$$

$$V_{BS} = t \cdot V_{bi} - M_J \cdot \frac{C_{bs,J}}{dC_{bs,J}/dV_{BS}} \quad (5b)$$

The linear fits regarding the bottom “envelopes” of the $C_{bs}(V_{BS})$ characteristics of the p- and n-channel MOSFETs are shown in Fig. 13. Good quality of these fits is beneficial for the parameter extraction. The calculated values of the parameters are listed in Table 2.

As expected, in the case of the nMOSFET the capacitance C_J and the built-in voltage V_{bi} are larger than in the pMOSFET, because of the substrate doping concentration levels in both device types. There is also a significant difference between the grading coefficients M_J . In the nMOSFET a step junction has been revealed, whereas in the pMOSFET the junction is rather linear. However it should be mentioned, that the C_{bs} capacitance comprises the plane and sidewall components. In the finger-type device the side-wall capacitance component is large because of the large perimeter/area ratio. On the other hand the sidewall capacitance is less sensitive to V_{BS} voltage, than the plane one. Following this we may expect that the parameter extraction results are affected by the test structure non-ideality.

4.3. Gate-induced component of junction capacitance

In the previous section the bottom “envelopes” $C_{bs,J}(V_{BS})$ of the $C_{bs}(V_{BS}, V_{GS})$ characteristics corresponding to the lower plateaus in Figs. 10b and 11b have been used for the junction capacitance parameter extraction. In this section the top “envelopes” are considered. According to the interpretation given in Section 3 the junction capacitance in excess of $C_{bs,J}$ capacitance under the strong inversion conditions is attributed to the gate-induced depletion area. In Fig. 14 the differences between top and bottom “envelopes” are shown. These curves may be described with (6).

$$\begin{aligned} C_{bs} - C_{bs,J} &= A_{G,eff} \cdot \epsilon_{Si}/x_{d,max} \\ &= A_{G,eff} \cdot \epsilon_{Si} / \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_{sub}} \cdot (2\phi_F - t \cdot V_{BS})} \end{aligned} \quad (6)$$

where $A_{G,eff}$ is the effective gate area, and $\epsilon_{Si}/x_{d,max}$ is the gate-induced depletion region capacitance per unit area in the strong inversion conditions below the gate. Other parameters have the usual meaning. After a simple transformation we arrive at (7).

$$(C_{bs} - C_{bs,J})^{-2} = \frac{1}{A_{G,eff}^2} \cdot \frac{-2t}{q \cdot \epsilon_{Si} \cdot N_{sub}} (V_{BS} - t \cdot 2\phi_F) \quad (7)$$

Eq. (7) is suitable for the extraction of the Fermi voltage and of the substrate doping concentration. It is a favorable situation because the $2\phi_F$ values for the n- and pMOSFETs calculated in the Section 4.1 based on the $V_{th}(V_{BS})$ characteristics are far from the expected values probably due to the data noise. The linear regressions (7) are shown in Fig. 15. A very good linear fit is notice-

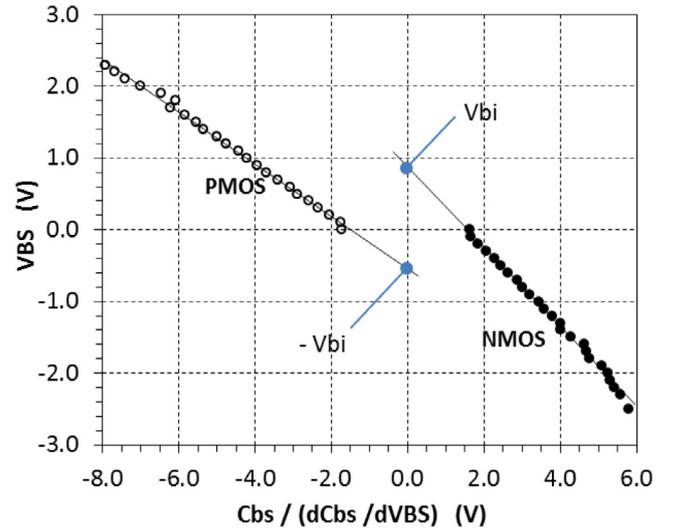


Fig. 13. Extraction of the junction capacitance parameters according to (5) for the n- and p-channel MOSFETs, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$.

Table 2Junction capacitance parameters for $4800 \times 10 \mu\text{m}$ MOSFETs.

Param	NMOS	PMOS
C_J (F)	$9.9 \cdot 10^{-12}$	$3.7 \cdot 10^{-12}$
M_J (-)	0.555	0.364
V_{bi} (V)	0.869	0.543

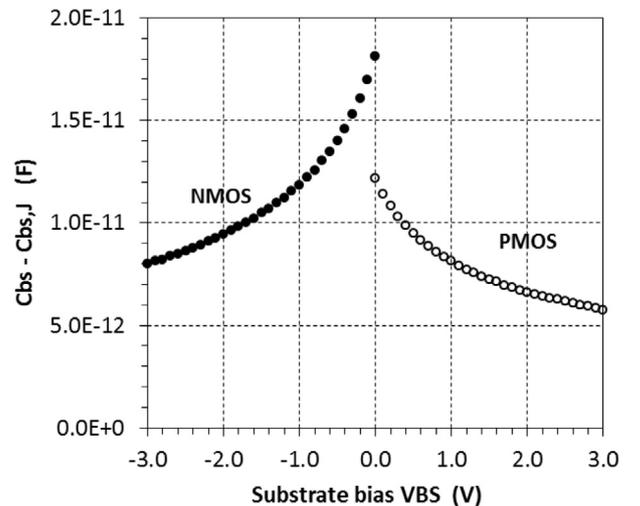


Fig. 14. Gate-induced components of the junction capacitances under strong inversion conditions for the n- and p-channel MOSFETs, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$.

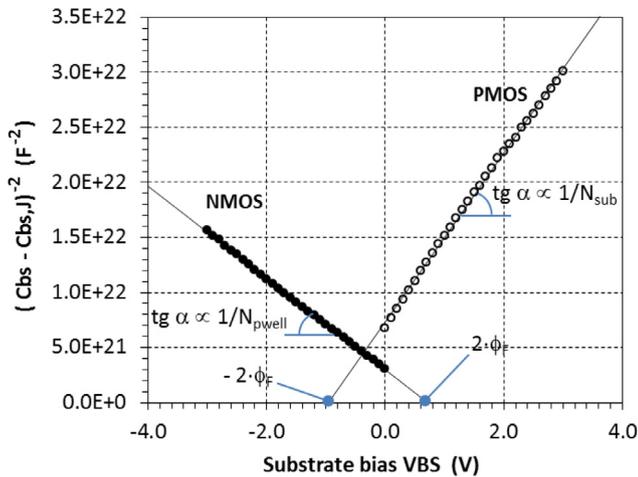


Fig. 15. Extraction of the Fermi voltage and substrate doping concentration for the n- and p-channel MOSFETs, $W = 4800 \mu\text{m}$, $L = 10 \mu\text{m}$.

Table 3

Threshold voltage parameters and doping concentrations determined with (7) for $4800 \times 10 \mu\text{m}$ MOSFETs.

Param	NMOS	PMOS
$N_{\text{sub}} (\text{cm}^{-3})$	$1.2 \cdot 10^{16}$	$7 \cdot 10^{15}$
$2\phi_F (\text{V})$	0.71	0.94
$\gamma (\text{V}^{0.5})$	1.25	0.92

able. The extraction results are summarized in Table 3. The body factor γ has been calculated using the nominal gate oxide thickness 65 nm.

The calculated doping concentrations remain in a good agreement with the CMOS process specification given at the beginning of Section 3. On the other hand a large discrepancy between the Fermi voltages and body factors obtained based on $V_{\text{th}}(V_{\text{BS}})$ and $C_{\text{bs}} - C_{\text{bs},J} = f(V_{\text{BS}})$ data is clearly visible. As it has been discussed earlier, the first data set is saddled with the large spread of data points originating from the raw $V_{\text{th}}(V_{\text{BS}})$ characteristics and shown in Fig. 12. The spread may affect the validity of the parameters in Table 1. The second data set is smooth. This can be a convincing argument for the reliability of the second data set.

Another issue visible in particular in Table 3 is that the Fermi voltage for the nMOSFETs is lower than for the pMOSFETs irrespective of the extracted doping concentration levels. Such an effect is not observed in the built-in voltages V_{bi} calculated based on the junction capacitance characteristics and shown in Table 2. A possible explanation for this observation is that due to boron segregation during the gate oxidation (partially compensated by the boron correction implantation) the acceptor concentration below the nMOSFET gate is lower than below the nMOSFET source junction. The situation for the pMOSFETs is opposite.

Finally, it is worth mentioning that the two data sets used for the parameter extraction, i.e. $V_{\text{th}}(V_{\text{BS}})$ and $C_{\text{bs}} - C_{\text{bs},J} = f(V_{\text{BS}})$, have been obtained using different characteristics and for different measurement conditions. The $V_{\text{th}}(V_{\text{BS}})$ characteristics obtained either from $I_{\text{D}}(V_{\text{GS}})$ or from $C_{\text{bs}}(V_{\text{GS}})$ data and used in (2), correspond to the onset of inversion. On the other hand, the $C_{\text{bs}} - C_{\text{bs},J} = f(V_{\text{BS}})$ data used in (7) correspond to strong inversion conditions. Therefore, by definition the Fermi voltage extracted based on (7) should be larger than the one determined based on (2).

5. Summary

Based on the numerical simulation results the methodology for a combined extraction of the MOSFET threshold voltage (including

body factor and Fermi voltage) and junction capacitance from the $C_{\text{bs}}(V_{\text{GS}})$ characteristics has been introduced. During the measurements the drain terminal of the device under test remains open to eliminate the effect of the channel current flow (via mobility or series resistances) on the threshold voltage extraction. According to the numerical simulations the C_{bs} -based method of the threshold extraction should be equivalent to the method based on gate capacitance measurements. It is worth mentioning that the proposed measurements may be done using gate-controlled diodes as well.

Based on the nonlinear models of the threshold voltage $V_{\text{th}}(V_{\text{BS}})$ and of the junction capacitance $C_{\text{bs}}(V_{\text{BS}})$ three simple algorithms for the threshold voltage and junction capacitance parameter extraction have been proposed. Using simple arithmetic transformations the closed-form, non-iterative formulae for the parameter extraction have been derived. They may be useful for automatic measurements combined with on the fly parameter extraction. There is no risk of a potential negative term in (1) or (3) breaking the automated measurement/extraction procedure.

Two of the proposed extraction methods concern the threshold voltage characterization. The first one, relying on $V_{\text{th}}(V_{\text{BS}})$ characteristics extracted from the $C_{\text{bs}}(V_{\text{GS}})$ data corresponds to the onset of inversion. It allows for a full characterization of the V_{th} of large area MOSFETs. The second one, based on $C_{\text{bs}}(V_{\text{GS}})$ data, is complementary. It has been derived under the assumption of strong inversion conditions below gate. It provides more reliable values of the Fermi voltage and body factor.

There are important conditions and limitations for using the proposed methods. First, for the measurements of the $C_{\text{bs}}(V_{\text{BS}})$, $V_{\text{GS}})$ characteristics we need the MOSFETs or gate-controlled diodes with the junction area not smaller than approximately $1000 \mu\text{m}^2$. Such a large junction is necessary in order to have reliable C_{bs} measurements (range of 0.5 pF) of the devices manufactured using our process. Furthermore, we need also devices with a large gate area not smaller than approximately $1500 \mu\text{m}^2$. Such a large gate is needed for the measurement of the gate-induced C_{bs} component in the strong inversion conditions. The reliable capacitance measurements are necessary to obtain smooth $V_{\text{th}}(V_{\text{BS}})$ and $C_{\text{bs}}(V_{\text{BS}})$ data, used in fitting procedures (2) and (7). The proposed methods have been tested using the standard CMOS devices manufactured in ITE.

References

- [1] Tsividis Y. Operation and modeling of the MOS transistor. 2nd ed. Oxford University Press; 1999.
- [2] Arora N. MOSFET modeling for VLSI simulation theory and practice. World Scientific Publishing; 2007.
- [3] Bucher M, Lallement C, Enz CC. An efficient parameter extraction methodology for the EKV MOST model. Proc. 1996 IEEE int. conf. on microelectronic test structures, Trento, Italy, March 26–28, vol. 9. p. 145–50. doi: <http://dx.doi.org/10.1109/ICMTS.1996.535636>.
- [4] Ortiz-Conde A, García-Sánchez FJ, Muci J, Terán Barrios A, Liou JJ, Ho C-S. Revisiting MOSFET threshold voltage extraction methods. Microelectron Reliab 2013;53(1):90–104. doi: <http://dx.doi.org/10.1016/j.microrel.2012.09.015>.
- [5] Ghibaudo G. New method for the extraction of MOSFET parameters. Electron Lett 1998;24(9):543–5. doi: <http://dx.doi.org/10.1049/el:19880369>.
- [6] Flandre D, Kilchytska V, Rudenko T. Gm/Id method for threshold voltage extraction applicable in advanced MOSFETs with nonlinear behavior above threshold. IEEE Electron Device Lett 2010;31(9):930–2. doi: <http://dx.doi.org/10.1109/LED.2010.2055829>.
- [7] Rudenko T, Kilchytska V, Md Arshad MK, Raskin J-P, Nazarov A, Flandre D. On the MOSFET threshold voltage extraction by transconductance and transconductance-to-current ratio change methods: Part I-Effect of gate-voltage-dependent mobility. IEEE Trans Electron Devices 2011;58(12):4172–9. doi: <http://dx.doi.org/10.1109/TED.2011.2168226>.
- [8] Rudenko T, Kilchytska V, Md Arshad MK, Raskin J-P, Nazarov A, Flandre D. On the MOSFET threshold voltage extraction by transconductance and transconductance-to-current ratio change methods: Part II-Effect of drain voltage. IEEE Trans Electron Devices 2011;58(12):4180–8. doi: <http://dx.doi.org/10.1109/TED.2011.2168227>.

- [9] Sugii N, Tsuchiya R, Ishigaki T, Morita Y. Local V_{th} variability and scalability in Silicon-on-Thin-BOX (SOTB) CMOS with small random-dopant fluctuation. *IEEE Trans Electron Devices* 2010;57(4):835–45. doi: <http://dx.doi.org/10.1109/TED.2010.2040664>.
- [10] Park C-K, Lee C-Y, Lee K, Moon B-J, Byun YH, Shur M. A unified current-voltage model for long-channel nMOSFET's. *IEEE Trans Electron Devices* 1991;38(2):399–406. doi: <http://dx.doi.org/10.1109/16.69923>.
- [11] Jasiński J, Głuszko G, Łukasiak L, Jakubowski A, Grabiński W, Tomaszewski D. Compact modeling of SiON MOST devices. In: Poster presented at MOS-AK/ESSDERC/ESSCIRC Workshop, Edinburgh, Sept 19, 2008, <http://www.mos-ak.org/edinburgh/posters/P10_Tomaszewski_MOS-AK_08.pdf>.
- [12] Tomaszewski D, Malesińska J, Głuszko G. Simple methods of threshold voltage parameter extraction for MOSFET models. In: Proc. 22nd int. conf. mixed design of integrated circuits & systems (MIXDES), June 25–27, 2015, Torun, Poland. p. 222–6. doi: <http://dx.doi.org/10.1109/MIXDES.2015.7208514>.



Low frequency noise assessment in n- and p-channel sub-10 nm triple-gate FinFETs: Part I: Theory and methodology



D. Boudier^{a,*}, B. Cretu^b, E. Simoen^c, R. Carin^a, A. Veloso^c, N. Collaert^c, A. Thean^c

^aUniversity of Caen Normandie, UMR6072 GREYC, F-14050 Caen, France

^bENSICAEN, UMR6072 GREYC, F-14050 Caen, France

^cImec, Kapeldreef 75, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Triple-gate

FinFET

Low frequency noise

1/f noise

Generation-recombination

Traps

ABSTRACT

The transfer characteristic at room temperature of FinFETs processed for sub-10 nm technologies could always be explained by solving Poisson equation throughout the channel – dielectric interface. Various methods for the MOSFET parameters estimation are proposed in the literature. In this paper, the electrical parameters extraction technique based on the Y-function methodology is reminded.

Low frequency noise is presented considering three major noise sources: 1/f noise associated to carrier trapping-detrapping in the gate oxide, channel carrier mobility fluctuations and generation-recombination noise related to traps located in the depletion zone of the device. Theory and methodology in order to identify the 1/f noise mechanism and to have information of the process induced traps in the silicon film using the noise spectroscopy technique are revisited.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Low frequency noise measurements can be used as a non-destructive diagnostic tool that leads to the identification of traps in the Si and the gate oxide, thus giving informations on the quality of the transistors fabrication. The study of the 1/f noise level with respect to the gate overdrive voltage V_{GT} leads to the identification of the 1/f noise origin. The study of the generation-recombination noise, if performed as a function of the temperature at fixed drain currents, gives access to information on the traps that are located in the depletion region of the transistor [1].

This paper summarizes the extraction methods and techniques that have been used in order to investigate triple-gate FinFETs issued from a sub-10 nm technological node.

Based on the expression of the drain current which takes into account surface roughness scattering impact on the effective carrier mobility, simple and accurate method for parameter extraction is based on the $Y = I_D/\sqrt{g_m}$ function. Indeed, the Y function exhibits a linear variation on the applied gate voltage and does not depend on the access resistance through the first mobility attenuation factor. Special attention is given for the threshold voltage,

low field carrier mobility and access resistance estimation which are needfully for 1/f low frequency noise modeling.

Then the low frequency noise theory is discussed, 1/f noise characterisation is described as it leads to the identification of the 1/f noise mechanism and an estimation of the gate oxide trap density. Finally the study of the depletion region traps is explained using a low frequency noise spectroscopy as a function of temperature.

2. Static parameters extraction

2.1. MOSFET model in linear operation region

MOSFETs are composed from three resistive regions: the source, the channel and the drain, as seen in Fig. 1. In the linear region of operation, the MOSFET drain current I_D in strong inversion can be expressed by [2]

$$I_D = \frac{W}{L} \mu_{eff} |Q_i| V_{D'S'} \quad (1)$$

W and L are the gate effective width and length, respectively, $V_{D'S'}$ is the voltage across the channel (between the points S' and D' of Fig. 1), and the inversion charge is usually taken as $Q_i = C_{ox}(V_{GS'} - V_{th} - V_{D'S'}/2)$, where $V_{GS'}$ is the gate-source voltage, V_{th} is

* Corresponding author.

E-mail address: dimitri.boudier@ensicaen.fr (D. Boudier).

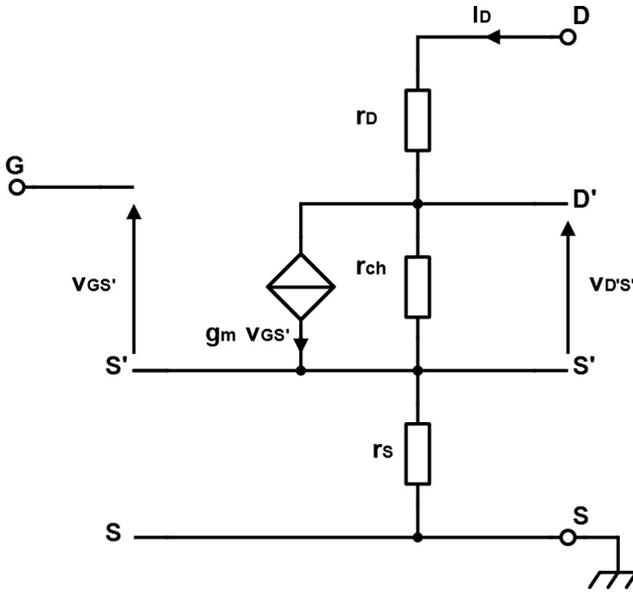


Fig. 1. MOSFET resistive model. The source (between nodes S and S'), the channel (between S' and D') and the drain (between D' and D) are the three resistive regions that compose the MOSFET.

the threshold voltage, and C_{ox} is the gate oxide capacitance per unit area.

The carrier mobility can be altered by several phenomena such as phonon scattering (thermal vibrations of the crystal lattice), Coulomb scattering (ionized impurities that deviates carriers) and surface roughness scattering. The mobility limited by the phonon scattering μ_{ph} , by the Coulomb scattering μ_C and by the surface roughness scattering μ_{sr} lead to the effective carrier mobility according to Matthiessen's rule [2]

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_C} + \frac{1}{\mu_{sr}} \quad (2)$$

In strong inversion (i.e. for strong transversal electric field) the Coulomb scattering is negligible, leading to $\mu_{eff}^{-1} \approx \mu_{ph}^{-1} + \mu_{sr}^{-1}$. The effective carrier mobility of the inversion channel carriers μ_{eff} depends of the low-field mobility μ_0 , as seen in [3]

$$\mu_{eff} = \frac{\mu_0}{1 + \theta_{10} \left(V_{GS'} - V_{th} - \frac{V_{DS'}}{2} \right) + \theta_2 \left(V_{GS'} - V_{th} - \frac{V_{DS'}}{2} \right)^2} \quad (3)$$

θ_{10} is called the intrinsic mobility attenuation factor and is due to phonon scattering, while θ_2 is the second order mobility attenuation factor, which is due to surface roughness scattering. Assuming that $R_S = R_D = R_{access}/2$, the drain current expression in the linear operation regime can be obtained by replacing Q_i and μ_{eff} in (1)

$$I_D = \frac{G_M V_{DS} \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)}{1 + (\theta_{10} + G_M R_{access}) \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right) + \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2} \quad (4)$$

where $G_M = \mu_0 C_{ox} W / L$ is the transconductance parameter. The extrinsic mobility attenuation factor is defined by $\theta_1 = \theta_{10} + G_M R_{access}$ and takes into account the drain current attenuation due to the phonon scattering and the access resistance. The transistor transconductance is defined as the variation of the drain current with respect to the gate voltage $g_m = \partial I_D / \partial V_{GS}$ and yields

$$g_m = G_M V_{DS} \frac{1 - \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2}{\left[1 + \theta_1 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right) + \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2 \right]^2} \quad (5)$$

2.2. Y function method

The Y function has two main advantages: it does not depend on the extrinsic mobility attenuation factor θ_1 and thus is independent on the access resistance R_{access} , and does not use the drain current second derivative, in opposition to some other methods. However, it assumes that the low-field mobility μ_0 is independent of the gate length L . The Y function is based on the drain current and transconductance equations as [3–5]

$$Y(V_{GS}) = \frac{I_D}{\sqrt{g_m}} = \sqrt{\frac{G_M V_{DS}}{1 - \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2}} \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right) \quad (6)$$

In moderate inversion, when $\theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2 \ll 1$, $Y(V_{GS})$ should be linear and a linear fitting gives access to G_M and V_{th} from the slope and the x-intercept, respectively. These estimations allow building the effective mobility attenuation factor θ_{eff} and thus determining the values of θ_1 and θ_2 from

$$\theta_{eff}(V_{GS}) = \frac{G_M V_{DS}}{I_D} - \frac{1}{V_{GS} - V_{th} - \frac{V_{DS}}{2}} = \theta_1 + \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right) \quad (7)$$

This function should be linear and θ_1 and θ_2 can be extracted from the y-intercept and the slope, respectively. The following step consists in extracting the access resistance R_{access} and the intrinsic mobility attenuation factor θ_{10} from the θ_1 and G_M values linear plot, as

$$\theta_1 = \theta_{10} + G_M R_{access} \quad (8)$$

Once the parameters G_M , V_{th} , θ_1 and θ_2 have been extracted, another iteration of the Y_n function method can be performed with a new Y_{n+1} function as written in

$$Y_{n+1}(V_{GS}) = Y_n(V_{GS}) \sqrt{1 - \theta_2 \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right)^2} \quad (9)$$

This should eliminate the influence of θ_2 in (5) and new values of G_M , V_{th} , θ_1 and θ_2 can be extracted, until the values converge. Examples of drain current and transconductance concordance between the experimental data and models of the drain current and of the transconductance described by Eqs. (4) and (5) are shown in Fig. 2. The right y-axis indicates that the absolute relative error is less than 0.1% for the drain, which reflects the good precision of this extraction procedure.

Several parameters can be extracted once the Y function method has been applied to devices with different gate dimensions [3]. The effective gate dimensions L and W can be determined using G_M and the mask gate dimensions L_m and W_m , as

$$G_M = \frac{\mu_0 C_{ox} W}{L} = \frac{\mu_0 C_{ox}}{L} (W_m - \Delta W) \quad (10.a)$$

$$\frac{1}{G_M} = \frac{L}{\mu_0 C_{ox} W} = \frac{1}{\mu_0 C_{ox} W} (L_m - \Delta L) \quad (10.b)$$

Assuming μ_0 is constant, both equations lead to a straight line, where the slope and the x-intercept of $G_M(W_m)$ or $G_M^{-1}(L_m)$ lead to μ_0 and ΔW or ΔL , respectively.

The extrinsic mobility attenuation factor θ_1 depends of the gate dimensions W and L , and on the access resistance R_{access} according to (7). Consequently a $\theta_1(G_M)$ function will give access to the intrinsic mobility attenuation factor θ_{10} and the drain-source resistance R_{access} .

The channel length reduction ΔL can also be determined with another method that consists in tracing the evolution of the total

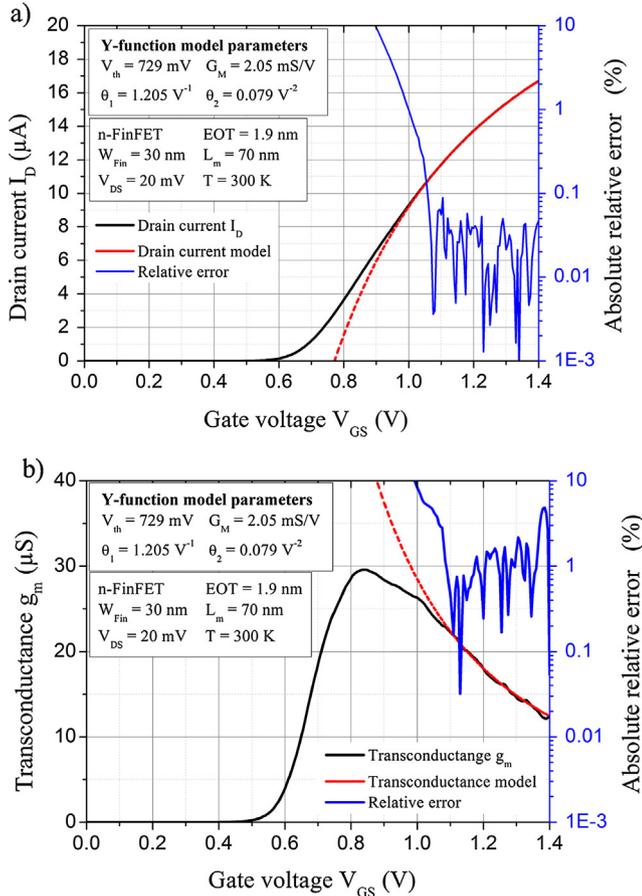


Fig. 2. MOSFET drain current (a) and transconductance (b). The Y function method leads to a model for both characteristics in strong inversion (in red). The relative error (in blue) in strong inversion remains below 5% for the transconductance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

resistance (i.e. of I_D/V_{DS}) at different gate voltages V_{GS} for different gate mask lengths, using the channel resistivity ρ_{ch} in the relation

$$R_{tot} = \frac{L}{W} \rho_{ch} + R_{access} = \frac{L_m - \Delta L}{W} \rho_{ch} + R_{access} \quad (11)$$

All lines should intercept at the same point, which coordinates are ΔL and R_{access} , as seen in Fig. 3.

3. Low frequency noise

3.1. Low frequency noise in MOSFETs

The characterization of the low frequency noise in field-effect transistors is performed from measurements of the noise power spectral density (PSD), which is usually referred either to the drain current (S_{ID} in A^2/Hz) or to the gate voltage (S_{VG} in V^2/Hz). Both quantities are linked together by the MOSFET transconductance as $S_{VG} \cdot g_m^2 = S_{ID}$.

The noise mainly originates from the resistive parts of the transistors, which are the source region (r_s), the drain region (r_D) and the channel (r_{ch}). The low frequency noise equivalent circuit is given in Fig. 4, where the resistive (noisy) parts have been modeled by uncorrelated current noise sources. The drain region is between the points D and D' , the channel is between D' and S' and the source region is between S' and S . From this schematic we can write the total drain current noise $i_{D,tot}$ as

$$i_{D,tot} = g_m \cdot v_{GS'} + g_{ch} \cdot v_{D'S'} + i_{ch} \quad (12)$$

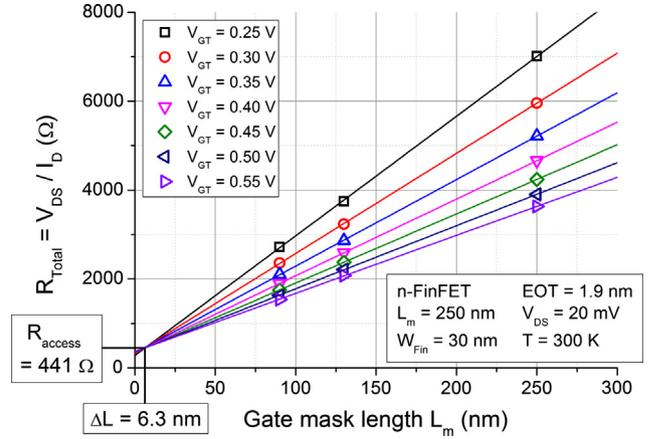


Fig. 3. MOSFET static parameters extraction. The access resistance R_{access} and the channel length reduction ΔL are obtained from the intercept of $R_{tot}(L_m)$ lines at different V_{GT} .

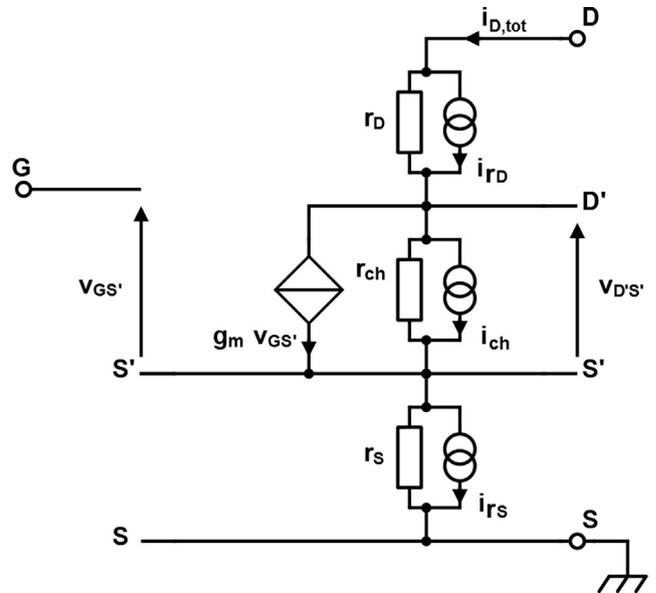


Fig. 4. MOSFET noise model with a representation of the three noisy region, which are the drain region (between nodes D and D'), the inversion channel (between nodes D' and S') and the source region (between nodes S' and S).

where $g_{ch} = r_{ch}^{-1}$ is the channel conductance. The $v_{GS'}$ and $v_{D'S'}$ voltages are

$$v_{GS'} = -v_{S'S} = -r_S(i_{D,tot} - i_{rS}) \quad (13.a)$$

$$v_{D'S'} = -v_{D'D} - v_{S'S} = -r_D(i_{D,tot} - i_{rD}) - r_S(i_{D,tot} - i_{rS}) \quad (13.b)$$

From (12) and (13) the total current noise in MOSFETs can be expressed by

$$i_{D,tot} = \frac{i_{ch} + i_{rD} \cdot g_{ch} r_D + i_{rS} \cdot r_S (g_m + g_{ch})}{1 + g_m r_S + g_{ch} (r_S + r_D)} \quad (14)$$

The total current noise PSD $S_{I_{D,tot}}$ can be written as a function of the channel current noise PSD $S_{I_{ch}}$, the drain region noise PSD $S_{I_{rD}}$ and the source region noise PSD $S_{I_{rS}}$

$$S_{I_{D,tot}} = \frac{S_{I_{ch}} + S_{I_{rD}} (g_{ch} r_D)^2 + S_{I_{rS}} [r_S (g_m + g_{ch})]^2}{[1 + g_m r_S + g_{ch} (r_S + r_D)]^2} \quad (15)$$

It is convenient to assume that the source and the drain region behave the same way, so that $r_D = r_S = r_{access}/2$ and $S_{I_{D}} = S_{I_{S}}$. In linear operation can be considered $g_m \ll g_{ch}$, then (15) yields

$$S_{I_{D,tot}} = S_{I_{ch}} \frac{(r_{tot} - r_{access})^2}{r_{tot}^2} + S_{I_r} \frac{r_{access}^2}{2r_{tot}^2} \quad (16)$$

where r_{tot} is the transistor total dynamic resistance (between the points D and S in Fig. 3).

3.2. White noise

The white noise is characterized by its level K_w , which is constant in the frequency domain, and has two main origins. The first one is the thermal noise (Johnson noise [6] or Nyquist noise [7]), which is due to the thermal motions of the electrons in the crystal lattice. Its voltage noise PSD is equal to $S_{V_G} = 4 k \cdot T \cdot R$, k being the Boltzmann constant, T the temperature and R the resistance of the piece of material.

The second one is the shot noise and has been described by Schottky in 1918 [8]. It is due to the random number of charge carriers that flows across a potential barrier at a given time. For an average current I flowing through a potential barrier, the current noise PSD is $S_{ID} = 2 q \cdot I$.

3.3. Generation-recombination noise

The generation-recombination (GR) noise is represented in the frequency domain by a Lorentzian shape, which consists in a plateau of level A followed by a $1/f^2$ decrease starting at the characteristic frequency f_0 . The latter parameter is linked to the characteristic time constant τ . GR noise is due to traps (in the gate oxide and in the depletion region) that randomly capture and release channel carriers, resulting in fluctuations of the available carriers for the current flow. Those traps may also induce changes in the carrier mobility, the scattering coefficient, the electric field and the potential barrier. It becomes significant when the Fermi energy level is close to the trap energy level, meaning that it is very sensitive to temperature [9].

Several Lorentzians can appear in the frequency spectrum, each having different values of A_i and $f_{0,i}$ (and τ_i) following (17) [10]. If the trap characteristic time constants are distributed in a precise way, the sum of the Lorentzians can have a $1/f$ shape over several decades of frequency.

$$S_{V_{G,lor_i}} = \frac{A_i}{1 + \left(\frac{f}{f_{0,i}}\right)^2} = \frac{A_i}{1 + (2\pi f \tau_i)^2} \quad (17)$$

If the gate surface is small enough (usually below $1 \mu\text{m}^2$) only a few traps are involved in the generation-recombination noise, meaning that the number of channel carriers being captured and released at a given time is also small. Thus the drain current displays a typical behavior in the time domain as a burst noise (also called Random-Telegraph-Signal, or RTS noise) appears superposed on the “classical” noise.

From [11] the trap characteristic time constant τ of a generation-recombination process is given by

$$\tau = \frac{1}{C_n(n_1 + n) + C_p(p_1 + p)} \quad (18)$$

with C_n and C_p the electrons and holes capture coefficients, respectively, n and p the free electrons and holes concentrations, respectively, n_1 and p_1 the concentrations of free electrons and holes under when the Fermi level coincides with the noisy trap level. In any case n , p , n_1 and p_1 vary with the temperature following an exponential dependency, which means that Lorentzians character-

istic time constants are temperature dependent. Nevertheless the variation of the characteristic time constants on the gate voltage depends on the localization of the traps in the transistor (either in the depletion film or at the Si-SiO₂ interface).

For Lorentzians that are related to interface traps, the Fermi level always coincides with the traps energy level [12] and $n = n_1$ and $p = p_1$, then $\tau = [2(C_n \cdot n_1 + C_p \cdot p_1)]^{-1}$. As the gate voltage $|V_{GS}|$ increases, the distance between the Fermi level and the edge of the majority carrier band increases, which leads to an increase of τ . However for a further increase of $|V_{GS}|$ the Fermi level approaches the minority carrier band and τ then decreases.

In the depletion region, free electrons and holes are practically absent, so the approximation $n = p = 0$ can be used in (18) and yields $\tau = (C_n \cdot n_1 + C_p \cdot p_1)^{-1}$. Traps close to the midgap have a concentration of $n_1 \approx p_1 \approx n_i$, where n_i is the intrinsic carrier concentration. This leads to a value of $\tau = [n_i \cdot (C_n + C_p)]^{-1}$, which is independent of the gate voltage $|V_{GS}|$.

3.4. 1/f noise

Several models have been proposed in order to explain $1/f$ noise in the inversion layer. The carrier number fluctuations model (ΔN) was first described by McWhorter in 1957 and it suggests that carrier tunneling into the oxide is the origin of the $1/f$ noise [13]. Carrier mobility fluctuations model ($\Delta\mu$) described by Hooge in 1969 explains that the $1/f$ noise is due to phonon scattering [14]. A carrier number fluctuations correlated to mobility fluctuations model ($\Delta N + \Delta\mu$) has been suggested in 1990 [15,16]. It states that oxide traps that capture and release channel carriers change the number of available carriers, which leads to surface mobility fluctuations.

In the case of carrier mobility fluctuations model, the gate voltage noise PSD of the channel can be written as [3]

$$S_{V_G} = \frac{q\alpha_H}{fWLC_{ox}} (V_{GS} - V_{th}) [1 + \theta_1 (V_{GS} - V_{th})]^2 \quad (19)$$

where α_H is the Hooge parameter and depends of the crystal quality. In weak inversion, the normalized drain current noise PSD S_{I_D}/I_D^2 is inversely proportional to $V_{GS} - V_{th}$.

For the $\Delta N + \Delta\mu$ model, the gate voltage noise PSD can be expressed by [17]

$$S_{V_G} = S_{V_{fb}} (1 \pm \alpha_C \mu_{eff} C_{ox} V_{GT})^2 \quad (20)$$

where $S_{V_{fb}}$ is the flat-band spectral density and α_C is the Coulomb scattering coefficient. The + and the - signs apply for donor and acceptor traps, respectively. This equation requires no assumption about the mechanism behind the fluctuations in the flat-band voltage [18]. However if $\alpha_C \mu_{eff} C_{ox} V_{GT} \ll 1$, then $S_{V_G} = S_{V_{fb}}$ and only carrier number fluctuations noise is observed.

As the noise of the access resistance can be modeled by the formula $S_{I_D,access} = \frac{K_r}{f} I_D^2$ [19], with K_r being the access resistance noise parameter, we can write from (16), (19) and (20) the total gate voltage spectral density in ohmic region operation

$$S_{V_G} = \frac{(r_{tot} - r_{access})^2}{r_{tot}^2} \frac{q\alpha_H}{fWLC_{ox}} (V_{GS} - V_{th}) [1 + \theta_1 (V_{GS} - V_{th})]^2 + \frac{K_r}{f} \frac{r_{access}^2}{2r_{tot}^2} \frac{I_D^2}{g_m^2} \quad (21.a)$$

$$S_{V_G} = \frac{(r_{tot} - r_{access})^2}{r_{tot}^2} S_{V_{fb}} (1 + \alpha_C \mu_{eff} C_{ox} V_{GT})^2 + \frac{K_r}{f} \frac{r_{access}^2}{2r_{tot}^2} \frac{I_D^2}{g_m^2} \quad (21.b)$$

The first term of (21.a) and (21.b) corresponds to the $\Delta\mu$ model and the $\Delta N + \Delta\mu$ model, respectively. The second term corresponds to the access resistance noise, usually observed in strong inversion.

4. Methodology

4.1. Experimental bench

The device under test (DUT) is placed into a Lakeshore TTP4 cryogenic probe station. The low frequency noise experimental bench uses home-made electronics and a HP 3562A dynamic signal analyzer. A simplified schematic of these electronics is shown in Fig. 5. The transistor is biased with V_{DS} and V_{GS} voltage sources, the resulting drain current I_D flows through the retroaction resistance of the transimpedance amplifier and the voltage measured by the voltmeter allows checking the bias point, as $V_{OUT} = V_{DS} + R \cdot I_D$.

The drain noise current i_D is converted into a voltage after the transimpedance stage, then this voltage passes through the high-pass filter (cut-off frequency of 0.08 Hz) and is amplified by an instrumentation amplifier. The spectrum analyzer finally provides the output voltage noise PSD $S_{V_{out}}$ from 1 Hz to 100 kHz.

In order to get the input-referred noise PSD (i.e. the gate voltage noise PSD), white noise is injected into the gate (v_{gs} voltage source in Fig. 5) after the output PSD measurements. The ratio between the output voltage and the white noise voltage gives the whole system gain, which allows to get the gate voltage noise PSD as $S_{V_G} = S_{V_{out}} / \text{gain}^2$ and to get rid of the system bandwidth limitation.

4.2. Noise model

The low frequency noise is generally due to the contributions of three distinct noises, which are the white noise (of a constant level K_w), the $1/f$ noise (of level K_f) and the generation-recombination noise. The latter can present several Lorentzians (of plateau level A_i and characteristic frequency $f_{0,i}$) in the noise PSD. The gate voltage noise PSD S_{V_G} can be seen as the sum of these three noise sources and is written

$$S_{V_G}(f) = K_w + \frac{K_f}{f^\gamma} + \sum_{i=1}^N \frac{A_i}{1 + (f/f_{0,i})^2} \tag{22}$$

A value of $\gamma = 1$ means that the traps are uniformly distributed in the gate oxide depth. Fig. 6 shows a gate voltage noise PSD model and its different components using (11).

However, in order to better extract the noise parameters, one can use a noise PSD normalized by the frequency, as displayed in Fig. 7. The $1/f$ noise is then represented by a constant value while the Lorentzians correspond to bumps centered around their characteristic frequency $f_{0,i}$. As it is more convenient to find $f_{0,i}$, it reduces the estimation errors on the plateau levels A_i . Consequently using the model described by (22) leads to the white noise level K_w , the $1/f$ noise level K_f and the Lorentzians parameters A_i and $f_{0,i}$.

4.3. $1/f$ noise mechanism and oxide trap density

In order to determine the $1/f$ noise mechanism, one has to perform noise measurements as a function of the gate voltage

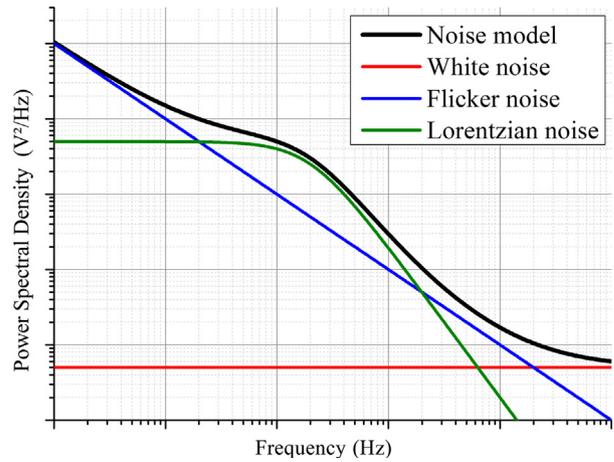


Fig. 6. Total noise PSD model (in black). The total noise PSD is the sum of the white noise, the $1/f$ noise and one Lorentzian.

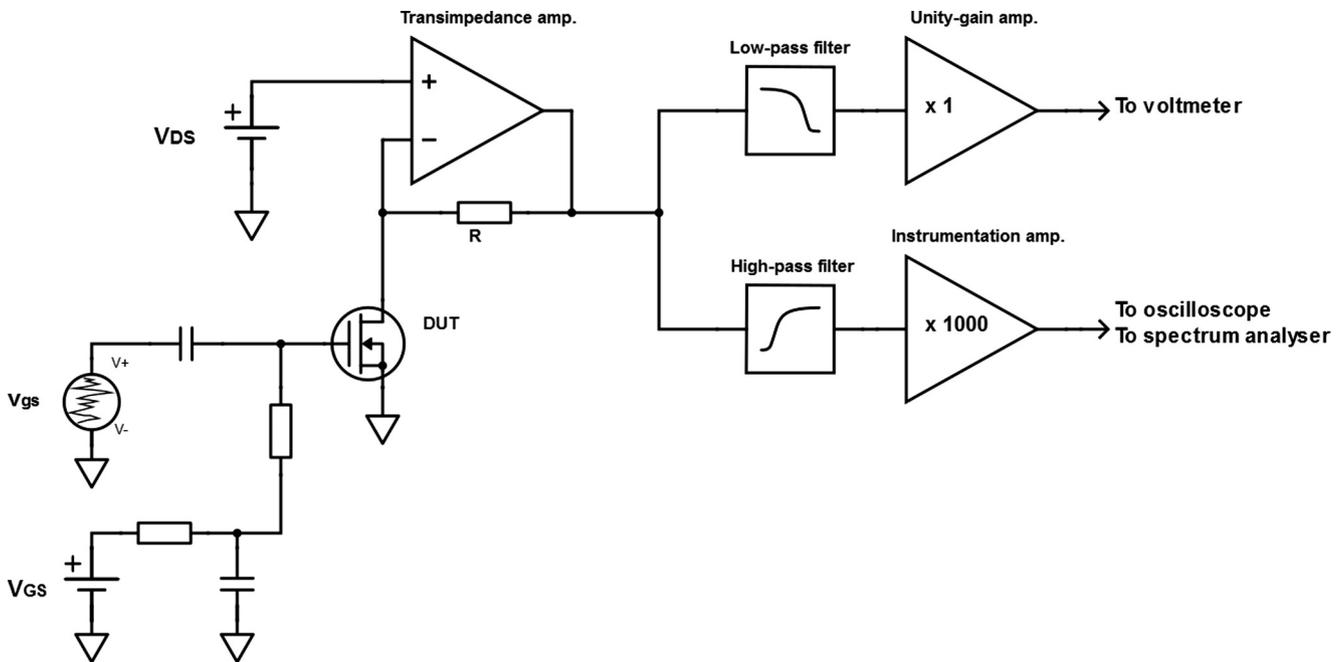


Fig. 5. Simplified schematic of the low frequency noise measurements bench.

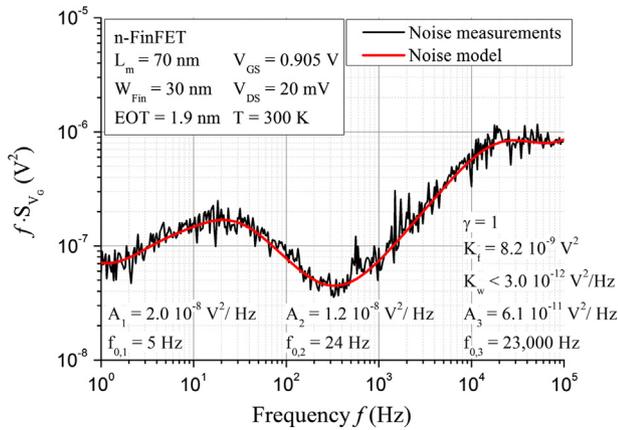


Fig. 7. Frequency-normalized gate voltage noise PSD. The noise model (20) allowed to extract the noise parameters (white noise level K_w , $1/f$ noise level K_f and characterization of three Lorentzians).

V_{GS} (or drain current I_D). For each noise PSD, a noise model must be performed using (21), which leads to the $1/f$ noise level K_f . The evolution of K_f as a function of V_{GS} (or I_D) leads to the $1/f$ noise mechanism, according to (21.a) and (21.b). Table 1 summarizes the evolution of S_{V_G} and S_{I_D}/I_D^2 using (21.a) and (21.b).

As the values of the effective mobility μ_{eff} , the mobility attenuation factor θ and the access resistance r_{access} have been extracted with static measurements, the coulomb scattering coefficient α_C and the flat-band voltage PSD $S_{V_{fb}}$ (for $\Delta N + \Delta\mu$ model) or the Hooge parameter α_H (for $\Delta\mu$ model) can be determined. In the case of number fluctuations correlated to mobility fluctuations, the flat-band voltage PSD leads to an estimation of the oxide trap density N_T , according to the formula

$$S_{V_{fb}} = \frac{q^2 k T \lambda N_T}{f^2 W L C_{ox}} \quad (23)$$

where λ is the tunneling attenuation length in the gate oxide ($\approx 1 \text{ \AA}$ for Si/SiO₂).

4.4. Silicon film traps: noise spectroscopy

As for studying the $1/f$ noise, the Lorentzian parameters (plateau level A_i and characteristic frequency $f_{0,i}$) are obtained from noise PSD modelling at different values of V_{GS} and at fixed temperature, using (22). As explained above, Lorentzians with a characteristic frequency that is independent of the gate voltage are considered to be related to traps in the depletion region whereas Lorentzians with a characteristic frequency that depends on V_{GS} are located in the gate oxide. However their characteristic frequency should vary with the temperature in both cases. Lorentzians with fixed $f_{0,i}$ as a function of V_{GS} have been studied in order to identify the physical nature of the Si film traps. To do so, one must perform noise spectroscopy, i.e.

noise measurements at different temperatures and at constant drain current I_D .

The characteristic time constant τ_i of a single trap varies with temperature, according to the expression [1]

$$\ln(\tau_i T^2) = (E_C - E_T) \frac{q}{kT} + \ln \left(\frac{h^3}{4k^2 \sigma_n \sqrt{6\pi^2 M_c m_e^{*1/2} m_h^{*3/2}}} \right) \quad (24)$$

where σ_n is the trap capture cross section and $E_C - E_T$ is the difference between the conduction band energy level and the trap energy level. k is the Boltzmann constant, h is the Planck constant, m_e^* and m_h^* are the effective mass of electrons and holes, respectively, and M_c is the number of conduction band energy minima.

From the characteristic time constant τ_i and the temperature, one can trace the evolution of $\ln(\tau_i T^2)$ with respect to q/kT , which is an Arrhenius diagram and should give a straight line. By using (23) the estimated capture cross section σ_n and energy difference $E_C - E_T$ can be determined and compared with data in the literature. Indeed Arrhenius diagrams of common traps are known from other methods, like Deep Level Transient Spectroscopy. If experimental points fit one particular line, then the associated Lorentzians correspond to the same particular trap.

Once traps have been identified, the associated plateau levels A_i should be traced against the characteristic time constants τ_i . This should give a straight line and lead to the surface (effective) trap density N_{eff} for each trap, according to

$$A_i = \frac{q^2 N_{eff} \tau_i}{W L C_{ox}^2} = \frac{q^2 W_d N_T \tau_i}{W L C_{ox}^2 B} \quad (25)$$

where q is the elementary charge, N_T is the volume trap density, W_d is the depletion depth and B is a coefficient. B is usually predicted to be $1/3$ for planar devices [19], however this does not apply well for multi-gate devices [20]. Consequently, even though silicon film traps are related to a volume phenomenon, the surface trap density N_{eff} will be extracted as it requires no assumption of the coefficient B .

5. Conclusion

This paper has presented the main noise parameters that can be extracted, either from the $1/f$ noise or the generation-recombination noise. First static characteristics need to be performed in order to extract DC parameters that will be used for noise characterization, such as the threshold voltage V_{th} , the low-field mobility μ_0 , the mobility attenuation factor θ and the access resistance r_{access} . Then gate voltage noise measurements can be performed in order to determine the $1/f$ noise mechanism and the gate oxide trap density N_T . Finally a low frequency noise spectroscopy leads to the identification of depletion region traps and an estimation of their effective densities N_{eff} .

Table 1

Evolution of the $1/f$ noise PSD (S_{V_G} and S_{I_D}) with respect to the gate overdrive voltage V_{GT} or the drain current I_D , according to the McWhorter model ($\Delta N + \Delta\mu$), the Hooge model ($\Delta\mu$), and the access resistance noise model.

	Channel noise: $\Delta N + \Delta\mu$ fluctuations		Channel noise: $\Delta\mu$ fluctuations	Access resist. noise
	$\alpha_C \mu_{eff} C_{ox} V_{GT} \ll 1$	$\alpha_C \mu_{eff} C_{ox} V_{GT} \gg 1$		
$S_{V_G} \propto$	Constant	$(1 \pm \alpha_C \mu_{eff} C_{ox} V_{GT})^2$	$V_{GT} [1 + \theta_1 V_{GT}]^2$	$\frac{f_0}{(r_{tot} g_m)^2}$
$\frac{S_{I_D}}{I_D} \propto$	$\left(\frac{g_m}{I_D}\right)^2$	$\left(\frac{g_m}{I_D}\right)^2 [1 \pm \alpha_C \mu_{eff} C_{ox} V_{GT}]^2$	$\frac{1}{V_{GS} - V_{th}}$	$\frac{1}{r_{tot}^2}$

References

- [1] Grassi V, Colombo CF, Camin DV. Low frequency noise versus temperature spectroscopy of recently designed Ge JFETs. *IEEE Trans Electron Dev* 2001;48:2899–905.
- [2] Lee K et al. Physical understanding of low-field carrier mobility in silicon MOSFET inversion layer. *IEEE Trans Electron Dev* 1991;38:1905–12.
- [3] Ghibaudo G. Critical MOSFETs operation for low voltage/low power IC's: ideal characteristics, parameter extraction, electrical noise and RTS fluctuations. *Microelectron Eng* 1997;39:31–57.
- [4] Mourrain C et al. New method for parameter extraction in deep submicrometer MOSFETs. In: *IEEE ICMTS*. p. 181–6.
- [5] Fleury D et al. New Y-function-based methodology for accurate extraction of electrical parameters on nano-scaled MOSFETs. In: *IEEE ICMTS*. p. 160–5.
- [6] Johnson JB. Thermal agitation of electricity in conductors. *Phys Rev* 1928;32:97–109.
- [7] Nyquist H. Thermal agitation of electric charge in conductors. *Phys Rev* 1928;32:110–3.
- [8] Schottky W. Über spontane Stromschwankungen in verschiedenen elektrizitätsleitern. *Ann Phys* 1918;57:541–67.
- [9] Von Haartman M, Östling M. Low frequency noise in advanced MOS devices. Springer; 2007.
- [10] Murray DC, Evans A, Carter JC. Shallow defects responsible for GR noise in MOSFETs. *IEEE Trans Electron Dev* 1991;38:407.
- [11] Schokley W, Read W, Hall R. Statistics of recombination of holes and electrons. *Phys Rev* 1952;87(5):835–42.
- [12] Lukyanchikova N. Sources of the Lorentzian components in the low-frequency noise spectra of submicron metal-oxide-semiconductor field-effect transistors. In: Baladin A, editor. *Noise and fluctuations control in electronic devices*. Riverside, CA: American Scientific Publishers; 2001. p. 201–33.
- [13] McWhorter AL. *Semiconductor surface physics*. Pennsylvania: Pennsylvania University Press; 1957.
- [14] Hooge FN. $1/f$ noise is no surface effect. *Phys Lett* 1969;A 29a:139–40.
- [15] Hung KK, Ko PK, Hu C, Cheng YC. A unified model for the flicker noise in metal-oxide-semiconductor field-effect-transistors. *IEEE Trans Electron Dev* 1990;37:654–65.
- [16] Ghibaudo G, Roux O, Nguyen-Duc Ch, Balestra F, Brini J. Improved analysis of low frequency noise in field-effect MOS transistors. *Phys Stat Sol (a)* 1991;124:571–81.
- [17] Ghibaudo G et al. Improved analysis of low frequency noise in field-effect MOS transistors. *Phys Status Solidi (a)* 1991;124–571.
- [18] Talmat R et al. Low frequency noise characterization in n-channel FinFETs. *Solid-State Electron* 2012;70:20–6.
- [19] Hooge FN. $1/f$ noise sources. *IEEE Trans Electron Dev* 1994;41:1926–35.
- [20] Achour H et al. Identification of Si film traps in p-channel SOI FinFETs using low temperature noise spectroscopy. *Solid-State Electron* 2015;112:1–6.



Low frequency noise assessment in n- and p-channel sub-10 nm triple-gate FinFETs: Part II: Measurements and results



D. Boudier^{a,*}, B. Cretu^b, E. Simoen^c, R. Carin^a, A. Veloso^c, N. Collaert^c, A. Thean^c

^a University of Caen Normandie, UMR6072 GREYC, F-14050 Caen, France

^b ENSICAEN, UMR6072 GREYC, F-14050 Caen, France

^c imec, Kapeldreef 75, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Triple-gate

FinFET

Low frequency noise

1/f noise

Generation recombination

Traps

ABSTRACT

Low frequency noise measurements are used as a non-destructive diagnostic tool in order to evaluate the quality of the gate oxide and the silicon film of sub-10 nm triple-gate Silicon-on-Insulator (SOI) FinFETs. It was found that the carrier number fluctuations explain the 1/f noise in moderate inversion for n- and p-FinFETs, which allows estimating the gate oxide trap densities. The noise spectroscopy with respect to temperature (study of the generation-recombination noise) led to the identification of the traps located in the transistors silicon film.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In order to meet the ITRS specifications in terms of CMOS down-scaling, new materials, designs and structures are necessary [1]. Triple-gate FinFETs are known for their good electrostatic performances and their compatibility with CMOS processes as a continuation of Moore's law [2].

Low-frequency noise measurements can be used as a non-destructive diagnostic tool that leads to the identification of traps in the Si film and the gate oxide, thus giving information on the quality of the transistors fabrication. The study of the 1/f noise level with respect to the gate overdrive voltage V_{CT} leads to the identification of the 1/f noise origin. The study of the generation recombination (GR) noise is performed as a function of the temperature at fixed drain current. This low frequency noise spectroscopy can give access to information on the traps located in the depletion region of the transistor [3].

In this work, static and dynamic parameters have been extracted in n- and p-channel triple-gate Silicon-on-Insulator (SOI) FinFETs. Low-frequency noise has been investigated in order to determine the quality of those devices. First the main static parameters of the devices have been extracted at 300 K, then the

1/f noise level has been studied as a function of the gate voltage and, finally, the generation-recombination noise has been investigated in function of temperature. The theory and methodology that are useful for the parameters extraction are explained in an accompanying Part I paper [4].

2. Devices and experimental

The tested devices have been processed at imec (Belgium) for sub-10 nm technological nodes, in the framework of a comparative study between triple-gate FinFETs and gate-all-around (GAA) nanowire (NW) FETs, fabricated on SOI substrates [5]. The gate stack consists of a high- κ dielectric (HfO_2) on top of a SiO_2 interfacial layer. Each layer is 1.5 nm wide, which leads to an equivalent oxide thickness $EOT = 1.9$ nm. The gate stack is followed by EWF (TiN) and W-fill metal depositions. All tested transistors have 5 fingers of 22–23 nm height; the finger width varies from $W_{fin} = 5$ nm to 40 nm, which gives a total gate width going from $W_m = 245$ nm to 420 nm. The finger pitch is 200 nm and the gate length varies from $L_m = 45$ nm to 10 μm . The transistors are built on a buried oxide (BOX).

The static and noise measurements have been performed on chip using a Lakeshore TTP4 prober. The DC characteristics have been obtained with a HP 4156B. Output noise power spectral densities (PSDs) have been measured using a home-made set-up,

* Corresponding author.

E-mail address: dimitri.boudier@ensicaen.fr (D. Boudier).

which includes a low-noise transimpedance amplifier and a HP 3562A dynamic signal analyzer. This set-up allows to bias transistors by choosing the V_{GS} , V_{DS} and V_{BS} voltages. The noise spectral densities are measured from 1 Hz to 100 kHz. The input referred gate voltage noise PSD S_{VG} is obtained by dividing the measured noise voltage by the square of the measured voltage gain between the gate and the system output. This cancels the set-up bandwidth limitation.

The static parameters extraction has been performed using $I_D(V_{GS})$ characteristics at 300 K. The drain voltage has been set to $V_{DS} = 20$ mV so the transistors works in the linear operation regime and the bulk has been connected to the source $V_{BS} = 0$ V. The low frequency noise has been measured as a function of temperature from 220 K to 300 K by steps of 20 K at fixed drain current I_D . As for static measurement the linear operation regime is reached with $V_{DS} = 20$ mV and $V_{BS} = 0$ V, then the gate voltage V_{GS} was adjusted in order to set the fixed drain current from $I_D = 1$ μ A to 5 μ A by 0.5 μ A steps. This paper shows the results obtained for n-channel and p-channel FinFETs with a finger width of $W_{fin} = 20$ nm and 30 nm.

3. Static parameters extraction

The transistors have been studied in linear operation ($V_{DS} = 20$ mV) at room temperature. Good behavior of the transfer DC characteristics $I_D(V_{GS})$ and $g_m(V_{GS})$ is observed for both n- and p-channels, as shown in Fig. 1.

Static parameters extraction has been further processed using the Y function method ($Y = I_D/\sqrt{g_m}$ [3]), which gives access to the threshold voltage V_{th} and the transconductance parameter $G_M = \mu_0 C_{ox} W/L$. The values that will be discussed in this paper have been obtained for two iterations of the Y function method.

Once the basic DC parameters have been determined with the Y function method for different gate lengths L_m , the low-field mobility μ_0 and the access resistances R_{access} can also be extracted from a set of transistors with different gate lengths [4]. As the transconductance parameter G_M is defined by $G_M = \mu_0 C_{ox} W/L$, one can write

$$\frac{1}{G_M} = \frac{L}{\mu_0 C_{ox} W} = \frac{1}{\mu_0 C_{ox} W} (L_m - \Delta L) \quad (1)$$

The low-field mobility μ_0 and the channel length reduction ΔL values can be determined from the slope and the y-intercept of $G_M^{-1}(L_m)$, respectively. The points corresponding to different gate lengths are shown in Fig. 2, on which a linear regression has been performed in order to estimate μ_0 and ΔL . For these n-FinFETs with $W_{fin} = 30$ nm, we have found that $\mu_0 = 259$ cm²/(V·s) and $\Delta L = 4.8$ nm.

The effective (extrinsic) mobility attenuation factor θ_1 is defined by

$$\theta_1 = \theta_{10} + G_M R_{access} \quad (2)$$

thus the plot of $\theta_1(G_M)$ should give a straight line where the slope is equal to R_{access} and the y-intercept is equal to θ_{10} , as shown in Fig. 3. For the same n-FinFETs ($W_{fin} = 30$ nm), values of $R_{access} = 370$ Ω and $\theta_{10} = 707$ mV have been recorded.

The values of the access resistances have been confirmed using the linear variations of the total resistance for various gate overdrive voltage V_{GT} against the gate mask length L_m . The value of the access resistance is determined at the common intersection of all lines, as shown in Fig. 4. In this case values of $R_{access} = 375$ Ω and $\Delta L = 5.5$ nm have been derived, which confirms the results that have been obtained with the previously explained method.

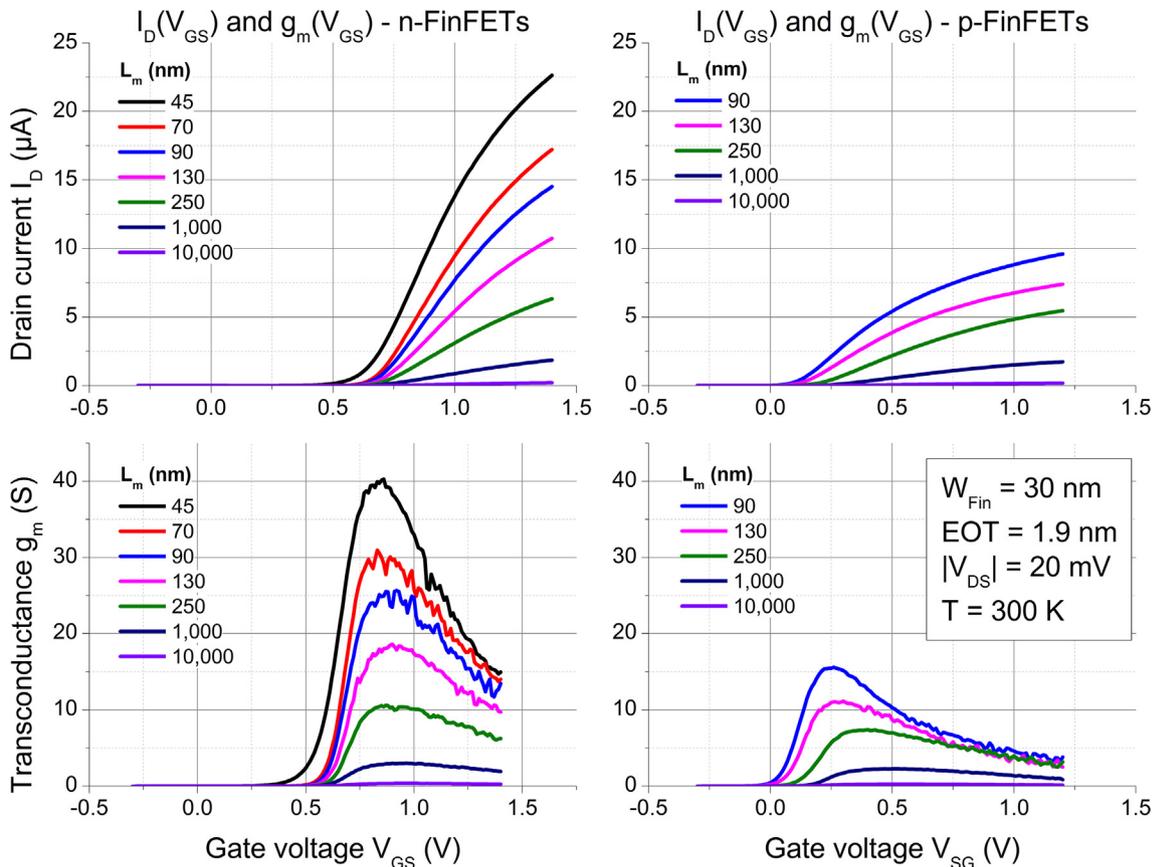


Fig. 1. Typical $I_D(V_{GS})$ and $g_m(V_{GS})$ characteristics of n- and p-FinFET for various gate lengths.

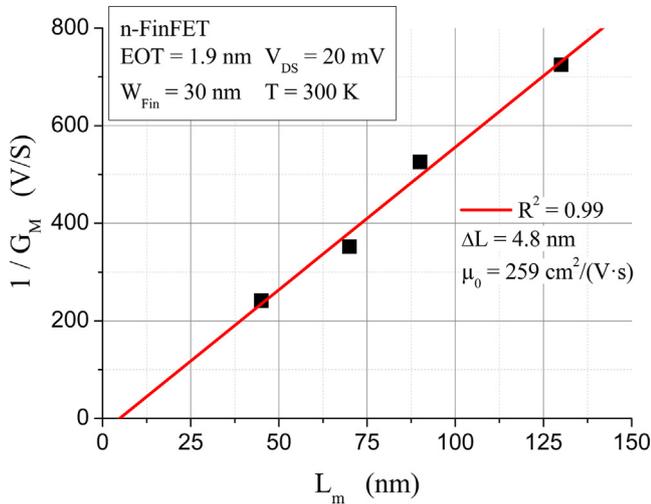


Fig. 2. Extraction of the low-field mobility μ_0 and gate length reduction ΔL from the linearisation of $G_M^{-1}(L_m)$.

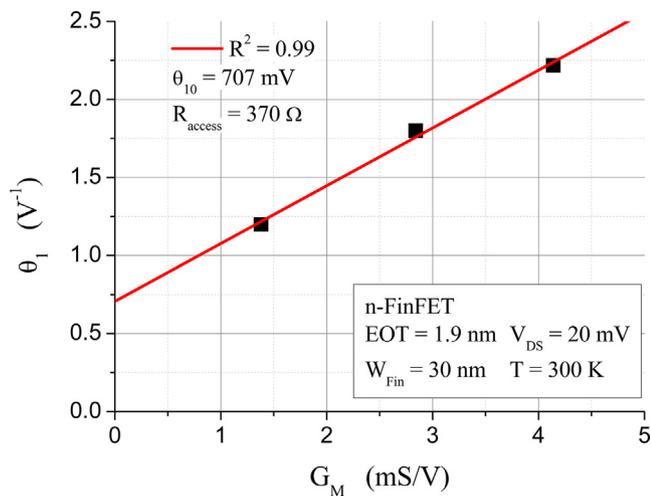


Fig. 3. Extraction of the intrinsic mobility attenuation factor θ_{10} and access resistance R_{access} from the linearisation of $\theta_1(G_M)$.

A synthesis of the estimated static parameters is showed in Table 1. It can be observed that the threshold voltage V_{th} is higher (in absolute value) for n-channel transistors than for p-channel transistors. The charge sharing effect is more pronounced for p-channel devices, in particular for smaller finger widths. Higher values of the low-field mobility μ_0 can be observed for thinner fingers in both n- and p-channel transistors; this could be related to the effective values of the gate width and to the relative contribution of the sidewall versus top surface conduction. The access resistances are higher for p-channel FinFETs compared to the n-channel ones. The increase of the access resistance with the finger width decrease from 30 nm to 20 nm, observed for p-FinFETs, could be related to the effective values of the gate width.

4. 1/f noise

All the noise PSDs can be modeled using a combination of three noise sources: white noise (of level K_w), 1/f noise (or flicker noise, of level K_f) and generation-recombination (GR) noise (of plateau A_i and characteristic frequency $f_{0,i}$). Assuming that the three noise sources are uncorrelated, the total noise PSD follows the equation

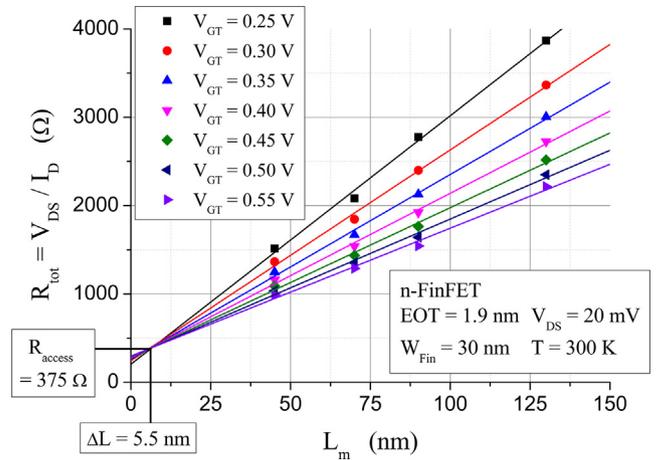


Fig. 4. Example of DC parameters extraction, giving the value of the access resistance R_{access} .

$$S_{V_G}(f) = K_w + \frac{K_f}{f} + \sum_{i=1}^N \frac{A_i}{1 + (f/f_{0,i})^2} \quad (3)$$

As seen in [4], the 1/f noise contribution of the total gate voltage spectral density S_{V_G} in ohmic operation can be expressed by

$$S_{V_G} = \frac{(r_{tot} - r_{access})^2}{r_{tot}^2} S_{V_{fb}} (1 + \alpha_C \mu_{eff} C_{ox} V_{GT})^2 + \frac{K_r r_{access}^2 I_D^2}{f 2r_{tot}^2 g_m} \quad (4)$$

where r_{tot} is the total dynamic resistance of the transistor between source and drain, r_{access} is the dynamic access resistance for both source and drain regions, $S_{V_{fb}}$ is the flat-band voltage power spectral density, α_C is the Coulomb scattering coefficient, μ_{eff} is the carrier effective mobility and K_r is an access resistance coefficient. The first term of (4) corresponds to the carrier number fluctuations correlated to mobility fluctuations model ($\Delta N + \Delta \mu$) in the inversion layer while the second term corresponds to the access resistance noise, usually observed in strong inversion.

From the gate voltage noise power spectral density S_{V_G} one can determine the 1/f noise level K_f with the model described in (3). By performing measurements at different values of gate voltage V_{GS} (thus at different fixed values of drain current I_D) at fixed temperature, one can study the variations of the flicker noise level as a function of the gate voltage in order to determine the 1/f noise mechanism.

The evolution of the 1/f noise level K_f with the gate overdrive voltage leads to the identification of the 1/f noise origin. Fig. 5 shows the evolution of K_f for both n- and p-channel devices and compares them to models given by (4). One can observe that n-FinFETs follow the carrier number fluctuations correlated to mobility fluctuations model ($\Delta N + \Delta \mu$). However, p-channel transistors trend to follow the number fluctuations model ΔN in moderate inversion while the total 1/f noise in strong inversion seems to originate from the access resistance noise contribution [6].

The extraction of the flat-band voltage spectral densities $S_{V_{fb}}$ leads to an estimation of the oxide trap densities N_T , according to the formula

$$S_{V_{fb}} = \frac{q^2 k T \lambda N_T}{f W L C_{ox}} \quad (5)$$

where λ is the tunneling attenuation length in the gate oxide ($\approx 1 \text{ \AA}$ for Si/SiO₂). Trap density values from $6.5 \cdot 10^{18}$ to $12 \cdot 10^{18} \text{ eV}^{-1} \cdot \text{cm}^{-3}$ for n-channel and from $0.8 \cdot 10^{18}$ to $4.6 \cdot 10^{18} \text{ eV}^{-1} \cdot \text{cm}^{-3}$ for p-channel FinFETs have been obtained, as shown in Table 2. These

Table 1
Synthesis of DC parameters of several n-FinFETs and p-FinFETs.

	W_{fin}/W_m (nm/nm)	L_m (nm)	V_t (mV)	μ_0 (cm ² /V·s)	R_{access} (Ω)
n-channel FinFETs	20/320	70	732	338	375
	20/320	250	793		
	20/320	1000	752		
	30/370	70	727		
	30/370	250	726		
	30/370	1000	740		
p-channel FinFETs	20/320	70	-69	190	1210
	20/320	250	-131		
	20/320	1000	-235		
	30/370	70	-144		
	30/370	250	-232		
	30/370	1000	-295		

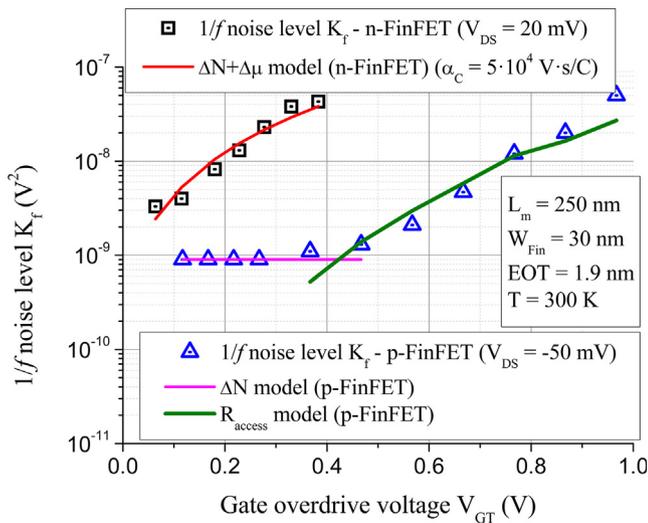


Fig. 5. Evolution of the $1/f$ noise level K_f with respect to V_{GT} , for a n- and a p-channel FinFET with the same gate dimensions. The red line corresponds to the $\Delta N + \Delta \mu$ model in moderate inversion for the n-FinFET (with $\mu_0 = 256$ cm²/Vs), the pink line corresponds to the ΔN model in moderate inversion for the p-FinFET and the green line represents the access resistance noise contribution in strong inversion for the p-FinFET. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

values show a small difference between n- and p-channel devices, the latter being slightly less impacted by $1/f$ noise. However, values of the oxide trap density around 10^{18} eV⁻¹·cm⁻³ show a good quality of the gate stack deposition process for both types.

5. Generation-recombination noise

The Lorentzians parameters can be studied once the gate voltage noise PSDs have been modeled using (3). Lorentzians with a characteristic frequency that is independent of the gate voltage

Table 2
Summary of flat-band voltage spectral densities S_{Vfb} and oxide trap densities N_T of several N- and P-FinFETs.

	W_{fin}/W_m (nm/nm)	L_m (nm)	S_{Vfb} (V ² /Hz)	N_T (eV ⁻¹ ·cm ⁻³)
n-channel FinFETs	20/320	70	$6 \cdot 10^{-9}$	$6.5 \cdot 10^{18}$
	20/320	250	$8 \cdot 10^{-9}$	$10 \cdot 10^{18}$
	30/370	70	$3 \cdot 10^{-9}$	$12 \cdot 10^{18}$
	30/370	250	$2 \cdot 10^{-9}$	$8.9 \cdot 10^{18}$
p-channel FinFETs	20/320	70	$2.2 \cdot 10^{-9}$	$2.4 \cdot 10^{18}$
	20/320	250	$0.6 \cdot 10^{-9}$	$0.75 \cdot 10^{18}$
	30/370	70	$1.2 \cdot 10^{-9}$	$4.6 \cdot 10^{18}$
	30/370	250	$0.7 \cdot 10^{-9}$	$3.1 \cdot 10^{18}$

are considered to be related to traps in the depletion region; however their characteristic frequency should vary with the temperature. Examples of the Lorentzians characteristic frequency $f_{0,i}$ evolution with respect to the gate overdrive voltage V_{GT} is shown in Fig. 6. For this n-FinFET at a fixed temperature, one can observe that the characteristic frequency of some Lorentzians increases with V_{GT} (may be related to traps located in the gate oxide), while the characteristic frequency of the other Lorentzians is constant with the applied gate voltage (may be related to traps in the depletion film) [7]. Only the latter have been studied for the noise spectroscopy, as shown in Fig. 7.

These Lorentzians have been studied in order to identify the physical nature of the Si film traps, by using [3]

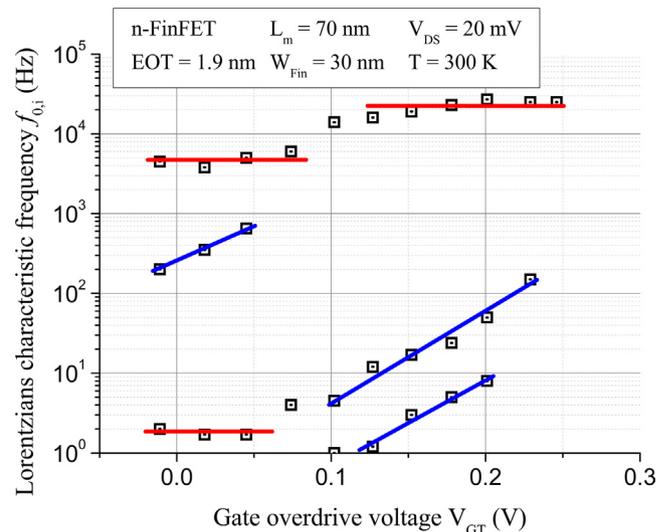


Fig. 6. Evolution of the Lorentzians characteristic frequency $f_{0,i}$ with respect to the gate overdrive voltage V_{GT} .

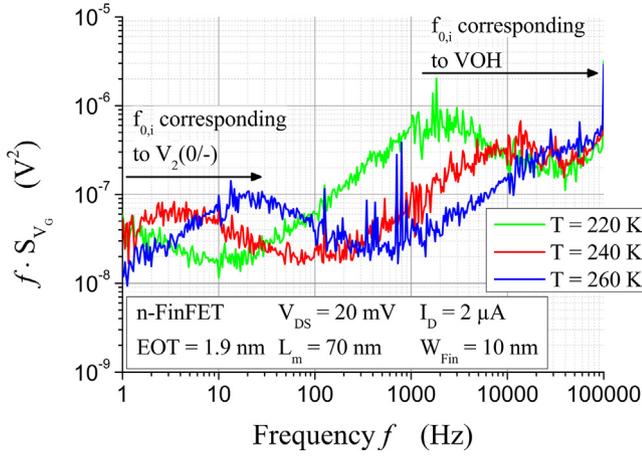


Fig. 7. Example of the gate noise PSD normalized by the frequency measured at different temperatures and at $V_{DS} = 20$ mV, $I_D = 2$ μ A. The variation of the Lorentzians characteristic frequency (associated to Si fin traps) with temperature is represented by the arrows.

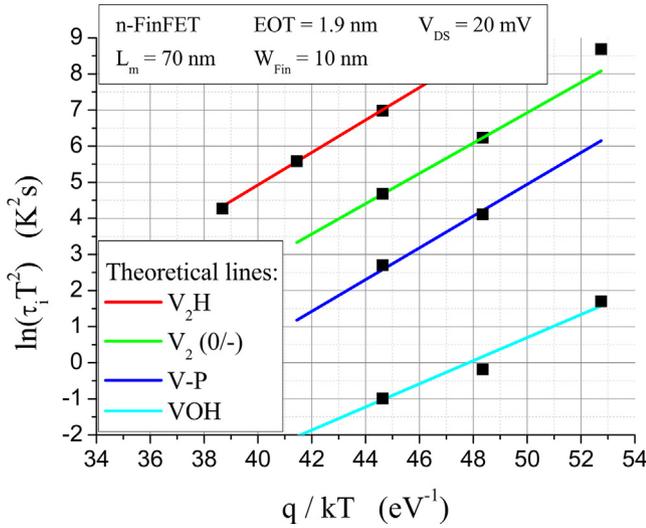


Fig. 8. Example of Arrhenius plot for a n-channel FinFET, leading to the identification of the Si fin traps.

Table 3
Summary of parameters of the identified traps.

	$E_C - E_T$ (eV)	σ_n (cm ²)	T (K)	N_{eff} (cm ⁻²)
V ₂ H	0.45	$1.4 \cdot 10^{-16}$	260–300	$4.32 \cdot 10^{10}$
V ₂ (0/-)	0.42	$3.8 \cdot 10^{-16}$	220–260	$0.862 \cdot 10^{10}$
V-P	0.44	$7.5 \cdot 10^{-15}$	240–260	–
VOH	0.32	$1.3 \cdot 10^{-15}$	220–260	$12.8 \cdot 10^{10}$

$$\ln(\tau_i T^2) = (E_C - E_T) \frac{q}{kT} + \ln \left(\frac{h^3}{4k^2 \sigma_n \sqrt{6\pi^2 M_c m_e^{*1/2} m_h^{*3/2}}} \right) \quad (6)$$

and comparing the estimated capture cross section σ_n and energy difference $E_C - E_T$ with data in the literature. The Lorentzians characteristic frequency evolution with the temperature leads to a typical Arrhenius plot, as shown in Fig. 8, for a n-channel FinFET with $W_m/L_m = 10$ nm/70 nm. The physical nature of these traps can be identified by comparing the energy and capture cross section of the traps with data in the literature. There are traps related to

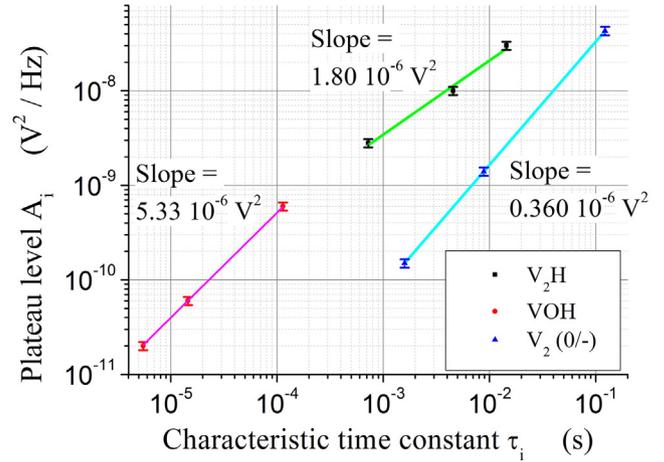


Fig. 9. Estimation of the effective trap densities N_{eff} from the linear variation of the Lorentzians plateau level A_i against the characteristic time constant τ_i (associated to the same trap).

hydrogen (V₂H and VOH) [8–10], the single negatively charged acceptor state (0/-) of the divacancy (V₂) [9,11] and the phosphorus-vacancy complex (VP) [10]. The presence of traps related to hydrogen could be explained by the hydrogen incorporated during the selective epitaxial growth of the raised source/drain from the SiH₄ precursors used in the chemical vapor deposition. Divacancies may be due to the recombination or the evolution to a stable state of the unstable defects like Frenkel pairs, which could be generated during the implantation [12]. The identified traps are shown in Table 3, which gives the energy difference between the conduction band and the trap level, the electron capture cross section and the temperature range of the traps.

Finally one can extract the surface trap density N_{eff} of the identified traps by plotting the Lorentzian plateau level against its characteristic frequency, according to

$$A_i = \frac{q^2 N_{eff} \tau_i}{WLC_{ox}^2} \quad (7)$$

The plateau levels A_i of Lorentzians corresponding to the identified traps have been traced against their characteristic time constant τ_i , as shown in Fig. 9. Using (7), the slope of the linear regressions leads to the effective surface trap density N_{eff} , which has been extracted for the V₂H, V₂ and VOH related traps. The estimated values of the effective surface trap densities are $4.32 \cdot 10^{10}$ cm⁻², $0.862 \cdot 10^{10}$ cm⁻² and $12.8 \cdot 10^{10}$ cm⁻² for V₂H, V₂(0/-) and VOH, respectively.

6. Conclusion

Low frequency noise studies showed that the $1/f$ noise mainly originates from the carrier number fluctuations. The study of the evolution of the $1/f$ noise level K_f with the gate overdrive voltage V_{GT} led to an estimation of the gate oxide trap density for n- and p-channel FinFETs. The estimated trap densities vary around $N_T \approx 10^{18}$ eV⁻¹·cm⁻³, which reflects a good quality of the gate stack processing as a comparison to 32-nm technology FinFETs [13].

The low frequency noise spectroscopy allowed studying the evolution of the generation-recombination noise as a function of the temperature. The evolution of the Lorentzians characteristic frequency led to the identification and the effective density of the Si film traps, which are either related to hydrogen or divacancies. From the nature of the traps, it can be suggested that the implantation and the use of selective epitaxial growth in the source and drain regions are mainly responsible for the observed traps.

References

- [1] International Technology Roadmap for Semiconductors. <<http://www.itrs.net>>.
- [2] Natarajan S et al. A 14nm logic technology featuring 2nd generation FinFET transistors, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size. IEDM Tech Dig 2014:71–3.
- [3] Grassi V, Colombo CF, Camin DV. Low frequency noise versus temperature spectroscopy of recently designed Ge JFETs. IEEE Trans Electron Dev 2001;48:2899–905.
- [4] Boudier D, et al. Low frequency noise assessment in n- and p-channel sub-10 nm triple-gate FinFETs: Part I: Theory and methodology. Solid-State Electron 2017;128:102–108.
- [5] Veloso A et al. Gate-all-around NWFETs vs. triple-gate FinFETs: junctionless vs. extensionless and conventional junction devices with controlled EWF modulation for multi-VT CMOS. VLSI Technol 2015:T138–9.
- [6] Lukyanchikova N. Sources of the Lorentzian components in the low-frequency noise spectra of submicron metal-oxide-semiconductor field-effect transistors. In: Baladin A, editor. Noise and fluctuations control in electronic devices. Riverside (CA): American Scientific Publishers; 2001. p. 201–33.
- [7] Ghibaudo G. Critical MOSFETs operation for low voltage/low power IC's: ideal characteristics, parameter extraction, electrical noise and RTS fluctuations. J Appl Phys 1997;39:31–57.
- [8] Tokuda Y, Shimizu, Usami A. Studies of neutron-produced defects in silicon by deep-level transient spectroscopy. Jpn J Appl Phys 1979;18(2):309–15.
- [9] Hallen A, Keskitalo N, Masszi F, Nagl V. Lifetime in proton irradiated silicon. J Appl Phys 1996;79(8):3906–14.
- [10] Claeys C, Simoen E. Radiation effects in advanced semiconductor material and devices. Springer Verlag; 2002.
- [11] Bains SK, Banbury PC. AC hopping conductivity and DLTS studies on electron-irradiated boron-doped silicon. Semicond Sci Technol 1978;2(1):20–9.
- [12] Agullo-Lopez F, Catlow CRA, Townsend PD. Point defects in materials. Academic Press; 1988.
- [13] Talmat R et al. Low frequency noise characterization in n-channel FinFETs. Solid-State Electron 2012;70:20–6.



Systematic method for electrical characterization of random telegraph noise in MOSFETs



Carlos Marquez^{a,*}, Noel Rodriguez^a, Francisco Gamiz^a, Akiko Ohata^b

^a Department of Electronics, CITIC-University of Granada, 18071 Granada, Spain

^b Collaboration Center for Research and Development, Utsunomiya University, Japan

ARTICLE INFO

Article history:

Available online 21 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Random telegraph noise
MOSFET reliability
Low frequency noise (LFN)

ABSTRACT

This work introduces a new protocol which aims to facilitate massive on-wafer characterization of Random Telegraph Noise (RTN) in MOS transistors. The methodology combines the noise spectral density scanning by gate bias assisted with a modified Weighted Time Lag Plot algorithm to identify unequivocally the single-trap RTN signals in optimum bias conditions for their electrical characterization. The strength of the method is demonstrated by its application for monitoring the distribution of traps over the transistors of a SOI wafer. The influence of the back-gate bias on the RTN characteristics of the SOI devices with coupled front- and back-interfaces has revealed unusual characteristics compatible with the carrier emission to the gate metal contact.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Random Telegraph Noise (RTN) prevails as a major constrain when the characteristic dimensions of the semiconductor devices are downscaled to the decananometer range [1,2]. Paired with the decrease of the signal levels, the introduction of new technologies, such as high-*k* metal-gate stacks and ultrathin Silicon-On-Insulator substrates, entails the appearance of additional contributions to the RTN, related with the quality of the new materials or the electrostatic coupling between the top-channel and the Si-film/BOX interface [3,4]. These new effects also require powerful characterization tools to cope with the massive monitoring of the transistors over the wafer, aiming to extract useful statistical information to nourish the optimization of the fabrication process flows. The systematic characterization protocol that we are proposing allows the detection of transistors with a single active trap, identifying the best operation region to characterize the RTN signature. This method is based on the Spectral Scanning by Gate Bias (SSGB) combined with a modified Weighted Time Lag Plot (w-TLP) approach.

On the following pages, Section 2 describes the methodology and the experimental setup on which this work is based. Section 3 introduces the Spectrum Scanning by Gate bias to easily trace the RTN in the time-domain, whereas in Section 4, we apply the Weighted Time Lag Plot to analyze the number of traps of the

devices over the wafer. Finally, in Section 5, we implement the technique to study the role of the substrate bias on the RTN characteristics of SOI transistors.

2. Methodology and experimental setup

The experimental method is summarized in the schematic of Fig. 1. At first, we characterize low frequency noise (LFN) of the device for different gate biases introducing the Spectral Scanning by Gate Bias approach. The output of this first stage, provides the optimum bias condition where the device is affected by RTN. Once the RTN signal is identified, the transient monitoring of the drain current, together with the application of the Weighted Time Lag Plot method, determine the number of active traps in the device. This selection of the suitable bias condition and the single-trap device identification, permit us to obtain the optimum RTN characterization of single-trap devices in an unambiguous protocol for the implementation in automatic testers.

All the experiments presented in this work are based on an ad hoc characterization setup consisting of:

- An Agilent B1517A high resolution Source-Measurement-Unit (SMU) monitoring repeatedly the drain current during periods of 400 s at a 2 ms sampling-rate (the schematic is shown in Fig. 2a). The acquisition periods were repeated until a minimum number of 100 transitions in the current signal were captured to guarantee the significance of the statistics.

* Corresponding author.

E-mail address: carlosmg@ugr.es (C. Marquez).

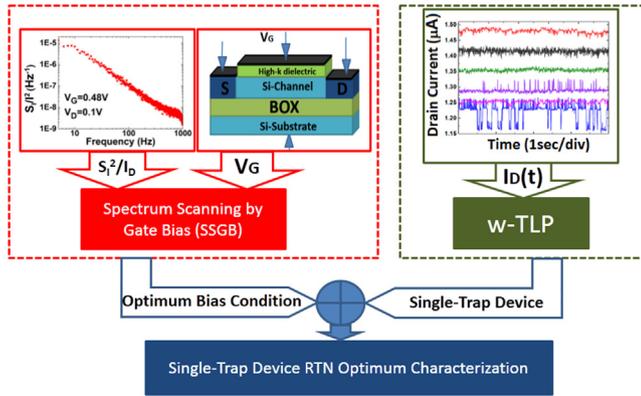


Fig. 1. Scheme of the experimental method proposed for the single-trap random telegraph characterization in transistors. The method permits the optimum RTN characterization of single-trap devices combining the Spectral Scanning by Gate Bias technique and the Time Lag Plot method.

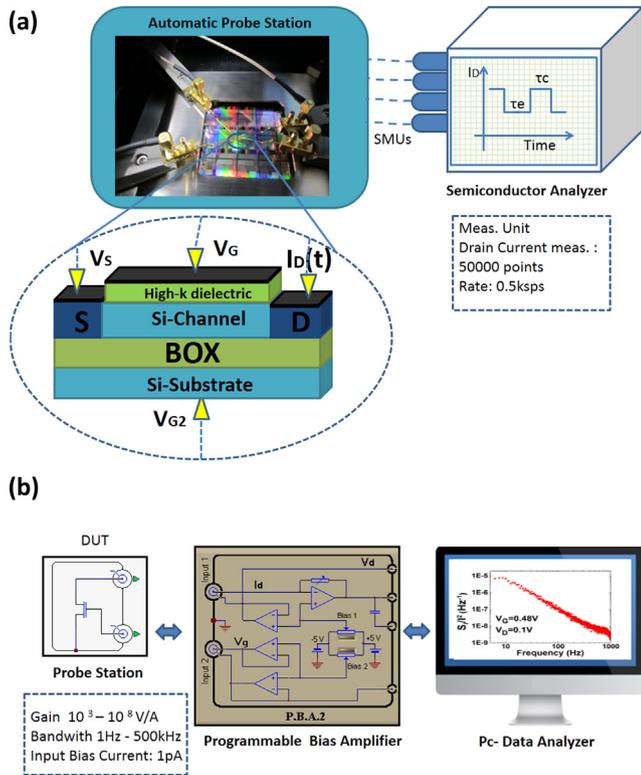


Fig. 2. (a) Schematic of the experimental setup developed for the RTN measurements. Each terminal of the transistor is monitored by an Agilent B1517A SMU at a rate of 500 samples per second. Automatic measurements around the wafer are implemented through a Süss semiautomatic probe station. (b) Schematic of the measurement setup used to extract the spectral noise power of the drain current based on a programmable low-noise amplifier.

- A Süss PA-300PS semiautomatic probe station GPIB-controlled by the Agilent SMU. This tool allows the fast location and scanning of all the devices over the wafer to extract information about the distribution of traps and the variability.
- A low-noise current amplifier connected to a software-based spectrum analyzer through a high resolution A/D converter (Fig. 2b). The system is used for the noise spectral density monitoring.

The devices used to tune the method (as well as for the subsequent experiments) were high- k metal-gate SOI nMOSFETs, fabricated at CEA-LETI in a 22 nm process [5]. The transistors feature an ultrathin body of $t_{\text{Si}} = 7$ nm, Buried-Oxide (BOX) thickness of $t_{\text{BOX}} = 145$ nm, gate length of $L = 100$ nm and gate width of $W = 80$ nm. The hafnium-based gate oxide has an equivalent oxide thickness (EOT) of $t_{\text{EOT}} = 1.3$ nm.

3. Determination of the optimum bias condition for the RTN appearance

The time-domain characterization of the electrical noise requires large time windows due to the extensive quantity of events needed to obtain meaningful statistics [6]. In addition, due to the dependence of the capture time (τ_c : average time at the high current level) and emission time (τ_e : average time at the low current level) of a trap with the carrier concentration of the channel [7], random telegraph noise is not easily observable in all the bias range of the transistor (see Fig. 3a). Therefore, RTN must be characterized at a bias point where the characteristic times, τ_e and τ_c , of the trap differ at most in three orders of magnitude. In this way, the transition events between states can be observed in a reasonable time frame.

This task can be facilitated by the first method that we are introducing in this work: Spectral Scanning by Gate Bias (SSGB). Initially the spectral noise density of the drain current (S_I) is obtained for a given bias (preferentially close to the threshold voltage, Fig. 3b). The corner frequency of the noise spectrum is determined by the sum of the inverse of the characteristic times ($f_c = 1/\tau_e + 1/\tau_c$) [8]. Once the corner frequency is located ($f_c = 3.75$ Hz for the case of Fig. 3b), the spectral density of the current noise, S_I , is measured while V_G is swept leading to the $S_I - V_G$ curve shown in Fig. 3c (SSGB). The bell shaped characteristic of Fig. 3c identifies the bias range where the RTN will be easily observable ($V_G \in [0.4$ V, 0.55 V] for this particular case). The reader can notice that this result is consistent with the drain current signals shown in Fig. 3a where the fluctuations are visible in the same voltage range given by the SSGB plot of Fig. 3c.

4. Identifying the number of traps involved in the RTN signals

Once the bias range for the experiments is unequivocally determined, identifying the RTN signals corresponding to a single-trap is one of the most challenging tasks during its electrical characterization. Examples of the intricate drain current signature of multiple-trap RTN are shown in Fig. 4. The current transients are obtained for different transistors for the same value of the gate bias.

Despite from the time domain representation of Fig. 4 is difficult to discern the contribution of a single-trap, this can be carried out unambiguously by a modified version of the Time Lag Plot method (TLP) partially based on the approach described in [9,10]. Each point of the TLP space (events) is given by the sample of the current at a specific moment, and the immediately next sample ($(I_D(i), I_D(i+1))$). Then, the TLP space event is weighted by the appearance function, $\mathcal{A}(I_D(i), I_D(i+1))$, which accounts for the number of events inside a certain circle of the TLP space and defined by:

$$\mathcal{A}(I_D(i), I_D(i+1)) = \sum_{j=1}^{N-1} \xi(I_D(j), I_D(j+1)) \quad (1)$$

where ξ is a function that eliminates the samples outside the appearance radius,

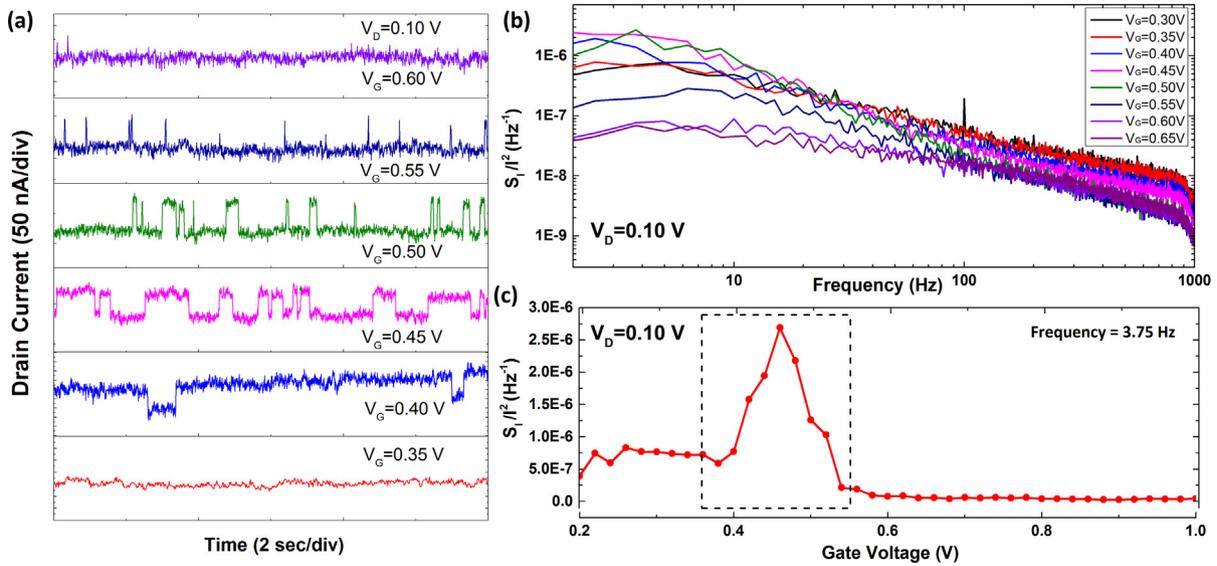


Fig. 3. (a) Drain current traces as a function of the time from a transistor affected by random telegraph noise for different gate voltages. (b) Normalized current noise power (S_i/I^2) dependence with frequency for different gate voltages (V_G). (c) Normalized current noise spectral density dependence with gate voltage (SSGB) for the transistor of figure (b).

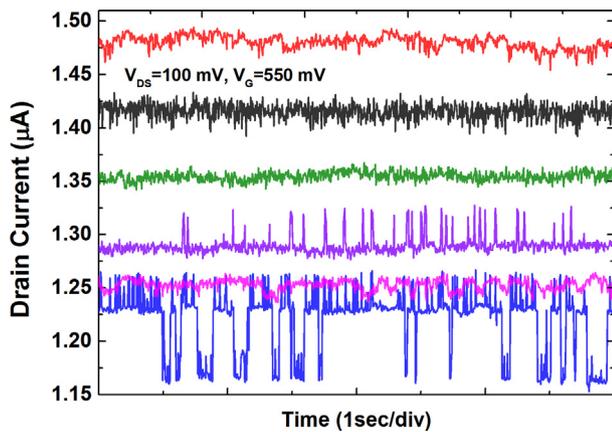


Fig. 4. Drain current traces of different devices presenting multi-trap RTN.

$$\xi(I_D(j), I_D(j+1)) = \begin{cases} 1 & \text{if } d\{[I_D(j), I_D(j+1)], [I_D(i), I_D(i+1)]\} \leq r \\ 0 & \text{if } d\{[I_D(j), I_D(j+1)], [I_D(i), I_D(i+1)]\} > r \end{cases} \quad (2)$$

N is the total number of samples inside the time frame analyzed; $d\{.,.\}$, is the Euclidean distance function; and r is the appearance radius, $r = 10^{-6}\sigma$ (A) (with σ the standard deviation of the samples within the time frame).

This sample-weighted approach of the conventional TLP method allows to identify the RTN levels clearly as populated regions in the diagonal of the TLP space, while populated regions outside the diagonal are related to the transitions between states. We show results of the application of the method in Fig. 5. When the device is not affected by traps (Fig. 5a), a single cloud appears in the TLP space; the appearance of a constellation with two lobes indicates the presence of a single-trap (Fig. 5b); a three-lobes constellation, reveals the existence of three predominant current levels (two traps, Fig. 5c); and when multiple current levels appear, a spread lobe indicates a multi-trap situation (Fig. 5d).

The strength of the method is fully exploited when it is combined with an automatic probe-station. The distribution of traps over the transistors of the wafer can be determined by the auto-

matic monitoring of the drain current of specific transistors and the simultaneous application of the w-TLP method described previously. Fig. 6 shows the map of traps for a 100 nm-length and 80 nm-width transistor fabricated in each die of the wafer (336 transistors in total) when the bias point is close to the threshold voltage ($V_D = 0.1$ V, $V_G = 0.5$ V). The map represents the number of traps detected in each transistor: failed transistors, for which the automatic contact was missed or simply because they were faulty (black); transistors with only one drain current level, i.e. transistor without trap (green¹); transistors with two current levels, i.e. single-trap devices (blue); transistors with three current levels, i.e. two-traps devices (red); and transistors with more than two active traps (pink). Examples of the current level for each type of transistor are shown in the annexed time domain representations of Fig. 6.

Fig. 7 presents the histogram summarizing the trap count over the wafer. From this result, and for the specific transistor studied (one $L = 80$ nm, $W = 100$ nm transistor per die), we can conclude that one out of three of the transistors are affected by one trap, and virtually the same number of transistors are not affected by traps leading to RTN. The case of devices with three current levels, i.e. two traps, represents the 7.5% of the devices characterized on the wafer. Finally, the case of multi-trap devices (transistors with more of three current levels), involves 12.5% of the devices.

The previous result demonstrates the large presence of single-trap devices within this wafer run. This information constitutes an useful feedback for process engineers in the path of the optimization of the technology.

5. Application to substrate bias influence on the RTN

Finally, the whole methodology is applied to study the effect of the substrate bias on the RTN characteristics of SOI transistors: SSGB, for RTN detection, and w-TLP, for the location of single-trap devices. Fig. 8 shows typical τ_c/τ_e ratios used to extract the physical parameters of the traps [11] for two different samples when a substrate bias is applied to the SOI wafer; the range of

¹ For interpretation of color in Fig. 6, the reader is referred to the web version of this article.

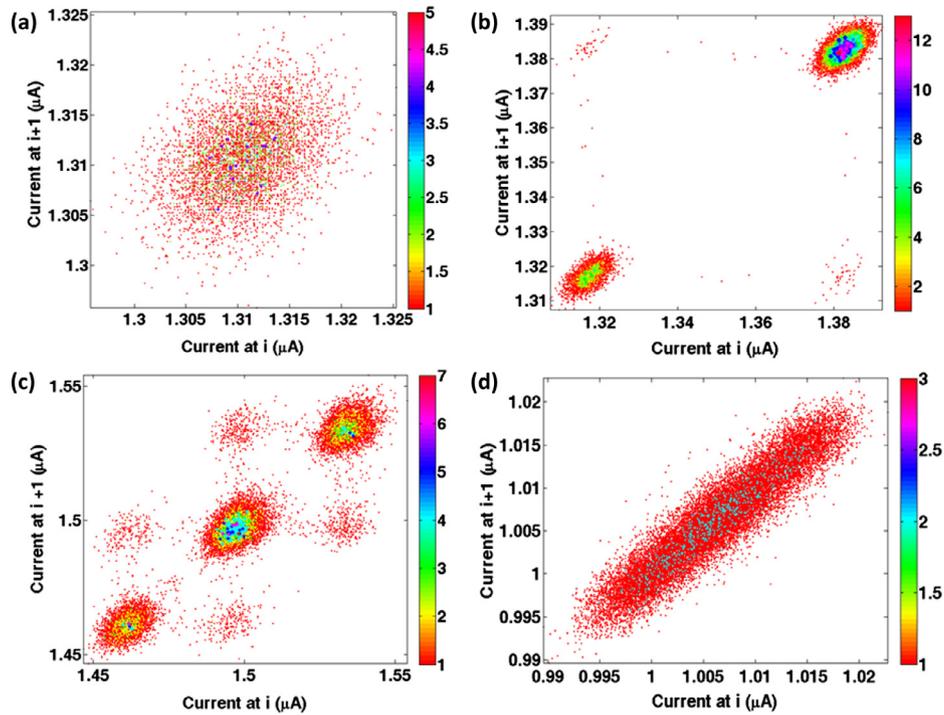


Fig. 5. Traps constellations in the weighted TLP space of drain current signals: (a) cloud: transistor lacking of RTN signature (only thermal noise is reflected). (b) Transistor with two lobes (states) in the TLP constellation result of the single active trap. (c) Transistor with a three-lobes constellation identifying the characteristic signature of two traps. (d) Spread cloud corresponding with a transistor with multi-trap events.

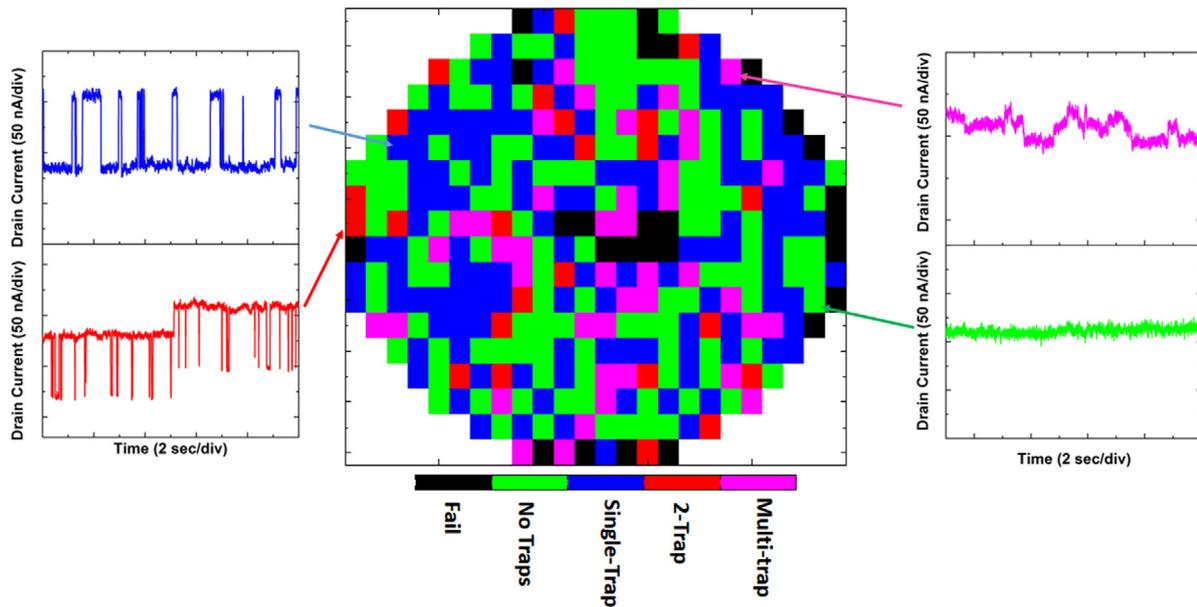


Fig. 6. Distribution of the number of current levels detected in the transistors over the wafer at $V_D = 0.1$ V and $V_G = 0.5$ V. (Left) Examples of transistors with two and three current levels (top and down respectively). (Right) Examples of multi-current level (>3) and RTN-free transistors (top and down respectively).

the front gate overdrive voltage is selected with the SSGB procedure, previously described in Section 3. As observed, although the curves are presented as a function of the overdrive voltage (note that V_T will be modified by V_{SUB} when the back interface is in depletion) [12], the τ_c/τ_e ratio depends on the particular value of V_{SUB} (even though the inversion charge is the same in all cases). Note also that this dependence on V_{SUB} , for a given inversion charge, is different for the particular trap considered (Fig. 8a vs. b).

Fig. 9 shows the values of τ_c and τ_e as a function of the overdrive voltage for the previous devices. From the analysis of this plot we can appreciate that for a given overdrive voltage, the capture time (τ_c) decreases whereas the emission time (τ_e) increases as V_{SUB} increases. From Fig. 9 one may also notice that this behavior is observed both for attractive (linear dependence of τ_c and τ_e with the gate bias, Fig. 9a) or neutral traps (linear dependence of τ_c and constant value τ_e with the gate bias, Fig. 9b) [13]. This effect

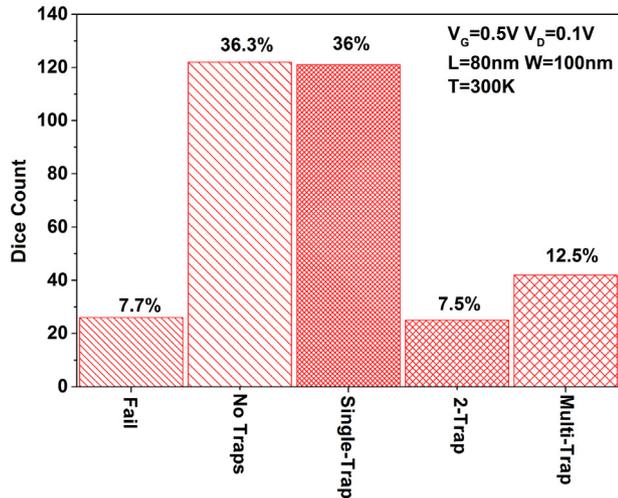


Fig. 7. Histogram of the number of dice presenting a determined number of traps in the device under study detected from the RTN characterization ($L = 100$ nm, $W = 80$ nm, one transistor per die). Same bias than in Fig. 6.

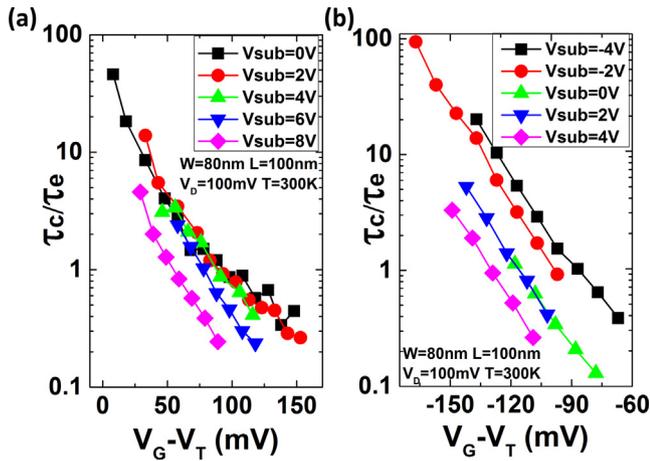


Fig. 8. τ_c/τ_e ratios as a function of the gate overdrive bias for different negative and positive substrate biases. Results from two different devices (a) and (b) and the same physical dimensions.

is contrary from what we may expect considering classical RTN models relying on the inversion charge and the decrease of the electric field in the Si-film while increasing the substrate bias [12]. In contrast, these results are compatible with the fact that the carrier is emitted to the metal gate contact since, for this mechanism, the lower the gate-oxide electric field intensity, the larger the probability of the trap to be filled [14].

6. Conclusion

We have introduced an exhaustive method to identify, experimentally, single-trap Random Telegraph Noise by combining the noise Spectral Scanning by Gate Bias technique (SSGB) with a modified Weighted Time Lag Plot method (w-TLP). The characterization protocol has been implemented in an automatic probe-station and applied for the massive test of the devices all over a wafer providing information about the distribution of traps. Finally, the procedure has been implemented for the study of the impact of the substrate bias on the RTN characteristics of ultrathin SOI transistors. The results show a non-intuitive trend leading to an increase of the emission time (decrease of capture time) when the electric

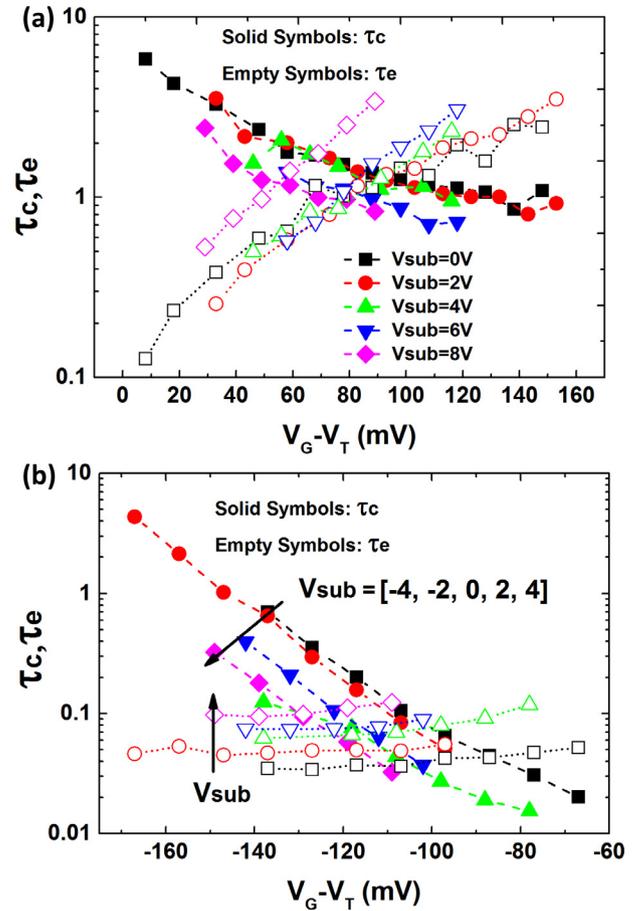


Fig. 9. τ_c and τ_e ratios as a function of the gate overdrive voltage for different substrate biases in the same devices than in Fig. 8.

field in the film is relaxed by increasing the substrate bias, for a given inversion charge. However, this behavior could be explained by the carrier emission to the gate metal contact.

Acknowledgment

The authors would like to thank Dr. O. Faynot and Dr. F. Andrieu from CEA-LETI for sample supplying. This work has been partially funded by Spanish Government through project TEC-2014-59730, regional government through P12-TIC-1996, the BBVA-Spain Foundation for the “Ayudas para Investigadores y Creadores Culturales” program and the European Union under project WAYTOGO FAST.

References

- [1] Tsai M-H, Ma T-P. The impact of device scaling on the current fluctuations in MOSFETs. *IEEE Trans Electron Dev* 1994;41(11):2061–8.
- [2] Fantini P, Ghetti A, Marinoni A, Ghidini G, Visconti A, Marmiroli A. Giant random telegraph signals in nanoscale floating-gate devices. *IEEE Electron Dev Lett* 2007;28(12):1114–6.
- [3] Ohata A, Pretet J, Cristoloveanu S, Zaslavsky A. Correct biasing rules for virtual DG mode operation in SOI-MOSFETs. *IEEE Trans Electron Dev* 2005;52(1):124–5.
- [4] Eminente S, Cristoloveanu S, Clerk R, Ohata A, Ghibaudo G. Ultra-thin fully-depleted SOI MOSFETs: special charge properties and coupling effects. *Solid-State Electron* 2007;51(2):239–44.
- [5] Ranica R, Villaret A, Fenouillet-Beranger C, Malinge P, Mazoyer P, Masson P, et al. A capacitor-less DRAM cell on 75 nm gate length, 16 nm thin fully depleted SOI device for high density embedded memories. In: *IEEE symposium on VLSI technology digest of technical papers*. p. 128–9.
- [6] Uren MJ, Day DJ, Kirton MJ. 1/f and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors. *Appl Phys Lett* 1985;47(11):1195.

- [7] Simoen E, Claeys C. Random telegraph signal: a local probe for single point defect studies in solid-state devices. *Mater Sci Eng: B* 2002;91–92:136–43.
- [8] Shi Z, Mieville JP, Dutoit M. Random telegraph signals in deep submicron n-MOSFET's. *IEEE Trans Electron Dev* 1994;41(7):1161–8.
- [9] Nagumo T, Takeuchi K, Yokogawa S, Imai K, Hayashi Y. New analysis methods for comprehensive understanding of random telegraph noise. In: 2009 IEEE international electron devices meeting (IEDM).
- [10] Martin-Martinez J, Diaz J, Rodriguez R, Nafria M, Aymerich X. New weighted time lag method for the analysis of random telegraph signals. *IEEE Electron Dev Lett* 2014;35(4):479–81.
- [11] Kirton MJ, Uren MJ. Capture and emission kinetics of individual Si:SiO₂ interface states. *Appl Phys Lett* 1986;48(19):1270.
- [12] Lim H-K, Fossum J. Threshold voltage of thin-film Silicon-on-insulator (SOI) MOSFET's. *IEEE Trans Electron Dev* 1983;30(10):1244–51.
- [13] Schulz M. Coulomb energy of traps in semiconductor space-charge regions. *J Appl Phys* 1993;74(4):2649–57.
- [14] Ji X, Liao Y, Zhu C, Chang J, Yan F, Shi Y, et al. The physical mechanisms of I_C random telegraph noise in deeply scaled pMOSFETs. In: 2013 IEEE international reliability physics symposium (IRPS). IEEE; 2013. p. XT.7.1–5.



RF SOI CMOS technology on 1st and 2nd generation trap-rich high resistivity SOI wafers



B. Kazemi Esfeh^{a,*}, S. Makovejev^b, Didier Basso^c, Eric Desbonnets^c, V. Kilchytska^a, D. Flandre^a, J.-P. Raskin^a

^aICTEAM, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

^bIncize, Louvain-la-Neuve, Belgium

^cSoitec, Parc Technologique des Fontaines, 38190 Bermin, France

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

High-resistivity (HR) SOI substrate
Trap-rich high-resistivity silicon
Enhanced signal integrity silicon-on-insulator (eSI HR-SOI)
Substrate effective resistivity
Silicon-on-insulator
DC and RF performance
Partially-depleted (PD) SOI MOSFETs
Crosstalk
Digital substrate noise

ABSTRACT

In this work three different types of UNIBOND™ Silicon-on-Insulator (SOI) wafers including one standard HR-SOI and two types of trap-rich high resistivity HR-SOI substrates named enhanced signal integrity high resistivity silicon-on-insulator (eSI HR-SOI) provided by SOITEC are studied and compared. The DC and RF performances of these wafers are compared by means of passive and active devices such as coplanar waveguide (CPW) lines, crosstalk- and noise injection-structures as well as partially-depleted (PD) SOI MOSFETs. It is demonstrated that by employing enhanced signal integrity high resistivity silicon-on-insulator (eSI HR-SOI) compared to HR-SOI wafer, a reduction of 24 dB is measured on both generations of trap-rich HR-SOI for 2nd harmonics. Furthermore, it is shown that in eSI HR-SOI, digital substrate noise is effectively reduced compared with HR-SOI. Purely capacitive behavior of eSI HR-SOI is demonstrated by crosstalk structure. Reduction of self-heating effect in the trap-rich HR-SOI with thinner BOX is finally studied.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

During last decades, CMOS technology scaling-down has enabled millimeter wavelength operation and low-cost integration of digital, analog and RF systems on the same wafer for system-on-chip or system-in-package applications [1–3]. In this context, the most special advantage of SOI CMOS compared to bulk Si is the availability of high-resistivity silicon (HR-Si) substrate to achieve low crosstalk between passive and active devices and high-quality passive elements thanks to effective reduction of substrate coupling and losses in RF circuits [4,5]. However, HR-SOI substrate suffers from resistivity degradation due to the formation of parasitic surface conduction (PSC) beneath the buried oxide layer (BOX) [6–9] due to fixed oxide charges (Q_{ox}) within the oxide. One of the most efficient techniques to overcome this problem is to introduce a trap-rich layer at the Si/SiO₂ interface compatible with industrial SOI wafer production and thermal budget of standard CMOS process [8]. Such layer aims at capturing the free carriers forming the PSC and thus retaining the substrate nominal high

resistivity. In this work two types of trap-rich HR-SOI wafers named 1st and 2nd generation of enhanced signal integrity (eSI HR-SOI) substrate having respectively a BOX thickness of 400 nm and 200 nm developed by Soitec are studied and compared with the classical HR-SOI substrate with a BOX thickness of 1 μm. One of the motivations of using trap-rich HR-SOI substrates with thinner BOX is the reduction of self-heating effect. Moreover, it creates a pathway for further ultimate BOX thinning used in advanced nano-scaled ultra-thin body and BOX (UTBB) fully depleted MOSFETs which allows threshold voltage control by means of back-gate biasing voltage V_{bg} [10]. Therefore, trap-rich HR-SOI with thinner BOX could be considered as a promising candidate.

2. Device description

In this work two types of trap-rich HR-SOI substrates denoted eSI Gen1 and eSI Gen2 as 1st and 2nd generations with 400 nm and 200 nm-thick BOX respectively and one standard HR-SOI with 1 μm BOX (all provided by Soitec) are characterized and compared for non-linearity effects, crosstalk, noise coupling, DC/RF figures of merit and self-heating. The test structure devices include

* Corresponding author.

E-mail address: Babak.kazemiesfeh@uclouvain.be (B. Kazemi Esfeh).

0.52 μm -thick CPW lines and PD SOI nMOSFETs fabricated using TowerJazz 0.18 μm SOI CMOS process (Fig. 1).

The lateral dimensions of the CPW lines are 20, 18 and 100 μm for the central conductor, slot space and ground plane, respectively. The PD SOI nMOSFETs have 145 nm-thick active silicon film with a nominal operating voltage of 2.5 V. The studied RF body-tied MOSFET has a gate length (L_g) of 0.24 μm with 16 gate fingers of 2 μm each (W_f). The studied single-finger DC nMOSFET has a gate length (L_g) of 0.26 μm with 1.5 μm width (W_f). To investigate substrate coupling we used a passive crosstalk structure consisting of two identical metallic pads embedded into RF pads for probe measurement. Active crosstalk structure is used to study the propagation of a digital noise signal through the substrate [11]. This characterization is performed by measuring the frequency spectra at the nMOSFETs drain as the output port when a square noise signal is injected in the vicinity of the nMOSFET via a metallic RF pad with a certain distance from the transistor.

3. Transistor characteristics

3.1. DC and RF performance

The DC on-wafer measurements have been done using an Agilent B1500. As shown in Fig. 2, the drain current versus gate voltage (I_D - V_G) and transconductance vs gate voltage (g_m - V_G) curves in linear regime are almost identical for the DC transistor on the 3 different wafers. To eliminate the threshold voltage variations effect and fairly compare these results, g_m/I_D ratio versus $I_D/(W/L)$ curves for the same transistors are plotted in Fig. 3.

Fig. 3 shows that similar values of maximum $g_m/I_D \sim 31 \text{ V}^{-1}$ are obtained for all substrate types, as well as similar characteristics in strong inversion, with very slight deviations in weak inversion which could be related to CMOS process and measurements variability. From Figs. 2 and 3 it can be seen that neither the existence of trap-rich layer nor the BOX thickness affects the DC characteristics of the PD SOI MOSFETs.

High-frequency measurement of studied RF body-tied nMOSFET is performed to extract the current gain cutoff frequency (f_T) as one of the major RF figures of merit. f_T is known as unity current gain frequency at which the short circuit current gain (H_{21}) becomes unity (0 dB) [12–14]. RF measurements are done in the frequency range from 40 MHz up to 40 GHz using Anristu 37369A in combination with HP4145 semiconductor-parameter analyzer. By using the off-wafer line-reflect-match (LRM) calibra-

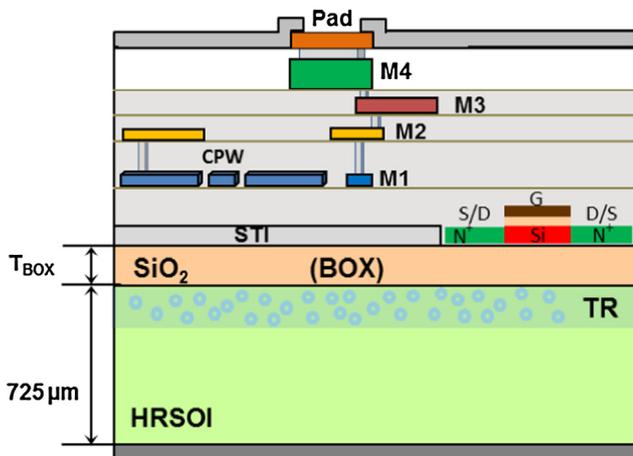


Fig. 1. Cross-section details of 4-metal layer 0.18 μm SOI CMOS process by TowerJazz on top of high resistivity SOI substrates having different BOX thickness (T_{BOX}) provided by SOITEC.

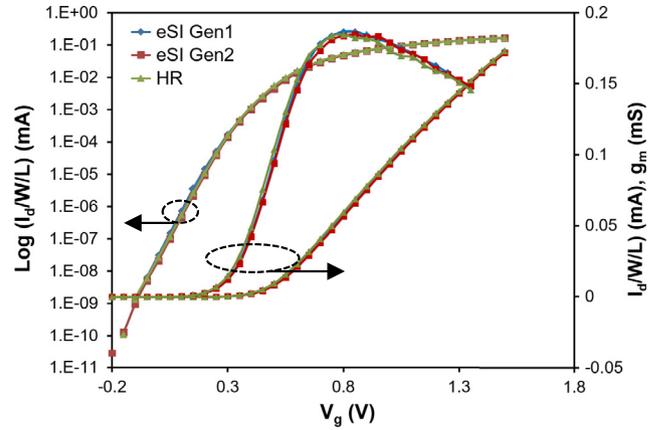


Fig. 2. Normalized I_D - V_G (linear and logarithmic) and g_m - V_G characteristics in linear regime ($V_{DS} = 50 \text{ mV}$) of DC transistor for 3 different wafers of 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

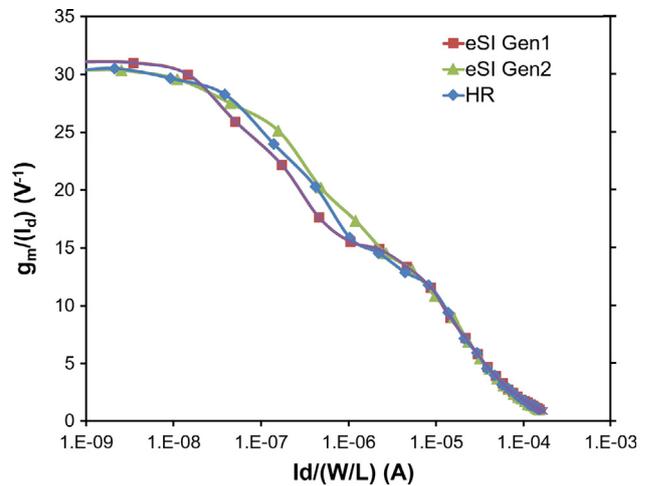


Fig. 3. g_m/I_D ratio versus $I_D/(W/L)$ DC transistor in linear regime ($V_{DS} = 50 \text{ mV}$) on 3 different wafers of 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

tion technique, the reference planes at the probe tips are determined. Then by means of on-wafer dedicated de-embedding structures, the unwanted parasitic effects introduced by the RF pads are removed. H_{21} is extracted from the measured S-parameters of the transistor in saturation regime at V_g bias corresponding to the maximum of g_m [15,16]. As shown in Fig. 4, cut-off frequencies f_T on all 3 wafers are almost the same.

According to the MOSFET small-signal equivalent circuit f_T can be in a first order expressed as [13,14,16]

$$f_T \approx \frac{g_m}{2\pi C_{gg}} \quad (1)$$

where C_{gg} is the total gate capacitance (i.e. $C_{gs} + C_{gd}$). From Eq. (1), Fig. 4 and the results of DC measurements, it can be also concluded that total gate capacitances of PD MOSFETs fabricated on three different wafers are identical.

3.2. Self-heating and coupling through the substrate

Self-heating in SOI devices becomes a critical issue because of device downscaling and use of materials with low thermal conductivity like SiO_2 [17]. Thermal conductivity of SiO_2 ($1.4 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$)

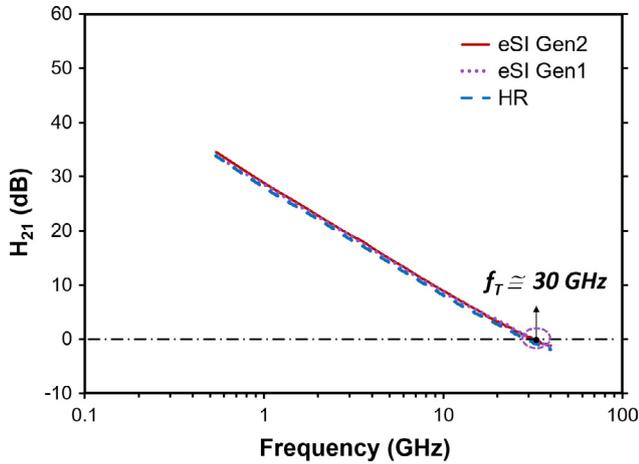


Fig. 4. Current gain cutoff frequency f_T at $V_{DS} = 1.2$ V and V_{GS} at which g_m (transconductance) is maximum for body-tied RF transistor for 3 different wafers of 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

is two orders of magnitude lower than that of Si ($148 \text{ W}\cdot\text{m}^{-1}\cdot\text{K}^{-1}$) [18]. Consequently, the thicker the buried oxide layer (BOX) the higher the thermal resistance (R_{th}) of the device. In [19] it is shown that the thermal resistance is proportional to the square root of the BOX thickness. Therefore, BOX thinning is seen as a good solution for thermal properties improvement. Moreover, thin BOX enables other useful features in some advanced technology such as ultra-thin body ultra-thin BOX (UTBB) devices, notably threshold voltage control from the backgate [20–22] and better control of short channel effects [23].

In this section, the self-heating effect of studied RF body-tied PD nMOSFET is investigated by applying the RF technique [24]. In this technique, S-parameters are measured over a frequency range from 40 kHz up to 3 GHz at room temperature (25°C), de-embedded using dedicated open structures and converted to Y-parameters. The output conductance g_{ds} is obtained from the real part of the Y_{22} parameter. In order to extract device thermal impedance from g_{ds} variation with frequency, hot chuck DC measurements are required. The output characteristics (I_d/V_{ds}) for different gate voltages (V_{gs}) at different temperatures were measured. The thermal resistance R_{th} is proportional to the g_{ds} transition amplitude and inversely proportional to the drain current temperature dependence [19,24]:

$$R_{th} = \frac{\Delta g_{ds}}{(V_{ds}g_{dT} + I_d)\partial I_d/\partial T_A} \quad (2)$$

where Δg_{ds} is the g_{ds} difference at low and high frequency (~ 50 MHz) and g_{dT} is g_{ds} at high frequency. $\partial I_d/\partial T_A$ is the dependence of drain current I_d on the ambient/chuck temperature T_A and can be extracted from I_d/V_{gs} or I_d/V_{ds} measurements at different temperatures. The average temperature rise ΔT is obtained from

$$\Delta T = R_{th}I_dV_{ds} \quad (3)$$

The studied devices are body-tied to prevent the floating-body or kink effect as a common feature in PD SOI devices causing the history effect [25]. Moreover, body-tied devices do not feature frequency variation of output conductance related to the floating-body and thus allow for more straightforward interpretation of output conductance frequency response in terms of self-heating and substrate effect and corresponding parameters extraction.

Fig. 5 shows g_{ds} variation with frequency when the transistor is biased in saturation regime ($V_{gs} = V_{ds} = 1.4$ V). DC output conductance values extracted from the I_dV_d measurements are also pointed. As discussed in [26–31], transitions in g_{ds} versus fre-

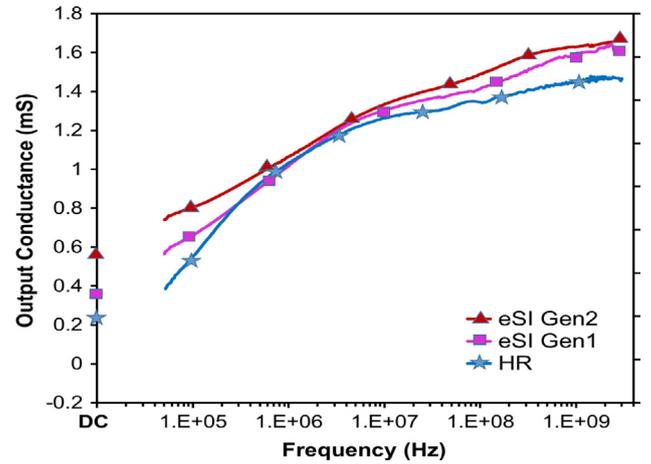


Fig. 5. Output conductance (g_{ds}) variation with frequency at $V_{gs} = V_{ds} = 1.4$ V.

quency curve are caused by various effects. The g_{ds} transition in tens of kHz to hundreds of MHz range is generally considered to be caused by self-heating [19,24,26–28]. As the frequency increases, device temperature stops following electrical oscillations and isothermal condition is reached [26]. As can be seen in Fig. 5, transistors on different substrates exhibit similar g_{ds} values at ~ 50 MHz where dynamic self-heating is removed, whereas at DC and low frequencies, HR substrate shows the lowest output conductance. This lowest output conductance in HR substrate case can be misleadingly interpreted as the better performance of devices on HR substrates. However, just contrarily, lower DC g_{ds} values obtained for the devices on HR substrate is a result of stronger self-heating caused by thicker BOX employed and thus worse performance can be expected.

Values of R_{th} , the average temperature rise ΔT in the device and the magnitude of g_{ds} transitions are plotted in Fig. 6. R_{th} and ΔT values are in line with previously measured values for other PD SOI devices [19,24]. As expected, R_{th} and ΔT are larger in devices with thicker BOX. In Fig. 5, the transition manifested in the GHz range (f_{sub}) is related to the substrate effect, which appears as a result of source and drain coupling through the substrate under the BOX [28]. As discussed in more details in [28,29,31], this

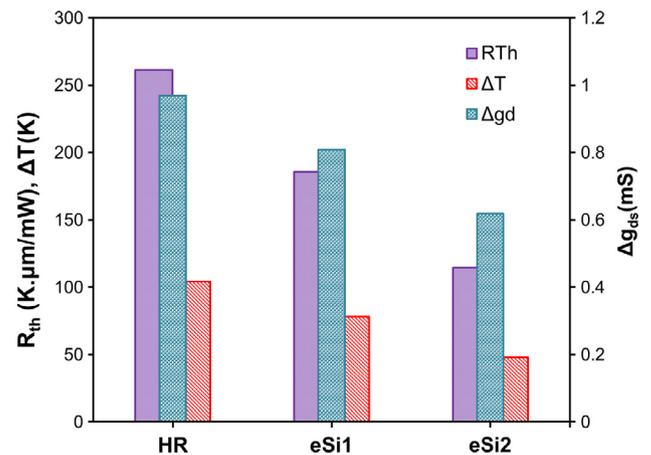


Fig. 6. Average temperature rise (ΔT), thermal resistance (R_{th}) and output conductance transition amplitude due to self-heating (Δg_{ds}) in studied body-tied RF transistors biased at $V_{GS} = V_{DS} = 1.4$ V for 3 different wafers of 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

transition at high frequency can in a first order be modeled by the Si substrate resistance R_{Si} and its capacitance C_{Si} , when it can be considered as dielectric.

$$f_{sub} \sim \frac{1}{(R_{Si} \cdot C_{Si})} \quad (4)$$

It can be seen that the substrate-related transition (GHz range) shifts progressively to lower frequencies when HR substrate is replaced by eSI Gen1 and then by the eSI Gen2. This trend correlates with higher resistivity of eSI Gen1 and Gen2 discussed in next section (see Fig. 7). Indeed, characteristic frequency of this substrate transition was previously shown to be inversely proportional to the Si substrate resistance and capacitance [5,28,29,31] (see Eq. (4)). Furthermore, one can see that amplitude of substrate transition slightly increases in the case of eSI substrates Gen1 and Gen2. This is due to a stronger effect of parasitic source and drain coupling via the substrate in eSI Gen2/Gen1 devices due to the thinner BOX.

4. Substrate effective resistivity and harmonic distortion

The effective resistivity (ρ_{eff}) and total loss (α) on the 3 different types of substrate have been extracted by means of a 2100 μm -long CPW line S-parameters measurements [32]. On-wafer small- and large-signal measurements on CPW line are done based on a

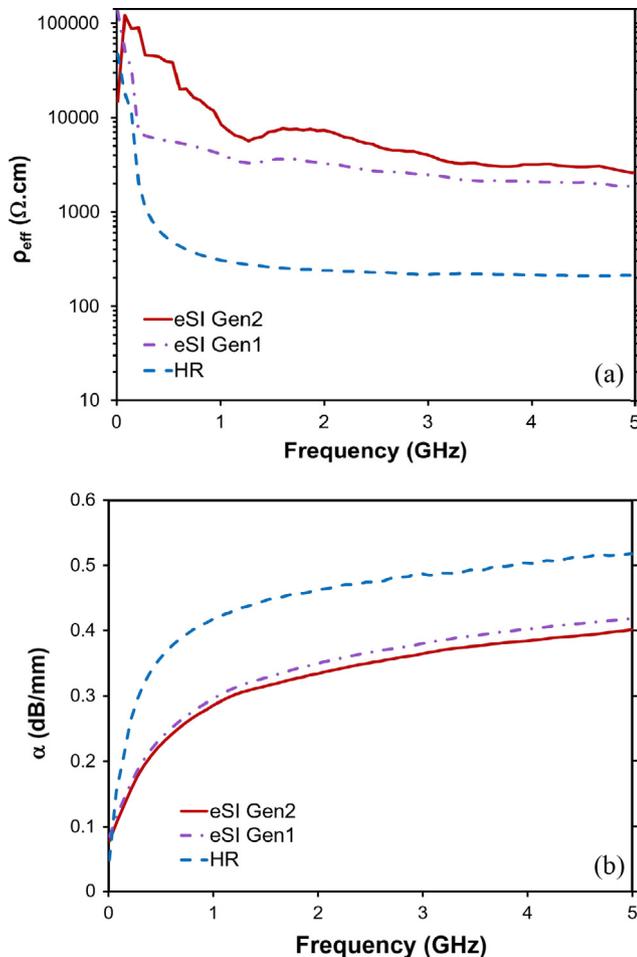


Fig. 7. (a) Effective resistivity, (b) Total Loss (Conductor and substrate) of 3 different substrates of 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

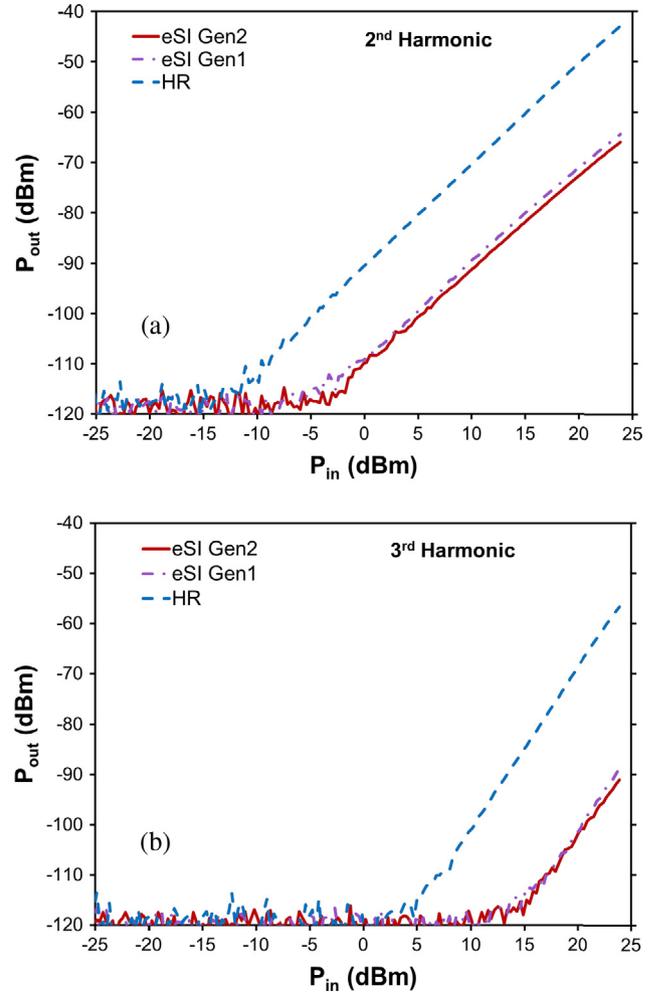


Fig. 8. (a) 2nd and (b) 3rd harmonic distortion measured at zero bias by CPW lines on 3 different wafers: 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

special setup [9] using an Agilent 4-port PNA-X vector network analyzer in the frequency range of 10 MHz up to 26.5 GHz. Fig. 7 shows that as stated before, due to the formation of PSC, the standard HR SOI substrate loses its nominal high resistivity and shows an effective resistivity of only 200 $\Omega \cdot \text{cm}$, whereas in 1st and 2nd generations of trap-rich eSI HR-SOI the substrate has kept its high resistivity of more than 2 $\text{k}\Omega \cdot \text{cm}$ and 3 $\text{k}\Omega \cdot \text{cm}$, respectively, up to 5 GHz after CMOS processing. Another point that can be seen in this figure is that despite its thinner BOX, the eSI Gen2 HR-SOI substrate still shows a slightly higher ρ_{eff} and lower α compared to eSI Gen1 HR-SOI which could be explained by the higher nominal substrate resistivity in 2nd generation.

Fig. 8 illustrates the 2nd and 3rd harmonics distortion at the output of the CPW line on different wafers fed by a 900 MHz input RF signal for a power level ramp of -25 dBm up to 25 dBm and zero bias applied on the substrate. Compared to HR-SOI wafer, a reduction of 24 and 35 dB is measured on both generations of trap-rich eSI HR-SOI for 2nd and 3rd harmonics, respectively. From Figs. 8 and 7 it can be clearly seen that the level of the harmonics is inversely proportional to the substrate resistivity. This non-linear behavior can be explained by the PSC layer which changes the distribution of free carriers inside the substrate generating a modulated charge density at the Si/SiO₂ interface [33]. Consequently, the wafer becomes highly bias voltage dependent [34].

Fig. 9 demonstrates RF performance insensitivity of trap-rich eSI HR-SOI substrates to the applied bias voltages on CPW lines. It can be seen that under different bias conditions, the maximum variation of 2nd and 3rd harmonics distortion in HR-SOI wafer is much higher compared to trap-rich substrates.

5. Crosstalk analysis

For the RF system-on-chip (SoC) applications, less coupling through the substrate between devices fabricated on the same substrate, and therefore a good isolation between them, is desired [6]. In mixed-mode high-frequency integrated circuit in which RF analog circuits are integrated with baseband digital circuitry, the crosstalk coupling through the substrate is an important limiting factor especially for the analog part of the chip which is very sensitive to voltage variations on the power supply and substrate ground rails [5]. The strength of the coupling between different devices depends on device type (active or passive) and the characteristics of the substrate like effective resistivity, effective permittivity, BOX thickness, etc. [35]. Various methods are proposed in literature for reducing the coupling mechanisms between digital and analog parts through the common substrate like metal Faraday cages, guard ring, etc. [5,36–38]. In this work we use enhanced signal integrity high resistivity silicon on insulator substrates for this purpose. Indeed, a trap-rich layer at the SiO₂/Si-substrate will reduce the coupling between devices through the substrate and hence makes better isolation between them [6].

In this section the substrate crosstalk performances of different Silicon-On-Insulator (SOI) technologies including standard HR and two generations of trap-rich high resistivity eSI Gen1 and eSI Gen2 are comparatively investigated and compared. By this comparison the influence of substrate effective resistivity and buried oxide thickness (BOX) are analyzed. The study of crosstalk is demonstrated by two types of test structures: passive crosstalk structures and noise to active devices. The test structures have been designed and fabricated using TowerJazz 0.18 μm SOI CMOS process as stated in Section 2 on 3 different SOI wafers provided by Soitec.

The test structure is composed by two rectangles of metal 1 as the closest metal to the substrate with a length of 150 μm (L) and a width of 50 μm (W) representing the noise injector and the sensitive node as shown in Fig. 10. The two ports are separated by a distance S of 50 μm considered as the nominal structure for which the

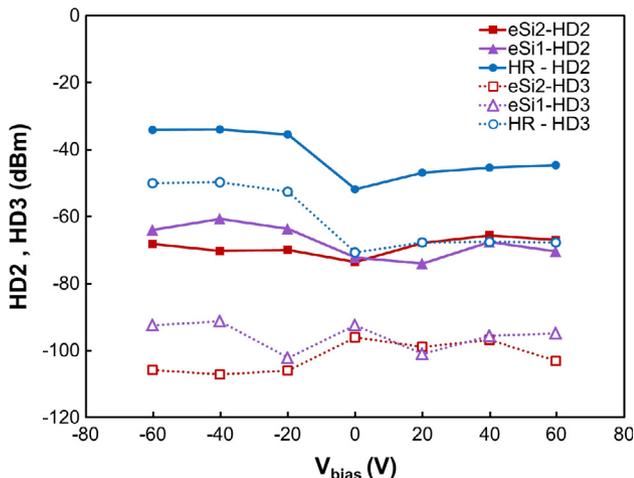


Fig. 9. The variation of 2nd harmonic distortion (HD2) and 3rd harmonic distortion (HD3) with bias changing from -60 V to +60 V at input power P_{in} = 20 dBm for 3 different wafers: 1st generation of trap-rich high resistivity SOI (eSI Gen1), 2nd generation of trap-rich high resistivity SOI (eSI Gen2) and high resistivity SOI (HR).

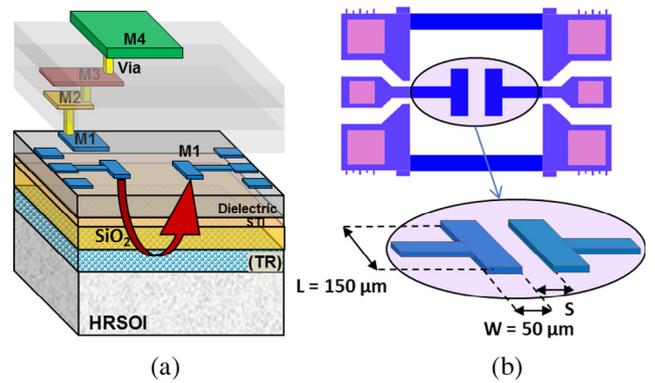


Fig. 10. Crosstalk structure in 0.18 μm TowerJazz CMOS process: (a) schematic; (b) top layout view.

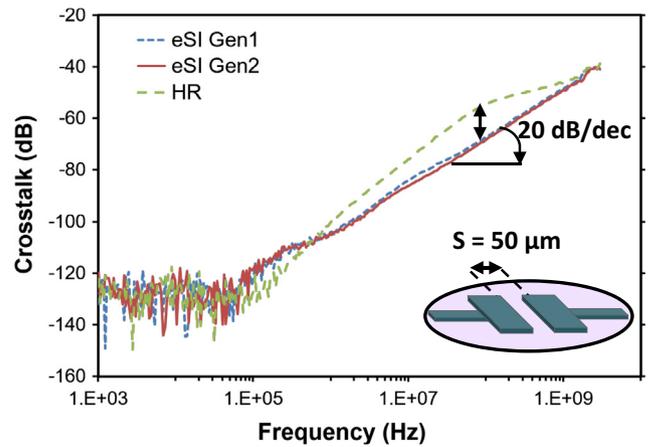


Fig. 11. Measured crosstalk versus frequency on HR-SOI, eSI Gen1 and eSI Gen2 HR-SOI substrates for a crosstalk structure with a pad distance of S = 50 μm and zero substrate bias.

measurements and comparison between the 3 different substrates are shown.

By means of S-parameters measurements (S₂₁) of crosstalk structure shown in Fig. 10(b) with S = 50 μm in the frequency range of 1 kHz up to 3 GHz, the crosstalk level of different wafers is assessed and plotted in Fig. 11. As can be seen, a pure capacitive coupling is observed for both eSI Gen1 and Gen2 with a typical 20-dB/decade slope over the frequency range of 150 MHz-to-3 GHz. This behavior highlights the lossless property of the eSI HR-SOI substrates. Also a reduction of 15 dB of crosstalk is observed at 200 MHz for trap-rich HR-SOI wafers compared with conventional HR-SOI counterpart. From Fig. 11 it can be clearly seen that HR-SOI substrate suffers from higher crosstalk level (slope >20 dB/dec) due to the parasitic surface conduction (PSC) effect which is suppressed in eSI HR-SOI wafers [8,22,39,40].

The large-signal characterization of the crosstalk structure is illustrated in Fig. 12. It shows the 2nd (HD2) and 3rd (HD3) harmonics of 3 different wafers. The 2nd and 3rd harmonics are measured at the output of a crosstalk structure with a pad distance of S = 50 μm fed by an input RF signal at 900 MHz and for a power level ramp from -25 dBm up to 25 dBm. Fig. 12 illustrates that harmonics level of signals coupled through the substrate is reduced by 30 dB when HR-SOI substrate is replaced by an innovative trap-rich HR-SOI (eSI HR-SOI) one. This low harmonics level of -100 dB is comparable with insulating substrates like quartz [41,42].

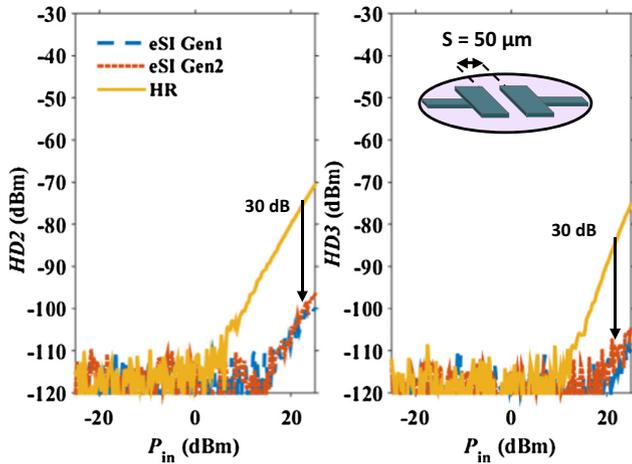


Fig. 12. Measured 2nd (HD2) and 3rd (HD3) harmonic distortions of a crossstalk structure with a pad distance of $S = 50 \mu\text{m}$ and zero DC substrate bias.

Next to that, we study the harmful effect of the digital switching noise emulated by a clock signal, on the performance of an nMOSFET representing the analog/RF part in an RF SoC application [11]. To simulate such an environment, the test structure shown in Fig. 13 is implemented.

The test structure consists of a PD SOI nMOSFET located at a fixed distance of $350 \mu\text{m}$ from an RF metal pad used for injecting a digital noise signal to the structure. The PD SOI nMOSFET has a gate length, gate finger width and number of the finger of $0.24 \mu\text{m}$, $2 \mu\text{m}$ and 64, respectively. It is biased in saturation ($V_{GS} = 1.2 \text{ V}$ and $V_{DS} = 1.2 \text{ V}$). A fundamental input signal at 900 MHz is applied at the gate. A 5 V peak-to-peak digital noise source with selected frequencies of 50 kHz, 100 kHz, and 500 kHz, 1 MHz, 10 MHz, 50 MHz and 80 MHz is connected to the RF noise pad. Fig. 14 shows the coupled noise spectra at the MOSFET drain port for the three different substrates for the 10 MHz clock frequency when gate RF input signal is off. In HR-SOI substrate despite a thicker BOX of $1 \mu\text{m}$ compared to eSI HR-SOI wafers, the output peak noise level is 19.25 dB and

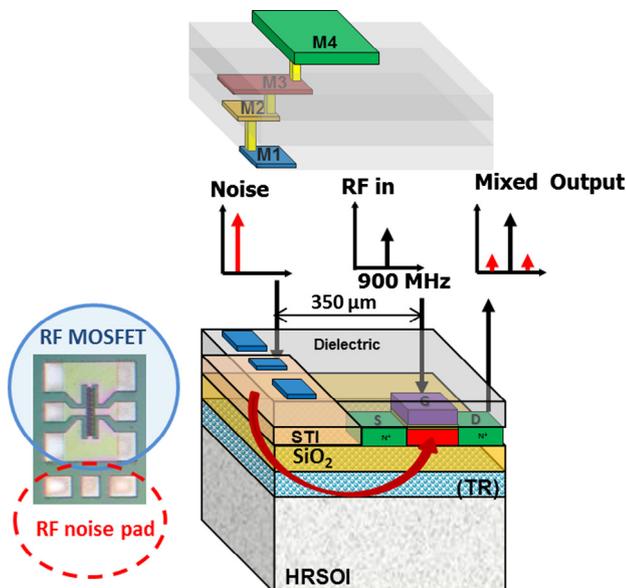


Fig. 13. Cross-section of PD SOI nMOSFET with a noise pad at a distance of $350 \mu\text{m}$ (center to center) lying on either eSI HR-SOI or HR-SOI wafers when a trap-rich layer exists or not respectively.

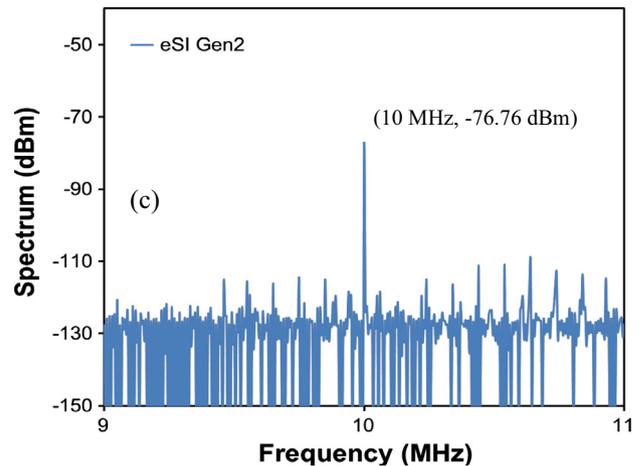
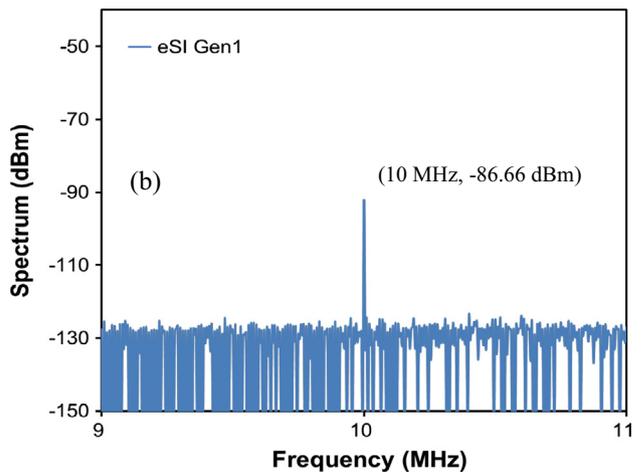
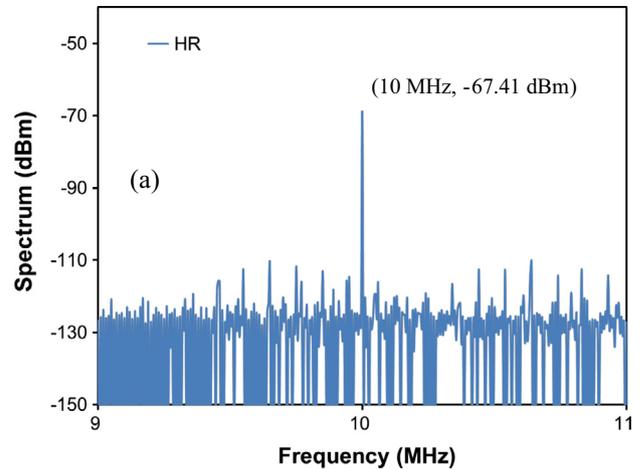


Fig. 14. Frequency spectrum of measured noise signal at the drain port of the PD nMOSFET having a gate length (L_g), gate finger width (W_f) and number of the finger (N_f) of $0.24 \mu\text{m}$, $2 \mu\text{m}$ and 64 respectively biased in saturation with no fundamental RF signal at the gate. A digital noise signal (5 V peak-to-peaks at 10 MHz) is injected to the RF noise pad on (a) HR-SOI, (b) eSI Gen1 HR-SOI and (c) eSI Gen2 HR-SOI substrates.

9.35 dB higher than eSI Gen1 and eSI Gen2, respectively. This is due to the presence of highly parasitic conductive surface layer (PSC) at the BOX/HR-SOI interface by which a strong coupling and propagation of the noise signal is observed at the output. However, comparing Fig. 14(b) and (c), eSI Gen1 shows a noise coupling reduction of 10 dB compared to eSI Gen2 because of

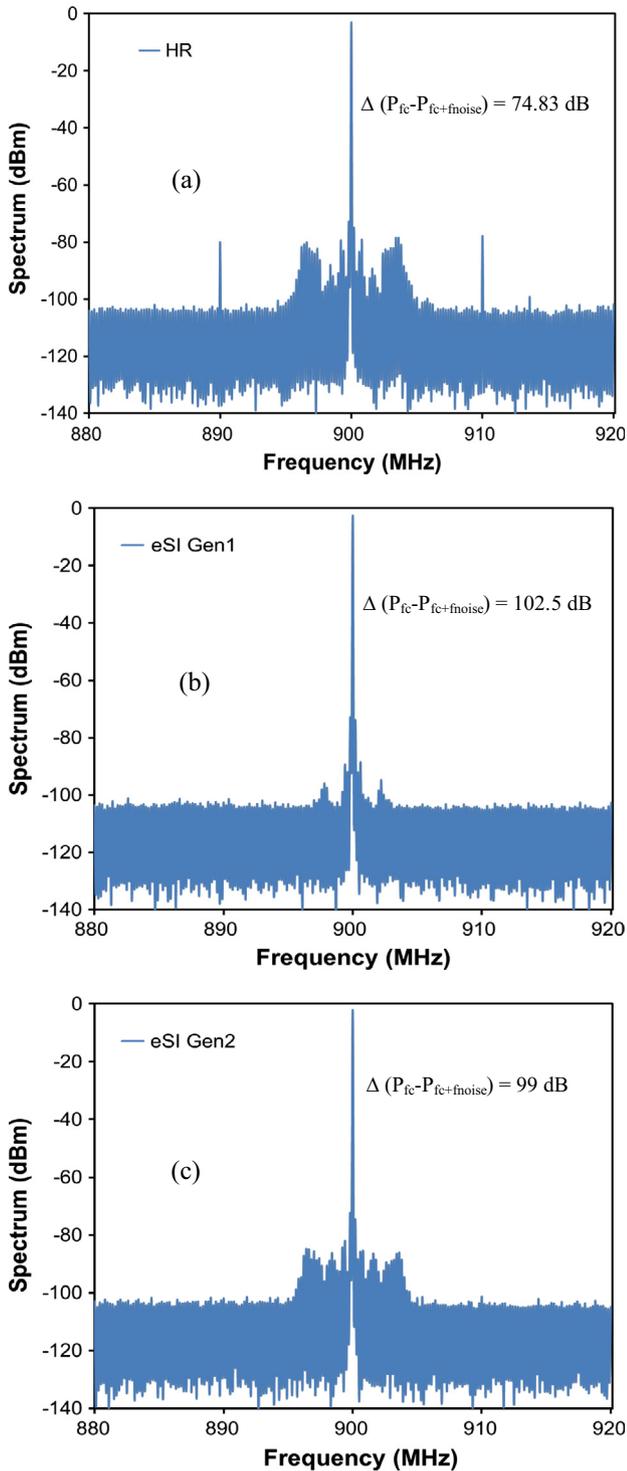


Fig. 15. Output spectrum of measured fundamental and noise signal at the drain port of the PD nMOSFET having a gate length (L_g), gate finger width (W_f) and number of the finger (N_f) of 0.24 μm , 2 μm and 64 respectively biased in saturation with a 900 MHz fundamental RF signal at the gate. A digital noise signal (5 V peak-to-peaks at 10 MHz) is injected to the RF noise pad on (a) HR-SOI, (b) eSI Gen1 HR-SOI and (c) eSI Gen2 HR-SOI substrates.

having a thicker BOX. Therefore, the 1st generation of trap-rich HR-SOI (eSI Gen1) shows a very good performance thanks to a good trade-off between substrate resistivity, from one side and a relatively “thick” BOX from another side.

When a fundamental RF signal with amplitude of A_{Fund} at the frequency of f_c is applied to the gate of the transistor, the digital

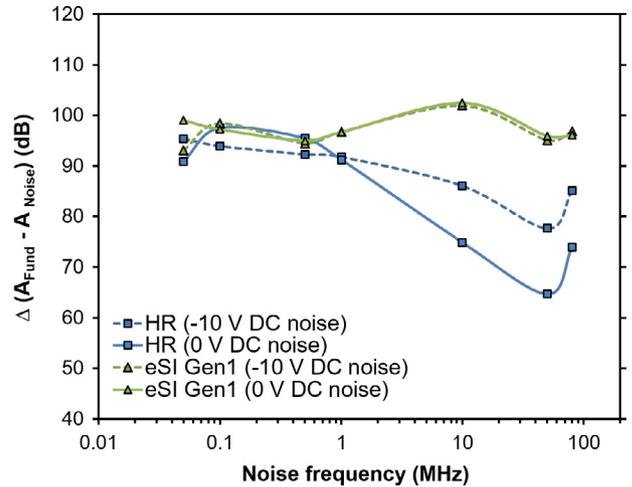


Fig. 16. Power level difference between 900 MHz fundamental RF input (A_{Fund}) and digital noise signal (A_{Noise}) on HR-SOI and eSI Gen1 substrates.

noise signal (A_{Noise} , f_{Noise}) induces a mixing product and generates 2N harmonics at $f_c - f_{\text{Noise}}$ and $f_c + f_{\text{Noise}}$. In this work, as illustrated in Fig. 15, the fundamental 900 MHz input signal and the square digital noise signal of 5 V peak-to-peak at $f_{\text{Noise}} = 10$ MHz result in two harmonics at 890 MHz and 910 MHz. Fig. 15 evaluates the substrate performance through the parameter Δ defined as the power difference between the fundamental output signal and the harmonic ($P_{fc} - P_{fc+f_{\text{Noise}}}$). The higher the Δ the better isolation will be achieved.

Fig. 15 shows an increase of Δ by 27.6 dB for eSI Gen1 and 24.2 dB on eSI Gen2 wafers compared with their HR-SOI substrate counterpart. Fig. 15 evidences that trap-rich eSI HR-SOI substrates filter well the digital noise signal from the output spectrum and show smoother response thanks to eliminating the effect of PSC at the BOX/HR-SOI interface, even though they have thinner BOX than HR-SOI. From Fig. 15(b) and (c), it can be clearly seen that 1st generation of trap-rich HR-SOI (eSI Gen1) shows a better trade-off between substrate resistivity and BOX thickness since most of the noise signals are suppressed. Fig. 16 demonstrates $\Delta = A_{\text{Fund}} - A_{\text{Noise}}$ measured at the drain pad of the transistor as the output when the injected digital noise signal changes from 50 kHz up to 80 MHz on HR-SOI and eSI-Gen1 (as the best substrate in our comparison).

A dc bias voltage of either 0 V or -10 V is also applied to the noise signal pad. By applying a negative dc bias voltage, the negative charges in PSC layer are repelled and a deep depletion will be formed. Therefore, it is expected that under these conditions, the coupled noise decreases and Δ increases. In Fig. 16 it can indeed be seen that application of negative dc bias voltage (-10 V) causes Δ increase by 11 dB at 10 MHz comparing to its level at 0 V bias in the case of HR-SOI substrate, whereas in the case of eSI Gen 1 it stays unchanged thanks to the traps that have frozen the free carriers in PSC. Moreover, from Fig. 16, one can see that in the lower frequency range below 1 MHz, since the carriers have enough time to relax, both HR-SOI and eSI HR-SOI substrates show similar behavior. In the frequency range higher than 1 MHz the effect of the traps is evident leading to almost constant high (w.r.t. HR-SOI) levels of Δ .

6. Conclusion

Use of a trap-rich layer underneath the BOX in HR SOI wafers allows the substrate to recover its high-resistivity properties and thus results in higher effective resistivity, lower losses, lower

crosstalk, lower harmonics level and hence higher linearity, all conserved after CMOS processing. It was shown that the presence of a trap-rich layer does not change the DC and RF characteristics of the MOSFET transistors. Moreover, with enhanced 2nd generation trap-rich eSI HR SOI substrate featuring thinner BOX of 200 nm, improved thermal properties can be achieved. Therefore, this technology is considered as a good candidate for SoC applications.

Acknowledgments

The authors acknowledge SOITEC for providing the wafers and TowerJazz for the fabrication of CPW lines and PD SOI CMOS devices.

References

- [1] International Technology Roadmap for Semiconductor (ITRS), 2006 Edition.
- [2] Buss D et al. SOC CMOS technology for personal internet products. *IEEE Trans Electron Dev* 2003;50(3):546–56.
- [3] Benaissa K et al. RF CMOS on high-resistivity substrates for system-on-chip applications. *IEEE Trans Electron Dev* 2003;50(3):567–76.
- [4] Vanada B. Study of floating body effect in SOI technology. *Int J Mod Eng Res* 2013;3:1817–24.
- [5] Raskin J-P, Viviani A, Flandre D, Colinge J-P. Substrate crosstalk reduction using SOI technology. *IEEE Trans Electron Dev* 1997;44(12):2252–61.
- [6] Ben Ali K, Roda Neve C, Gharsallah A, Raskin J-P. RF performance of SOI CMOS technology on commercial 200-nm enhanced signal integrity high resistivity SOI substrate. *IEEE Trans Electron Dev* 2014;61(3):722–8.
- [7] Lederer D, Raskin J-P. New substrate passivation method dedicated to high resistivity SOI wafer fabrication with increase substrate resistivity. *IEEE Electron Dev Lett* 2005;26:805–7.
- [8] Lederer D, Raskin J-P. RF performance of a commercial SOI technology transferred onto a passivated HR silicon substrate. *IEEE Trans Electron Dev* 2008;55(7):1664–71.
- [9] Neve CR, Raskin J-P. RF harmonic distortion of CPW lines on HR-Si and trap-rich HR-Si substrates. *IEEE Trans Electron Dev* 2012;59(4):924–32.
- [10] Makovejev S, Raskin J-P, Md Arshad MK, Flandre D, Olsen S, Andrieu F, et al. Impact of self-heating and substrate effects on small-signal output conductance in UTBB SOI MOSFETs. *Solid State Electron* 2012;71:93–100.
- [11] Bol D, Ambroise R, Roda Neve CR, Raskin J-P, Flandre D. Wide-band simulation and characterization of digital substrate noise in SOI technology. In: *IEEE international SOI conference*, Indian Wells, CA. p. 133–4.
- [12] Kazemi Esfeh B, Kilchytska V, Barral V, Planes N, Haond M, Flandre D, et al. Assessment of 28 nm UTBB FD-SOI technology platform for RF applications: figures of merit and effect of parasitic elements. *Solid-State Electron* 2016;117:130–7.
- [13] Raskin J-P. Modeling, characterization and optimization of MOSFETs and passive elements for the synthesis of SOI MMICs. PhD thesis. Université catholique de Louvain; 1997.
- [14] Colinge J-P, editor. *FinFETs and other multi-gate transistors*. Springer; 2008.
- [15] Guo JC, Huang CH, Chan KT, Lien WY, Wu CM, Sun YC. 0.13 μm low voltage logic based RF CMOS technology with 115GHz f_T and 80 GHz f_{max} . In: *Microwave conference*, 2003. 33rd European, 4–6 October. p. 683–6.
- [16] Kao HL et al. Limiting factors of RF performance improvement as down-scaling to 65-nm node MOSFETs. In: *Korea-Japan microwave conference (KJMW)*, April 2009 (CGU).
- [17] Makovejev S et al. Self-heating and substrate effects in ultra-thin body ultra-thin BOX devices. In: 2011, 12th international conference on ultimate integration on silicon (ULIS), Cork. p. 1–4.
- [18] Pop E, Sinha S, Goodson KE. Heat generation and transport in nanometer-scale transistors. *Proc IEEE* 2006;94(8):1587–601.
- [19] Jin W, Liu W, Fung SKH, Chan PCH, Hu C. SOI thermal impedance extraction methodology and its significance for circuit simulation. *IEEE Trans Electron Dev* 2001;48(4):730–6.
- [20] Makovejev S, Kazemi Esfeh B, Barral V, Planes N, Haond M, Flandre D, et al. Wide frequency band assessment of 28 nm FDSOI technology platform for analogue and RF applications. In: 2014, 15th international conference on ultimate integration on silicon (ULIS), Stockholm. p. 53–6.
- [21] Noel J-P et al. Multi-VT UTBB FDSOI device architectures for low-power CMOS circuit. *IEEE TED* 2011:2473–82.
- [22] Tsuchiya R, Horiuchi M, Kimura S, Yamaoka M, Kawahara T, Maegawa S, et al. Silicon on thin BOX: a new paradigm of the CMOSFET for low-power high-performance application featuring wide-range back-bias control. In: *Proceedings of the international electron devices meeting*. p. 631–4.
- [23] Liu Q et al. Ultra-thin-body and BOX (UTBB) fully depleted (FD) device integration for 22 nm node and beyond. In: *Symp on VLSI technology*; 2010. p. 61–2.
- [24] Makovejev S, Olsen SH, Kilchytska V, Raskin J-P. Time and frequency domain characterization of transistor self-heating. *IEEE TED* 2013:1844–51.
- [25] Colinge J-P. *Silicon-on-insulator technology: materials to VLSI*. 3rd ed. Kluwer Academic Publishers; 2004.
- [26] Tenbroek BM et al. Self-heating effects in SOI MOSFETs and their measurement by small signal conductance techniques. *IEEE Trans Electron Dev* 1996;43:2240–8.
- [27] Tenbroek BM et al. Impact of self-heating and thermal coupling on analog circuits in SOI CMOS. *IEEE J Solid-State Circ* 1998;33:1037–46.
- [28] Kilchytska V, Levacq D, Lederer D, Raskin JP, Flandre D. Floating effective back-gate effect on the small-signal output conductance of SOI MOSFETs. *IEEE Electron Dev Lett* 2003;24(6):414–6.
- [29] Kilchytska V, Levacq D, Lederer D, Pailioncy G, Raskin JP, Flandre D. Substrate effect on the output conductance frequency response of SOI MOSFETs. In: Hall S, Nazarov AN, Lysenko VS, editors. *Nanoscaled semiconductor-on-insulator structures and devices*. Springer; 2007. p. 221–38.
- [30] Tenbroek BM et al. Identification of thermal and electrical time constants in SOI MOSFETs from small signal measurements. In: *Proc 23rd ESSDERC*. p. 189–92.
- [31] Kilchytska V et al. Frequency variation of the small-signal output conductance of decananometer MOSFETs due to substrate crosstalk. *IEEE Electron Dev Lett* 2007;28(5):419–21.
- [32] Lederer D, Raskin J-P. Effective resistivity of fully-processed SOI substrates. *Solid-State Electron* 2005;49(3):491–6.
- [33] Kerr DC, Gering JM, McKay T, Carroll MS, Neve CR, Raskin J-P. The effect of a SiO_2 interface on RF harmonic distortion in CPW lines on silicon or passivated silicon. In: *Proc 8th topical meeting silicon monolithic integr circuits RF syst*, Orlando, FL, USA, January. p. 151–4.
- [34] Lederer D, Raskin J-P. Bias effects on RF passive structures in HR Si substrates. In: *Proc IEEE 6th topical meeting silicon monolithic integr circuits RF syst*, San Diego, CA, USA, January. p. 334–7.
- [35] Afzali-Kusha A, Nagata M, Verghese NK, Allstot DJ. Substrate noise coupling in SoC design: modeling, avoidance and validation. *Proc IEEE* 2012;94(12):2109–38.
- [36] Stefanou S, Hamel JS, Baine P, Armstrong BM, Gamble HS, Kraft M, et al. Ultralow silicon substrate noise crosstalk using metal faraday cages in an SOI technology. *IEEE Trans Electron Dev* 2004;51(3):486–91.
- [37] Juardar K. A simple approach to modeling crosstalk in integrated circuits. *IEEE J Solid-State Circ* 1994;29(10):1212–9.
- [38] Blalack T, Leclercq Y, Yue CP. On-chip RF isolation techniques. In: *Proc Bip/BiCMOS circuits technol meeting*, Monterey, CA, USA; 2002. p. 205–11.
- [39] Wu Y, Gamble HS, Armstrong BM, Fusco VF, Stewart JAC. SiO_2 interface layer effects on microwave loss of high-resistivity CPW lines. *IEEE Microw Guided Wave Lett* 1999;9(1):10–2.
- [40] Schollhorn C, Zhao W, Morschbach M, Kasper E. Attenuation mechanisms of aluminum millimeter-wave coplanar waveguides on silicon. *IEEE Trans Electron Dev* 2003;50(3):740–6.
- [41] Raskin J-P, Desbonnets E. High resistivity SOI wafer for mainstream RF system-on-chip. In: 2015 IEEE 15th topical meeting on silicon monolithic integrated circuits in RF systems (SIRF), San Diego, CA; 2015. p. 33–6.
- [42] Raskin J-P. High resistivity SOI wafer for mainstream RF system-on-chip. In: 2014 12th IEEE international conference on solid-state and integrated circuit technology (ICSICT), Guilin; 2014. p. 1–4.



Numerical investigation of plasma effects in silicon MOSFETs for THz-wave detection [☆]



C. Jungemann ^{a,*}, T. Linn ^a, K. Bittner ^b, H.-G. Brachtendorf ^b

^a Institute of Electromagnetic Theory, RWTH Aachen University, 52056 Aachen, Germany

^b Embedded Systems Research Group, University of Applied Sciences of Upper Austria, Austria

ARTICLE INFO

Article history:

Available online 20 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Drift-diffusion model
Resistive mixer
Plasma waves
THz detectors

ABSTRACT

Conventional silicon MOSFETs are used for THz detectors in order to facilitate fabrication of cost-efficient circuits with high integration density. Resistive mixers based on NMOSFETs are investigated by drift-diffusion simulations, which include the time derivative of the current densities usually neglected in TCAD codes. Different time-integration schemes are investigated for transient simulations and the modified backward differentiation formula is found to be the most CPU-efficient method for the periodic steady-state. By comparison with the Boltzmann transport equation it is shown that the drift-diffusion model can capture the salient aspects of transport in the THz range. The features of the device simulator are demonstrated by investigation of the current and voltage responsivity together with the noise-equivalent-power for a resistive mixer based on a quarter-micron NMOSFET.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Electromagnetic waves with a frequency in the range from 0.1 to 10 THz are referred to as THz waves [2]. Their strong absorption in air due to water vapor limits them to near-range applications. They are used in many fields (e.g. security, spectroscopy, medicine, biology, chemistry, communications, etc. [3]), but they are difficult to generate, because their frequencies are too high for efficient operation of solid-state devices and too low for optical generators (e.g. quantum cascade lasers) [4]. This is called the THz gap. In order to keep the price of the sources down, generators fabricated in standard semiconductor technologies are preferred. A good example is the versatile source with a single SiGe BiCMOS chip, which produces an output power of 1 mW at 0.525 THz [5]. The SiGe HBTs are operated below their maximum frequency of oscillation at 175 GHz and the periodic signal of the Colpitts oscillator is designed to be square-wave like, where the third harmonic at 525 GHz is extracted. Since this frequency is beyond their maximum frequency of oscillation, the HBTs show no power gain and the power efficiency is low. This has led to the proposal of a new kind of solid-state devices based on plasma waves in quasi-2D electron gases, where the THz signal is generated directly within

the device [6,7]. In analogy to an organ pipe the constant flow of electrons in these devices results in oscillations of the electron density and plasma waves are excited [6]. Yet no devices based on this effect have been demonstrated that work efficiently at room temperature. The best performance is obtained with III-V RF devices [2], albeit at very high cost. The above mentioned silicon-based RF devices are currently the best option for mass market applications. The same holds true for detectors. Cost-efficient detectors for THz waves, which can be integrated in a silicon technology, have been developed (e.g. [8]). Recently, a fully integrated 1 k-pixel CMOS camera for THz waves was demonstrated [9], where the detector is based on a resistive mixer, which contains silicon MOSFETs. In this work we restrict our investigation to silicon MOSFETs used in detector circuits for THz signals.

Typically, device operation is analyzed based on a 1D approach, where the Euler equation in various forms is solved in the small-signal regime together with the continuity equation under the gradual channel approximation leading to formulas similar to the Telegrapher's equations [7,10]. The Telegrapher's equations are solved under the assumption of piecewise homogeneity in the lateral direction and for every homogeneous part of the device a 2-port description similar to a transmission line is obtained. These 2-ports are then connected in series and a model for the total device results. These transmission line models fail to capture important effects due to the at least 2D nature of the devices. For example, important parasitic effects at the interfaces between the homogeneous device parts are neglected. Another drawback

[☆] This article is an extended version of a paper presented at EUROSOL-ULIS 2016 [1].

* Corresponding author.

E-mail address: cj@ithe.rwth-aachen.de (C. Jungemann).

of this approach is that it fails to account for the inhomogeneity of the electron channel in the FET structure. Furthermore, realistic modeling of the contacts is of the utmost importance for plasma waves, because the boundary conditions have a very strong impact on device performance [11]. In a FET the channel is bounded by the source/drain regions, which are inherently 2D. For accurate prediction of the device performance an approach beyond the transmission line models is therefore required. A first step in this direction was published in Ref. [12], where a 2D Poisson equation was solved with a 1D transport model in conjunction with still ideal boundary conditions. A fully 2D approach based on the Monte-Carlo method was presented in Ref. [13], but its CPU times are prohibitive for this type of investigation [14].

In this work we will present a deterministic approach for the simulation of silicon devices based on a CPU-efficient 2D drift-diffusion (DD) model, where the time-derivative of the current density is included in the constitutive equation for the current density [1]. In the next section the model and its numerical implementation are discussed, in the third the resistive mixer and in the fourth the model is applied to the simulation of THz signals in MOSFETs.

2. Numerical approach

The simulation model comprises the DD model [15]

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \vec{J}_n \quad (1)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \nabla \vec{J}_p \quad (2)$$

$$\frac{\partial \vec{J}_n}{\partial t} = -\frac{\vec{J}_n}{\tau_n} - \frac{q^2}{m_n} (n \nabla \varphi - V_T \nabla n) \quad (3)$$

$$\frac{\partial \vec{J}_p}{\partial t} = -\frac{\vec{J}_p}{\tau_p} - \frac{q^2}{m_p} (p \nabla \varphi + V_T \nabla p) \quad (4)$$

and the Poisson equation for the quasi-static potential

$$\nabla \cdot (\varepsilon \nabla \varphi) = -q(p - n + N_D - N_A), \quad (5)$$

where n is the electron density, q the positive electron charge, \vec{J}_n the electron current density, τ_n the macroscopic relaxation time of the electron momentum, m_n the electron conductivity mass ($m_n = 0.286m_0$ and $m_p = 0.372m_0$ for silicon), φ the quasi-static potential, V_T the thermal voltage, ε the permittivity, and N_D the donor concentration. The electron mobility is given by $\mu_n = q\tau_n/m_n$. Corresponding hole quantities are labeled by p and N_A is the acceptor concentration.

Eqs. (3) and (4) contain the time derivatives of the current densities, which are necessary to simulate plasma waves. In TCAD implementations of the DD model these derivatives are usually neglected, because the plasma waves cause numerical problems and increase the CPU time of transient simulations very much due to the required time steps in the order of femtoseconds. In silicon the macroscopic relaxation times for electrons and holes are shorter than a picosecond and the derivatives can be safely neglected for frequencies below 100 GHz [16,17], but not necessarily above.

Eqs. (1)–(5) are discretized on a 2D tensor-product grid with the finite-volume method and dimensional splitting (Fig. 1) [15,18]. The scalar quantities (n, p, φ) are evaluated on the nodes of the direct grid and the Cartesian components of the current densities ($J_{n,x}, J_{p,x}, J_{n,y}, J_{p,y}$) on the nodes of the adjoint grid. Due to the dimensional splitting the nodes of the adjoint grid are placed on the edges of the grid primitives between two nodes of the direct grid. Since the time derivatives are not neglected, the current den-

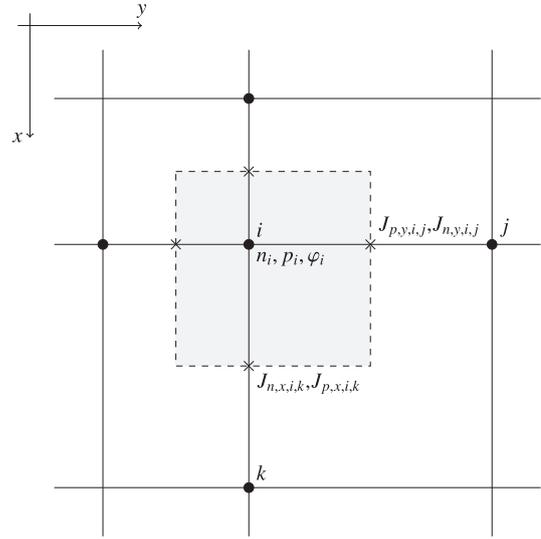


Fig. 1. 2D tensor-product grid. The box volume is shown in grey. The scalar quantities are evaluated on the direct grid nodes (solid circles) and the vector components on the adjoint points (crosses).

sities can no longer be eliminated from the set of equations and the number of unknown variables increases from three to seven per grid node. The second term on the right-hand side of (3) and (4) is stabilized by the Scharfetter-Gummel scheme [19], such that in the slowly varying case ($|\partial \vec{J} / \partial t| \ll |\vec{J} / \tau|$) the usual formulation of the DD model is obtained. The terminal currents are evaluated by the extended Ramo-Shockley theorem [20].

Stationary solutions of the DD model are obtained with the usual methods (Gummel loop [21] and Newton-Raphson method [15]). For the sinusoidal steady-state a small-signal analysis is implemented and for the periodic steady-state (large-signal) the single-tone harmonic balance (HB) method [22]. In addition, the DD model can be solved in the time domain by a transient approach. In this case Eqs. (1)–(4) are of the form

$$\frac{\partial u}{\partial t} = f(u, t). \quad (6)$$

Three different methods for time integration are used. The implicit (backward) Euler scheme (BE), the second-order backward differentiation formula based on polynomials (BDF2) [23,24] and the modified BDF2 scheme based on trigonometric functions (MBDF2) [25]. In all three cases the time derivative is approximated for a constant time step Δt by

$$\frac{\partial u}{\partial t} \Big|_{i\Delta t} \approx \frac{a_0 u_i + a_1 u_{i-1} + a_2 u_{i-2}}{\Delta t} = f_i \quad (7)$$

with $u(i\Delta t) \approx u_i$ and $f(u(i\Delta t), i\Delta t) \approx f_i$. The coefficients are for the BE scheme

$$a_0 = 1, \quad a_1 = -1, \quad a_2 = 0, \quad (8)$$

BDF2 scheme

$$a_0 = 1.5, \quad a_1 = -2, \quad a_2 = 0.5, \quad (9)$$

and MBDF2 scheme

$$a_0 = \frac{z \cos(2z) - z \cos(z)}{\sin(2z) - 2 \sin(z)}, \quad (10)$$

$$a_1 = \frac{z \sin(z)}{\cos(z) - 1}, \quad (11)$$

$$a_2 = \frac{z}{2 \sin(z)} \quad (12)$$

with $z = 2\pi/n_T$. The MBDF2 scheme is used only for the periodic steady-state and n_T is the number of time steps used per period, where n_T must be larger or equal to eight. All three formulas are A-stable, but show quite different spurious damping behavior in the case of plasma waves.

The dopant-dependent low-field mobility in silicon is modeled according to Caughey and Thomas [26] and the channel mobility in MOSFETs according to Lombardi et al. [27]. In all cases the resultant linear system of equations is solved by the direct sparse matrix solver PARDISO [28].

It is worth noting that without the time derivatives in (3) and (4) the particle densities will always be positive in the stationary and transient case for the above mentioned discrete system. As soon as the time derivatives of the current densities are included, second-order time derivatives of the particle densities occur implicitly and positive particle densities are no longer ensured by construction of the discrete system. If negative particle densities are encountered, the simulation is terminated. At small AC amplitudes this problem does usually not occur.

3. Mixer model

An NMOSFET is used as a passive mixer in Ref. [8] and we use an analogous circuit (Fig. 2). Due to the large externally added capacitance $C_{GD,ext}$ gate and drain are effectively short-circuited at high frequencies [8]. To capture this effect, the AC signal is applied to the source terminal, and the gate and drain are AC-wise grounded [1]. The applied DC drain/source bias is zero. The MOSFET is operated as a diode and it rectifies the AC signal, which leads to the DC drain current. The responsivity of the mixer is given by the ratio of the DC drain current and AC input power at the source

$$R_I = \frac{I_D^{DC}}{P_S^{AC}}, \quad (13)$$

where the AC input power is given by

$$P_S^{AC} = \frac{1}{T} \int_0^T V_S(t) I_S(t) dt. \quad (14)$$

The integral runs over one period with the duration T of the periodic input signal, and the input power is proportional to the square of the AC voltage amplitude V_{AC} similar to the DC drain current. The responsivity is thus constant over a wide range of AC input power. Only for very large AC voltages the responsivity depends on the AC power.

A similar responsivity can be defined for the drain voltage, if the drain is kept DC-wise open (i.e. by a very large resistance in series with the drain inductance in Fig. 2). If the resultant DC drain volt-

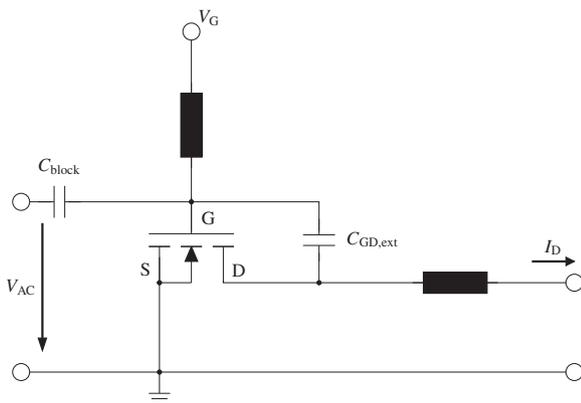


Fig. 2. Resistive mixer for AC power detection [8].

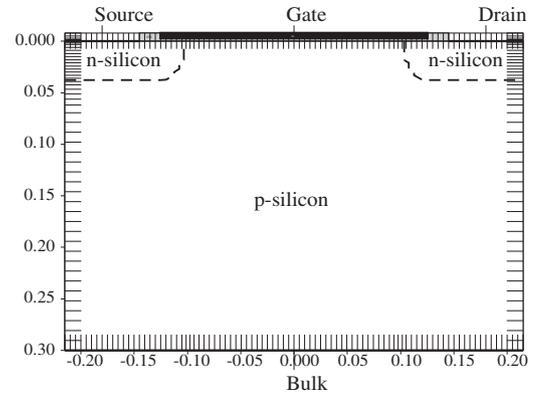


Fig. 3. 2D structure of the silicon NMOSFET. The ticks indicate the grid lines and the dashed thick lines are the metallurgical junctions. Numerical values are in μm .

age is small, the device response is linear and at zero frequency and zero drain/source bias the channel conductance is given by the drain self-admittance ($G_{DS} = 1/R_{DS} = \Re\{Y_{DD}\}$). The corresponding channel current, which compensates the DC drain current due to rectification, induces a DC drain voltage of $V_D^{DC} = R_{DS} I_D^{DC}$. The voltage responsivity is thus given by [8]

$$R_V = \frac{V_D^{DC}}{P_S^{AC}} = R_{DS} R_I. \quad (15)$$

Simulation of the two different mixer configurations confirms this relation for the AC voltage amplitudes used in this study. Since the device structures are 2D, the voltage responsivity has to be divided by the width of the device in the third dimension (the input power is proportional to the device width). In the case of the current responsivity the DC current and input power are both dependent on the width and it cancels in the responsivity.

In addition to the responsivity of a detector its noise-equivalent-power (NEP) is of importance and it should be as small as possible. It is defined as the input power, for which the signal-to-noise ratio at the output is one for a bandwidth of 1 Hz [8]

$$NEP_I = \frac{\sqrt{4k_B T_0 G_{DS}}}{R_I} = \frac{\sqrt{4k_B T_0 R_{DS}}}{R_V} = NEP_V \quad (16)$$

with the ambient temperature T_0 . At zero frequency and zero drain/source bias the NMOSFET can be replaced by the channel conductance G_{DS} , and the drain noise is given by the Johnson-Nyquist formula. In this calculation down-conversion of noise and $1/f$ -noise are neglected. Since our simulations are performed for 2D device structures, the NEP has to be multiplied with the square root of the width of the device in the third dimension.

4. Results

In Ref. [8] quarter-micron silicon NMOSFETs are used to detect THz radiation and we use therefore a similar device (Fig. 3). The grid for real space, which is indicated by tick marks, is non-equidistant to better resolve the junctions and the channel area. The length of the gate is 250 nm, the oxide thickness 3 nm and the threshold voltage 0.56 V. The 2D structure captures the most important parasitics (e.g. the overlap capacitances between gate and source/drain). The artificial boundary conditions often used in simulations for the channel (Dirichlet and inhomogeneous von-Neumann type, e.g. [12]) are avoided by this approach and the channel is realistically terminated by the highly-doped source/drain regions, into which plasma waves can enter from the channel. In addition, the channel is inhomogeneous, a fact

not easily included in 1D transmission line models. The interface between the oxide and silicon is assumed to contain no traps and corresponding effects (e.g. $1/f$ noise) are neglected.

The Boltzmann Transport Equation (BTE) yields the most fundamental description of semiclassical transport in semiconductor devices [29]. The more CPU-efficient DD model can be derived from the BTE by projection on certain moments, but this process involves approximations of unknown quality [30]. It is not clear how accurately the DD model can describe transport at frequencies, where plasma effects occur. In Fig. 4 the real part of the drain/gate admittance is shown as a function of the frequency for the NMOSFET. This small-signal result is calculated by the DD model with and without the time derivative of the current densities and by a spherical harmonics expansion of the BTE [31]. In this case the electron and hole mobilities are given by their respective bulk values [26] neglecting interface effects to facilitate a simpler comparison of the two models. Below 150 GHz the real part of the admittance is positive and all three models yield similar results. Above 1 THz differences occur. The impact of the plasma oscillations is manifested in the non-monotonic behavior of the results of the BTE and the DD model including the time derivative of the current densities. The DD model and the BTE show reasonable agreement even at very high frequencies. Thus, the DD model with the time derivative of the current densities is able to capture the most salient aspects of the plasma oscillations.

The responsivity is an inherently nonlinear effect and requires a large-signal simulation, which can be performed in the time or frequency domain for the periodic steady-state. The transient simulations in the time domain are based on the time-integration schemes presented in Section 2 and a sufficient number of periods is used to reach the periodic steady-state. The responsivity is calculated by integrating the drain current and the input power at the source over the last simulated period. In Fig. 5 the current responsivity of the NMOSFET is shown for the different time integration schemes as a function of the number of time steps per period for a very large electron mobility in order to emphasize the plasma effects. While the dependence of the MBDF2 result on the number of time steps is negligible, the other two schemes show a strong dependence and the responsivity is underestimated. This is due to the strong over-damping of these schemes, which decreases with increasing number of time steps [25]. For the periodic steady-state the MBDF2 scheme is the by far most efficient one and its results are in excellent agreement with the HB approach. This is due to the fact that the MBDF2 scheme is based on trigonometric functions similar to the HB method. Depending on the

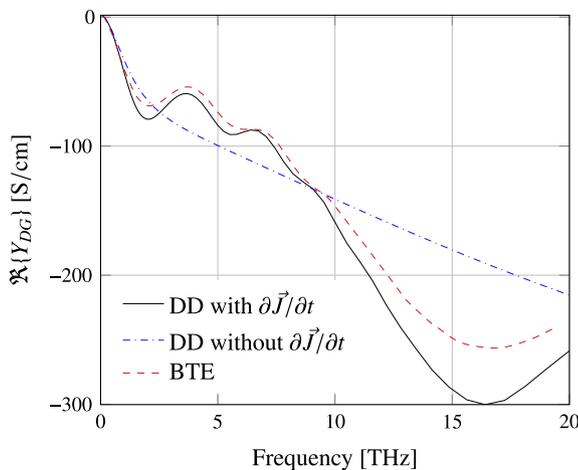


Fig. 4. Real part of the small-signal drain/gate admittance at $V_G = 2.0$ V, $V_D = 0.1$ V and room temperature for the NMOSFET.

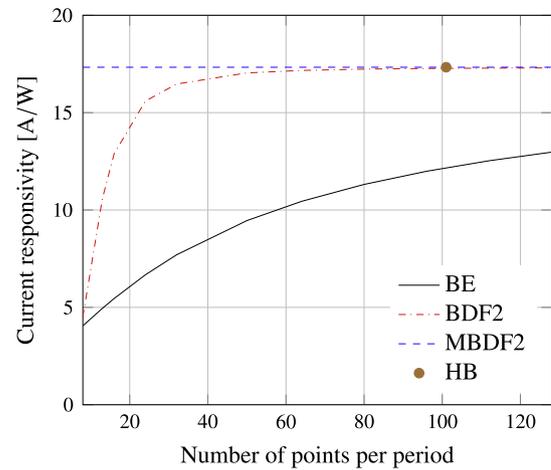


Fig. 5. Current responsivity at $V_G = 1.0$ V, $V_D = 0.0$ V, $V_{AC} = 1$ mV and room temperature for a $0.25 \mu\text{m}$ NMOSFET with a constant electron mobility $\mu_n = 10,000 \text{ cm}^2/\text{Vs}$ for 1 THz evaluated by transient simulations (BE, BDF2, MBDF2) and harmonic-balance calculations (HB).

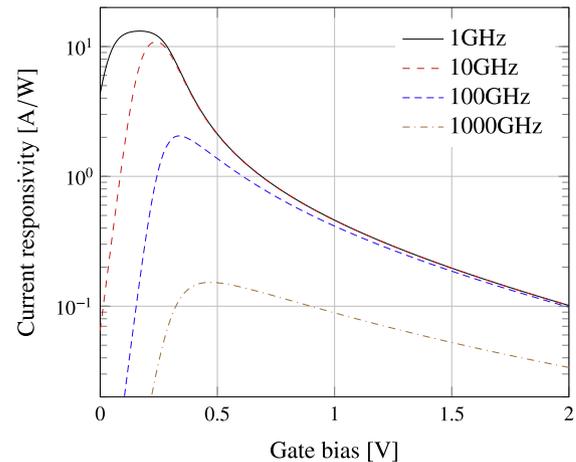


Fig. 6. Current responsivity at $V_D = 0.0$ V, $V_{AC} = 0.1$ mV and room temperature for the NMOSFET as a function of the gate bias for selected frequencies.

problem, either HB or MBDF2 simulations are more CPU efficient, where the MBDF2 simulations are often numerically more robust. This result also shows that the MBDF2 scheme should be superior to the BDF2 scheme in the case of active devices, where usually transient simulations with the BDF2 scheme are used (e.g. [12]) and the over-damping could suppress or delay the onset of active device behavior.

The current responsivity is shown in Fig. 6 for the NMOSFET for different frequencies as a function of the DC gate bias. For a given frequency the responsivity has a maximum in the sub-threshold region, where the channel resistance is very large and the relative resistive mixing strongest. The peak responsivity rapidly drops with increasing frequency and moves to higher gate voltages. In Fig. 7 the current responsivity is shown as a function of the frequency up to 10 THz for zero gate overdrive with and without the time derivatives of the current densities. In both cases it strongly drops with frequency and the differences are rather small due to the strong over-damping of the plasma waves by the low electron momentum relaxation time in silicon. Only weak enhancement of the responsivity at high frequencies due to plasma oscillations is found similar to Ref. [17].

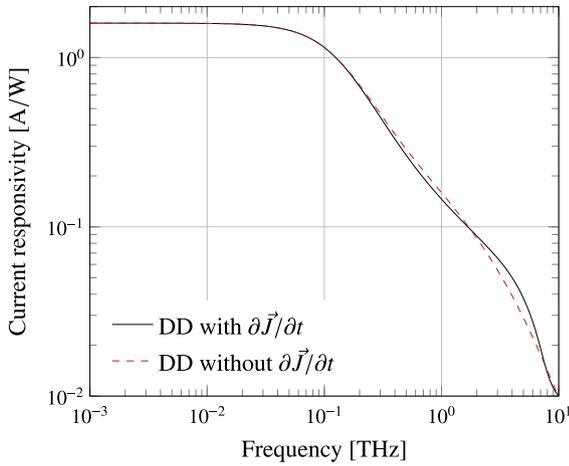


Fig. 7. Current responsivity at $V_G = 0.56$ V, $V_D = 0.0$ V, $V_{AC} = 0.1$ mV and room temperature for the NMOSFET as a function of frequency with and without the time derivative of the current density.

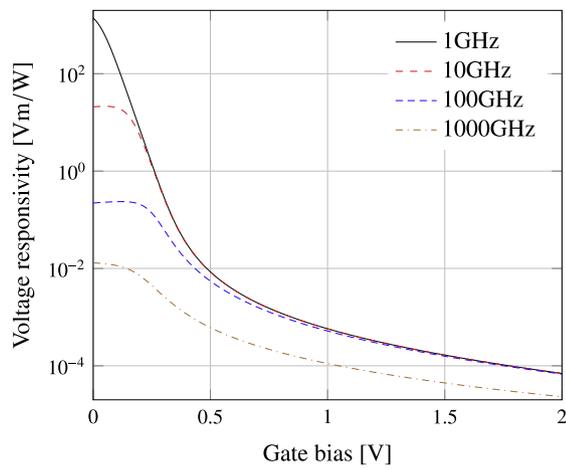


Fig. 8. Voltage responsivity at $V_D = 0.0$ V, $V_{AC} = 0.1$ mV and room temperature for the NMOSFET as a function of the gate bias for selected frequencies.

The voltage responsivity (Fig. 8) shows a similar behavior as the current one (Fig. 6), where peak values are shifted to lower gate voltages due to the multiplication with the zero-frequency channel resistance (15). Since the channel resistance is evaluated for zero frequency, the frequency dependence of the voltage and current responsivities are the same.

The noise-equivalent-power (Fig. 9) has distinct minima, which are found for gate biases below the gate voltages for peak current responsivity (Fig. 6), because it is proportional to the ratio of the square root of the zero-frequency channel conductance and current responsivity (16). The frequency dependence is therefore inversely proportional to the one of the responsivity and the noise-equivalent-power strongly increases with frequency.

The response of the DC component of the electric potential within the NMOSFET due to the AC signal is shown in Fig. 10 for the case of current and voltage sensing along the interface between the oxide and silicon bulk (the line with zero vertical coordinate in Fig. 3). At 1 GHz the potential response for current sensing $\Delta\phi_{CS}^{DC}$ is almost symmetric, whereas at 1 THz the maximum of the absolute value of the potential response moves towards the source. In the case of voltage sensing the induced DC drain voltage results in a roughly linearly varying potential in the channel region, which has to be added to the potential response

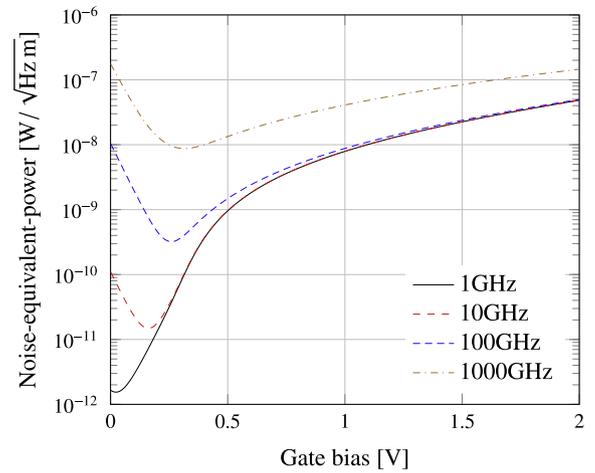


Fig. 9. Noise-equivalent-power at $V_D = 0.0$ V, $V_{AC} = 0.1$ mV and room temperature for the NMOSFET.

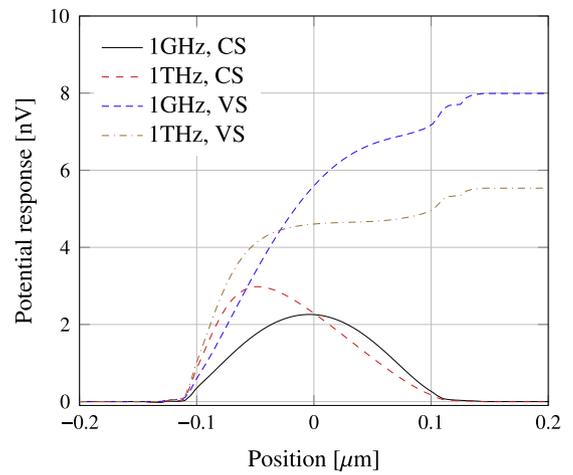


Fig. 10. DC shift of the electric potential along the oxide/silicon interface of the NMOSFET due to the AC signal at $V_G = 0.56$ V, $V_D = 0.0$ V, $V_{AC} = 0.1$ mV and room temperature for two frequencies and current (CS) and voltage (VS) sensing.

$$\Delta\phi_{VS}^{DC} = \Delta\phi_{CS}^{DC} + \frac{\partial\phi}{\partial V_D} V_D^{DC}, \quad (17)$$

where $\partial\phi/\partial V_D$ is the derivative of the potential w.r.t. the drain voltage for zero frequency and zero drainsource bias. The built-up of the DC response of the potential therefore occurs in this case mostly in the channel near the source.

5. Conclusions

The developed drift-diffusion model goes beyond the state-of-the-art, because it is 2D in real space and includes the time derivatives of the current densities. It is thus possible to capture many effects neglected by the usual quasi-1D approaches. We have demonstrated that the drift-diffusion model can be used to simulate NMOSFETs in the THz range and we have assessed its accuracy by comparison with the more fundamental Boltzmann transport equation. It appears that the backward-differentiation formula of the second order based on trigonometric functions is a very CPU-efficient time-integration scheme for the periodic steady-state.

The current and voltage responsivities have been investigated and the highest values are found for gate voltages below the threshold voltage even at one THz. Up to this frequency the impact

of plasma waves seems to be small due to the strong damping by the low electron momentum relaxation time in silicon.

References

- [1] Jungemann C, Bittner K, Brachtendorf HG. Simulation of plasma resonances in MOSFETs for THz-signal detection. In: 2016 joint international EUROSOI workshop and international conference on ultimate integration on silicon (EUROSOI-ULIS), p. 48–51.
- [2] Song H-J, Nagatsuma T, editors. Handbook of terahertz technologies: devices and applications. Singapore: Pan Stanford Publishing; 2015.
- [3] Woolard D, Brown E, Pepper M, Kemp M. Terahertz frequency sensing and imaging: a time of reckoning future applications? Proc IEEE 2005;93(10):1722–43.
- [4] Otsuji T, Shur M. Terahertz plasmonics: good results and great expectations. Microwave Mag, IEEE 2014;15(7):43–50.
- [5] Pfeiffer U, Zhao Y, Grzyb J, Al Hadi R, Sarmah N, Forster W, et al. A 0.53 THz reconfigurable source module with up to 1 mW radiated power for diffuse illumination in terahertz imaging applications. IEEE J Solid-State Circ 2014;49(12):2938–50.
- [6] Dyakonov M, Shur M. Shallow water analogy for a ballistic field effect transistor: new mechanism of plasma wave generation by dc current. Phys Rev Lett 1993;71:2465–8.
- [7] Dyakonov M, Shur M. Plasma wave electronics: novel terahertz devices using two dimensional electron fluid. IEEE Trans Electron Dev 1996;43(10):1640–5.
- [8] Öjefors E, Pfeiffer U, Lisauskas A, Roskos H. A 0.65 THz focal-plane array in a quarter-micron CMOS process technology. IEEE J Solid-State Circ 2009;44(7):1968–76.
- [9] Hadi RA, Sherry H, Grzyb J, Zhao Y, Forster W, Keller HM, et al. A 1 k-pixel video camera for 0.7–1.1 Terahertz imaging applications in 65-nm CMOS. IEEE J Solid-State Circ 2012;47(12):2999–3012.
- [10] Lisauskas A, Bauer M, Ramer A, Ikamas K, Matukas J, Chevtchenko S, et al. Terahertz rectification by plasmons and hot carriers in gated 2D electron gases. In: 2015 international conference on noise and fluctuations (ICNF), p. 1–5.
- [11] Crowne FJ. Contact boundary conditions and the Dyakonov-Shur instability in high electron mobility transistors. J Appl Phys 1997;82(3):1242–54.
- [12] Hong S-M, Jang J-H. Numerical simulation of plasma oscillation in 2-d electron gas using a periodic steady-state solver. IEEE Trans Electron Dev 2015;62(12):4192–8.
- [13] Mateos J, Gonzalez T. Plasma enhanced Terahertz rectification and noise in InGaAs HEMTs. IEEE Trans Terahertz Sci Technol 2012;2(5):562–9.
- [14] Jungemann C, Meinerzhagen B. In-advance CPU time analysis for stationary Monte Carlo device simulations. IEICE Trans Electron 2003;E86-C(3):314–9.
- [15] Selberherr S. Analysis and simulation of semiconductor devices. Wien: Springer; 1984.
- [16] Jungemann C, Meinerzhagen B. On the high frequency limit of the impedance field method for Si. International conference on noise in physical systems and 1/f fluctuations, AIP Conf Proc, vol. 780, p. 799–802.
- [17] Lisauskas A, Pfeiffer U, Öjefors E, Bolvar PH, Glaab D, Roskos HG. Rational design of high-responsivity detectors of terahertz radiation based on distributed self-mixing in silicon field-effect transistors. J Appl Phys 2009;105(11).
- [18] Varga RS. Matrix iterative analysis. Series in automatic computation. Englewood Cliffs (New Jersey): Prentice-Hall; 1962.
- [19] Scharfetter DL, Gummel HK. Large-signal analysis of a silicon read diode oscillator. IEEE Trans Electron Dev 1969;ED-16(1):64–77.
- [20] Kim H, Min HS, Tang TW, Park YJ. An extended proof of the Ramo-Shockley theorem. Solid-State Electron 1991;34:1251–3.
- [21] Gummel HK. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. IEEE Trans Electron Dev 1964:455–65.
- [22] Kundert KS, Sangiovanni-Vincentelli A. Simulation of nonlinear circuits in the frequency domain. IEEE Trans Comput-Aided Des 1986;CAD-5(4):521–35.
- [23] Curtiss CF, Hirschfelder JO. Integration of stiff equations. Proc Nat Acad Sci USA 1952;38:235–43.
- [24] Gear CW. Numerical initial value problems in ordinary differential equations. Englewood Cliffs (New Jersey): Prentice-Hall; 1971.
- [25] Brachtendorf H, Bittner K. Grid size adapted multistep methods for high Q oscillators. IEEE Trans Comput-Aided Des Integr Circ Syst 2013;32(11):1682–93.
- [26] Caughey DM, Thomas RE. Carrier mobilities in silicon empirically related to doping and field. Proc IEEE 1967;55:2192–3.
- [27] Lombardi C, Manzini S, Saporito A, Vanzi M. A physically based mobility model for numerical simulation of nonplanar devices. IEEE Trans Comput-Aided Des 1988;7:1164–70.
- [28] Schenk O, Bollhöfer M, Römer RA. On large-scale diagonalization techniques for the Anderson model of localization. SIAM Rev 2008;50(1):91–112.
- [29] Jacoboni C, Lugli P. The Monte Carlo method for semiconductor device simulation. Wien: Springer; 1989.
- [30] Lundstrom M. Fundamentals of carrier transport. Modular series on solid state devices, vol. 10. New York: Addison-Wesley; 1990.
- [31] Hong S-M, Pham AT, Jungemann C. Deterministic solvers for the Boltzmann transport equation. Computational Microelectronics. Wien (New York): Springer; 2011.



Anisotropic interpolation method of silicon carbide oxidation growth rates for three-dimensional simulation



Vito Šimonka^{a,b,*}, Georg Nawratil^c, Andreas Hössinger^d, Josef Weinbub^{a,b}, Siegfried Selberherr^b

^a Christian Doppler Laboratory for High Performance TCAD at the Institute for Microelectronics, TU Wien, Gußhausstr. 27-29/E360, 1040 Wien, Austria

^b Institute for Microelectronics, TU Wien, Gußhausstr. 27-29/E360, 1040 Wien, Austria

^c Institute of Discrete Mathematics and Geometry, TU Wien, Wiedner Hauptstraße 8-10, 1040 Wien, Austria

^d Silvaco Europe Ltd., Compass Point, St Ives, Cambridge PE27 5JL, United Kingdom

ARTICLE INFO

Article history:

Available online 20 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Silicon carbide
Oxidation
Growth rates
Anisotropy
Interpolation

ABSTRACT

We investigate anisotropic and geometrical aspects of hexagonal structures of Silicon Carbide and propose a direction dependent interpolation method for oxidation growth rates. We compute three-dimensional oxidation rates and perform one-, two-, and three-dimensional simulations for 4H- and 6H-Silicon Carbide thermal oxidation. The rates of oxidation are computed according to the four known growth rate values for the Si- (0001), a- (1120), m- (1100), and C-face (0001). The simulations are based on the proposed interpolation method together with available thermal oxidation models. We additionally analyze the temperature dependence of Silicon Carbide oxidation rates for different crystal faces using Arrhenius plots. The proposed interpolation method is an essential step towards highly accurate three-dimensional oxide growth simulations which help to better understand the anisotropic nature and oxidation mechanism of Silicon Carbide.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Silicon Carbide (SiC) has excellent physical properties and has received significant attention in recent years as a Silicon (Si) replacement material for power device applications due to a high electrical breakdown voltage and a high thermal conductivity. Compared to Si, SiC has approximately a three times wider band gap, ten times larger electrical breakdown voltage, and three times higher thermal conductivity [1–3]. Taking advantages of these properties, the on-state resistance for unipolar devices such as metal-oxide-semiconductor field-effect-transistors (MOSFET) can be reduced by a factor of a few hundreds when replacing Si with SiC [4,5]. Aside from the theoretical advantages in SiC devices, the need for numerical simulation based on accurate models is indispensable to further the success of modern power electronics.

Among the numerous polytypes of SiC, most popular for device applications are 3C-SiC, 4H-SiC, 6H-SiC, and 15R-SiC. These polytypes are characterized by the stacking sequence of the bi-atom layers of the SiC structure. Changing the stacking sequence has a profound effect on the electrical properties. See Fig. 1a for an

atomic view of a 4H-SiC. In this work, we focus on 4H- and 6H-SiC as they have been recognized as the most promising polytypes and are currently commercially available for high power, high frequency, and high temperature applications [6,7].

Thermally grown oxide layers (SiO₂) play a unique role in device fabrication, e.g., lateral structures in planar technology and passivation of device surfaces. Therefore, it is necessary to have a solid understanding of oxidation growth rates and the dependence on the crystallographic planes of SiC. Among the wide bandgap semiconductors, SiC is the only compound semiconductor which can be thermally oxidized in the form of SiO₂, similar to conventional Si substrate. This is seen as one of the most important technological properties of SiC and has motivated considerable effort in its development. The following reaction governs the oxidation of SiC [1]:



As opposed to the relatively simple oxidation of Si, the thermal oxidation of SiC includes five steps [2] (discussed in the following) and is about one order of magnitude slower under the same conditions [8,9]:

1. Transport of molecular oxygen gas to the oxide surface.
2. In-diffusion of oxygen through the oxide film.
3. Reaction with SiC at the SiO₂/SiC interface.

* Corresponding author at: Christian Doppler Laboratory for High Performance TCAD at the Institute for Microelectronics, TU Wien, Gußhausstr. 27-29/E360, 1040 Wien, Austria.

E-mail address: simonka@iue.tuwien.ac.at (V. Šimonka).

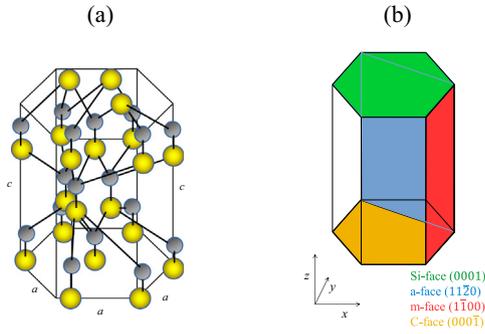


Fig. 1. (a) A schematic illustration of an atomic view of a 4H-SiC polytype with sequence ABAC. Yellow (big) spheres show the Si atoms, gray (small) spheres C atoms, a is the crystal dimension, and c is the crystal height. The ratio between c and a for 4H-SiC is approximately three and for 6H-SiC approximately five. (b) A schematic illustration of common faces of a hexagonal structure. Green (top), blue (diagonal), red (right), and orange (bottom) shapes show the Si-, a-, m-, and C-face, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Out-diffusion of product gases through the oxide film.
5. Removal of product gases away from the oxide surface.

The last two steps are not involved in the oxidation of Si. The first and the last steps are relatively fast and are not rate-controlling steps.

Another unique phenomenon has been observed: The oxidation of SiC is a face-terminated oxidation, i.e., the top and the bottom face have different oxidation rates [10–12]. Additionally, the oxidation of SiC varies also with other crystallographic planes as reported in [7,13,14], see Fig. 1b for common faces of 4H-SiC. The dependence of the oxidation rates on crystal orientation has significant consequences for a non-planar device structure, e.g., the trench design of a U-groove MOSFET, where the oxide is located on all crystallographic planes [7].

In Section 2 we discuss the thermal oxidation process and physical models of Si and SiC, in Section 3 we introduce the temperature dependence of the oxidation growth rates with Arrhenius plots, in Section 4 we discuss anisotropic and geometrical aspects of 4H- and 6H-SiC as well as our proposed interpolation method, and in Section 5 one-, two-, and three-dimensional simulations and calculations for initial and linear growth rates of SiC thermal oxidation are discussed.

2. SiC oxidation models

Thermal oxidation of SiC can be mathematically described with the Deal-Grove model [15], which has been originally proposed to explain the Si oxidation process. According to this model, the oxidation occurs by diffusion of the oxidant to the SiO₂/Si interface, where it reacts with Si. The relation between the oxide thickness X and oxidation time t is thus expressed by the following equation:

$$X^2 + AX = B(t + \tau), \quad (2)$$

where B/A , B , and τ are the linear rate constant, parabolic rate constant, and the constant related to the initial oxide thickness, respectively. Eq. (2) can be rewritten as ordinary differential equation:

$$\frac{dX}{dt} = \frac{B}{A + 2X} \quad (3)$$

In the Deal-Grove model, the linear rate constant B/A is the oxidation rate when $2X \ll A$, in which the interface reaction is the rate-controlling step [15]. The parabolic rate constant B is the oxidation rate when $2X \gg A$, in which the diffusion of oxygen through the oxide film SiO₂ is the rate-controlling step [15].

The oxidation process cannot be characterized by the Deal-Grove model for the thin oxide region in Si and SiC, hence Massoud et al. [16] have proposed an empirical relation to describe the growth rate enhancement in a thin oxide regime. This model includes an additional exponential term [16],

$$\frac{dX}{dt} = \frac{B}{A + 2X} + C \exp\left(-\frac{X}{L}\right), \quad (4)$$

where C and L are the exponential prefactor and the characteristic length, respectively.

It has been reported that the linear rate constant B/A and initial growth rate $B/A + C$ highly depend on the crystal orientation of SiC [1,6,7,14], i.e., the growth rate values are different for the surface oxidation on different faces of the crystal. On the other hand, the parabolic rate constant B does not depend on the crystal orientation [1].

The Deal-Grove model and Massoud's empirical relation were originally proposed for Si oxidation, but can be applied in a modified form to SiC oxidation [2]. For SiC oxidation the Massoud empirical relation can reproduce the oxide growth better than Deal-Grove model [17,18]. However, due to the one-dimensional nature both models fail to correctly predict the oxide growth for three-dimensional SiC structures. Our approach extends these models by incorporating the crystal direction dependence into the oxidation growth rates, thus enabling accurate three-dimensional modeling.

3. Temperature dependencies

Rates of chemical reactions depend on various physical quantities, e.g., temperature and pressure. The collision theory and transition state theory implies that chemical reactions typically proceed faster at higher temperature and pressure, and slower at lower temperature and pressure. The molecules move faster as the temperature increases and therefore collide more frequently, which changes the properties of the involved chemical reactions.

The relation between the absolute temperature T and the rate constant k is given via an Arrhenius equation [3,19]:

$$k = Z \exp\left(-\frac{E_a}{k_B T}\right) \quad (5)$$

Z is the pre-exponential factor discussed below, E_a is the activation energy of the chemical reaction, and k_B is the Boltzmann constant. Recalling that kT is the average kinetic energy, it becomes apparent that the exponent is the ratio of the activation energy E_a to the average energy of colliding molecules. The larger the ratio, the smaller the reaction rate. This means that high temperature and low activation energy favor larger rate constants, and thus speed up the reaction. The pre-exponential factor Z is known as the frequency or collision factor and can be calculated from kinetic molecular theory. In other words, Z is equal to the fraction of molecules which are involved in a chemical reaction, if (1) the activation energy $E_a = 0$ or (2) the kinetic energy of all molecules exceeds E_a [20,21].

The Arrhenius equation can be used to determine the activation energy of the oxidation growth rates [1]. The equation can be written in a non-exponential form by applying the natural logarithm on both sides of the equation:

$$\ln(k) = -\frac{E_a}{k_B} \frac{1}{T} + \ln(Z) \quad (6)$$

In this form, the Arrhenius equation is more convenient to use and to interpret graphically, as it appears as a linear function

$$\psi = m\chi + n, \quad (7)$$

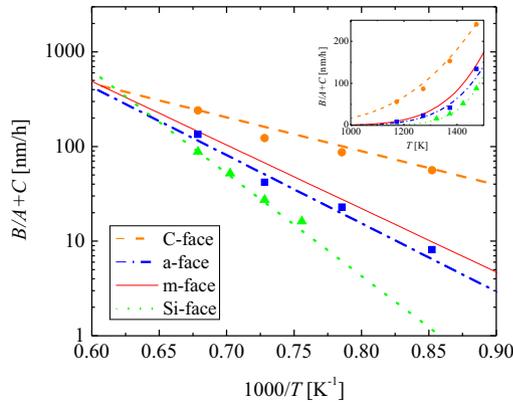


Fig. 2. Arrhenius plot for initial growth rates $B/A + C$ for the Si- (green), m- (red), a- (blue), and C-face (orange) of dry thermal oxidation of 4H-SiC. Experimental data for the Si-, a-, and C-face (symbols) are obtained from [1] and the data for the m-face (solid red lines) are approximated from [14]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where $\psi = \ln(k)$ is the dependent variable, $\chi = 1/T$ is the independent variable, $m = -E_a/R$ is the slope, and $n = \ln(Z)$ is the intercept. The activation energy is thus determined from the growth rate values at different temperatures by plotting $\ln(k)$ as a function of $1/T$.

Fig. 2 shows an exemplary Arrhenius plot for initial growth rates of 4H-SiC dry thermal oxidation. The data points are measured values, the dashed lines are fits using Massoud's empirical relation (4), and the solid lines are approximated values. We have obtained the growth rates and activation energies of the Si-, a-, and C-face from experimentally measured data [1] and approximated the growth rate and activation energy for the m-face based on published oxide thicknesses [14], as there are no experimental data available. We use these data sets to analyze the effect of temperature on oxidation and to obtain the fixed growth rate values for the proposed interpolation method.

4. Interpolation method

With respect to the interpolation method, the geometrical aspects of SiC are mathematically described according to basic crystallography and experimental findings [14,22]. A unit cell of the hexagonal crystal structure includes six $(1\bar{1}00)$ and six $(11\bar{2}0)$ crystallographic faces symmetric with respect to the z axis, while there is only one (0001) face on the top and one $(000\bar{1})$ face on the bottom of the crystal (see Fig. 1b).

We propose a direction dependent interpolation method [23] to convert an arbitrary crystal direction into a growth rate for oxidation, according to a set of known growth rate values. For fixed points of oxidation growth rates we use growth rates of experimentally examined crystallographic faces of SiC [1,6,13]: Si-, m-, a-, and C-face, which correspond to the (0001) , $(1\bar{1}00)$, $(11\bar{2}0)$, and $(000\bar{1})$ crystal directions, respectively.

The proposed interpolation method consists of a symmetric star shape in the x - y plane and a tangent-continuous union of two half-ellipses in z direction. See Fig. 3 for a schematic representation of the method in the x - y and x - z plane. Dark blue lines show the proposed interpolation between fixed points, which takes the symmetry of the hexagonal structure of SiC into account. Arrow directions and lengths represent crystal directions toward SiC faces and oxidation growth rate values, respectively. We could also consider a less accurate linear interpolation (shown with dotted black lines) with sharp edges, which would also fit the geometry of SiC, but the non-linear method offers considerable higher accuracy and is thus further used in this work.

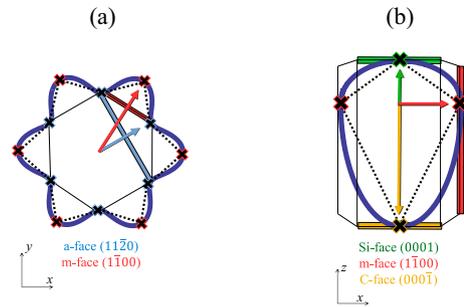


Fig. 3. Schematic representation of the proposed interpolation method in the (a) x - y and (b) x - z plane. A linear (black dotted) and a non-linear (dark blue line) interpolation is calculated according to four known growth rate values (black crosses) of Si- (green), m- (red), a- (blue), and C-face (orange square). Colored arrows represent crystal directions towards the corresponding faces. The arrow lengths are proportional to the oxidation growth rates. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The parametric expression of the three-dimensional interpolation method is

$$\begin{aligned} x &= (k_y + (k_x - k_y) \cos^2(3t)) \cos(t) \cos(u), \\ y &= (k_y + (k_x - k_y) \cos^2(3t)) \sin(t) \cos(u), \\ z &= k_z \sin(u), \end{aligned} \quad (8)$$

where $t \in [0, 2\pi]$ and $u \in [-\pi/2, \pi/2]$ are arbitrary parametric variables and $k_{x,y,z}$ are known oxidation growth rates in x , y , and z direction, respectively. In our case we consider: $k_x = k_m$, $k_y = k_a$, and $k_z = k_c$ or k_{Si} .

As shown in several studies [1,24,25], the oxide growth on top and bottom of the crystal is different, thus we need to calculate the positive and negative z coordinates separately:

$$\begin{aligned} z &= k_z^+ \sin(u) \text{ for } u \geq 0 \\ z &= k_z^- \sin(u) \text{ for } u < 0 \end{aligned} \quad (9)$$

k_z^+ and k_z^- correspond to the growth rate in the direction of the Si- and C-face, respectively. Thus, we define that $k_z^+ = k_{Si}$ and $k_z^- = k_c$.

The parametric expression of the proposed interpolation method can be converted into an explicit expression, which describes the surface as the zero set of equation $F(x,y,z) = 0$, where x , y , and z are the variables, e.g., vector coordinates. The explicit representation is more general and more suitable for one- and two-dimensional calculations, and is more closely related to the concepts of constructive solid geometry and modeling. However, the parametric form is more useful for three-dimensional calculations and remains dominant in computer graphics and geometrical modeling.

5. Results and discussion

We have performed several one-, two-, and three-dimensional calculations of growth rates and simulations of thermal SiC oxidation using the proposed interpolation method together with Massoud's empirical relation and Arrhenius plots. Fig. 4 shows schematic representations of the hexagonal crystal structure and variables for simulations which are used in the following discussion. Out of simplicity, the two-dimensional simulations are performed either in the x - y or in the x - z plane. The input of the interpolation method is an arbitrary crystal direction vector \vec{v} contained in the x - y or x - z plane, for which the oxidation growth rate has to be computed. By denoting the angle between \vec{v} and the x -axis by α in x - y plane or by β in x - z plane, we get trivial relations

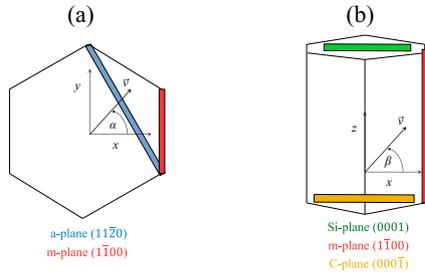


Fig. 4. Schematic representation of the hexagonal crystal structure in (a) x - y and (b) x - z plane. \vec{v} is crystal direction vector, α is an angle between the x and y -axis, and β is an angle between the vector and the x and z -axis. Blue (diagonal), red (right), green (top), and orange (bottom) squares represent a-, m-, Si-, and C-face, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$x = |\vec{v}| \cos \alpha, \quad (10)$$

$$y = |\vec{v}| \sin \alpha,$$

and

$$x = |\vec{v}| \cos \beta, \quad (11)$$

$$z = |\vec{v}| \sin \beta,$$

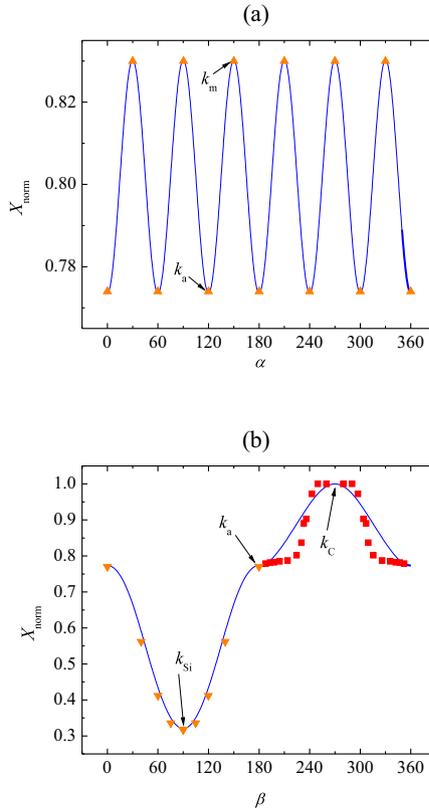


Fig. 5. Thermally grown oxide thickness as a function of (a) angle α in x - y plane and (b) angle β in x - z plane as shown in Fig. 4. The oxide thickness X_{norm} is normalized using the maximal oxide thickness from individual simulations and measurements for direct comparisons. Blue solid lines are simulations performed with available oxidation models using the proposed interpolation method. Orange triangles and red squares are experimental measurements from [14,22], respectively. Black arrows show fixed points for the interpolation method: k_m , k_a , k_{Si} , and k_c . Simulations are performed for the wet thermal oxidation of 6H-SiC (0001) Si-face (n-type, on-axis) at $T = 1100^\circ\text{C}$ for 720 min. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where x, y , and z are the crystal direction vector coordinates $\vec{v}(x, y, z)$ and $|\vec{v}|$ is the vector length. The crystal direction vector \vec{v} is normalized, thus $|\vec{v}| = 1$.

5.1. One-dimensional simulations

Thermally grown oxide thicknesses as a function of α and β , the angle between the crystal direction vector and the x -axis, are shown in Fig. 5. The oxide thicknesses have been normalized with the maximal oxide thickness from individual simulations and measurements for direct comparisons, i.e., $X_{\text{norm}} = 1$ corresponds to the maximum oxide thickness, whereas $X_{\text{norm}} = 0$ corresponds to no oxide at all. Fig. 5a shows simulation results (blue lines) for the crystallographic plane x - y and Fig. 5b for the crystallographic plane x - z . Orange triangles are measurements by Christiansen and Helbig [14] and red squares are measurements by Tokura et al. [22]. The calculation of growth rates as well as the simulations of thermal oxidation are performed for the angle from 0° to 360° for both planes.

In Fig. 5a, we observe six maxima and six minima in the x - y plane, which correspond to the m- and the a-face, respectively. On the other hand, in Fig. 5b we observe one maximum and one minimum in the x - z plane, which correspond to the C- and the Si-face, respectively. From comparing the normalized oxide thicknesses with the measurements from Christiansen and Helbig [14] we can argue that the proposed interpolation method fits experimental data very well. On the other hand, comparing results with measurements from Tokura et al. [22], the simulations do not fit all experimental data perfectly, but the shape and the extreme values are properly consistent.

5.2. Two-dimensional simulations

Fig. 6 shows the two-dimensional interpolation of the linear growth rates B/A of the wet thermal oxidation at $T = 1100^\circ\text{C}$ in

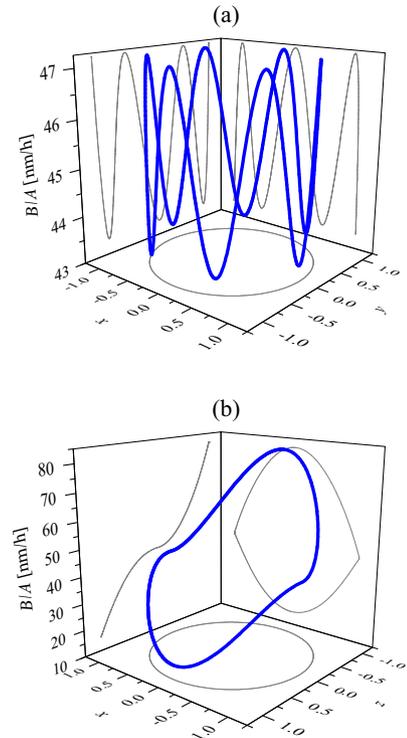


Fig. 6. Two-dimensional calculations of the SiC linear oxidation growth rates B/A using the proposed interpolation method in the (a) x - y and (b) x - z plane. x, y , and z are normalized crystal direction vector coordinates. Interpolation is performed for the wet thermal oxidation of 6H-SiC (0001) Si-face (n-type, on-axis) at $T = 1100^\circ\text{C}$.

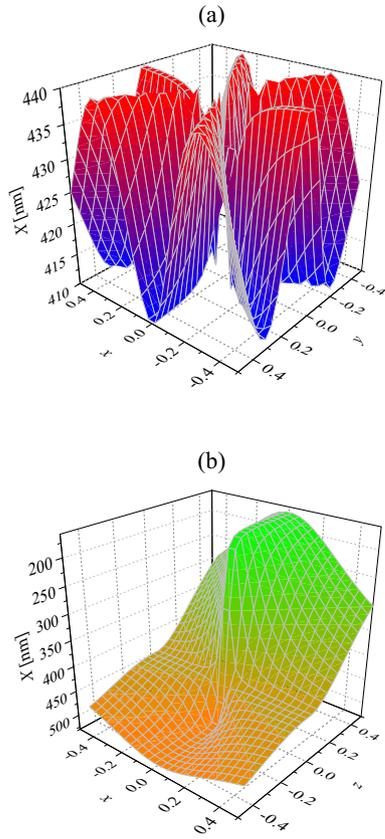


Fig. 7. Two-dimensional simulations of the wet thermal oxidation of SiC in the (a) x - y and (b) x - z plane. The figures show the final oxide thicknesses X as a function of the normalized crystal direction vector coordinates x, y , and z . The final oxide thicknesses are obtained using the available oxidation models and the results from previous plots (cf. Fig. 6). Red, blue, orange, and green colors represent oxide thicknesses for the m -, a -, C -, and Si -face, respectively. Simulations are performed for the wet thermal oxidation of 6H-SiC (0001) Si-face (n-type, on-axis) at $T = 1100^\circ\text{C}$ for 720 min. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the x - y and the x - z plane using normalized crystal direction vector coordinates x, y , and z as input for the interpolation method. The combination of vector coordinates is set in a way that it gives all possible crystal direction vectors to compute growth rates, which is clearly seen by the gray circle below the plots. Fixed points for the interpolation were approximated from [14,22] and calibrated according to the available data [26–28], thus the linear growth rates B/A towards the Si-, a -, m -, and C -faces are $k_{Si} = 15.25$ nm/h, $k_a = 43.33$ nm/h, $k_m = 47.19$ nm/h, and $k_C = 81.60$ nm/h, respectively. The results from the interpolation are used in the following simulations.

Oxide thicknesses thermally grown in a wet circumstance as a function of the vector coordinates x, y , and z are summarized in Fig. 7. These results show final oxide thicknesses X in the x - y and the x - z plane of a crystal. The combination of vector coordinates (x, y) or (x, z) define the crystal direction for which the oxide thickness is calculated. The results from Fig. 7a and b cover the whole x - y and x - z plane full of the given crystal directions. In the first plot we can observe six maxima and six minima, which correspond to the m - and the a -face, respectively. The final oxide thickness in the direction of the m -face is approximately 442 nm, and in the direction of the a -face 412 nm. On the second plot we observe a maximal and minimal oxide thickness, which correspond to the C - and the Si -face, respectively. The final oxide thickness in the direction of the Si -face is approximately 173 nm and in the direc-

tion of the C -face 531 nm. The results are in agreement with [14] for wet thermal oxidation of 6H-SiC.

5.3. Three-dimensional simulations

Fig. 8 shows the three-dimensional interpolation of the initial growth rates $B/A + C$ of the dry thermal oxidation at $T = 1100^\circ\text{C}$ using the parametric expression of the proposed interpolation method (8) and (9). Fixed oxidation growth rate values are $k_{Si} = 27.36$ nm/h, $k_a = 41.83$ nm/h, $k_m = 66.64$ nm/h, and $k_C = 122.8$ nm/h, obtained from the Arrhenius plot (Fig. 2) at $T = 1100^\circ\text{C}$. The growth rate surface is given by a nonlinear interpolation between these known growth rate values and follows the geometry of SiC, i.e., the crystallographic planes tangent to the growth rate surface at k_{Si}, k_m, k_a , and k_C are parallel to the corresponding faces. The distance from the origin $(0,0,0)$ to any point on the growth rate surface gives the oxidation rate in direction to this point. The set of growth rate values together with the SiC oxidation models are used for the following three-dimensional simulations of 4H-SiC dry thermal oxidation.

The surface of the oxide thicknesses of the dry initial thermal oxidation of 4H-SiC is shown in Fig. 9. The simulations are performed using Massoud’s empirical relation and previously calculated initial growth rate values $B/A + C$ (Fig. 8). The distance from the origin $(0,0,0)$ to any point on the surface gives the oxide thickness in the direction of this point. For the oxidation time of 15 min, the final oxide thicknesses in the direction of the common

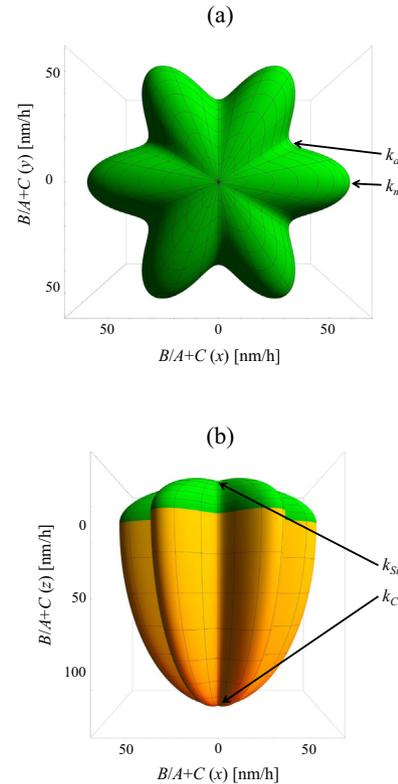


Fig. 8. Three-dimensional calculations of the SiC initial oxidation growth rates $B/A + C$ obtained with the parametric expression of the proposed interpolation method. The figure shows the (a) top and (b) front view of the growth rates’ surface. An arbitrary direction growth rate is calculated according to the four known growth rates (k_{Si}, k_m, k_a , and k_C) shown with black arrows. The surface color shows calculations for positive (green) and negative (orange) z direction. Interpolation is performed for the dry thermal oxidation of 4H-SiC (0001) Si-face (n-type, on-axis) at $T = 1100^\circ\text{C}$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

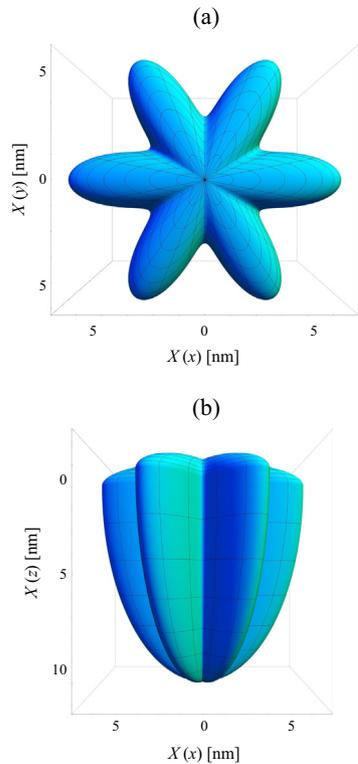


Fig. 9. Three-dimensional simulations of the dry thermal oxidation of SiC. The figure shows the (a) top and (b) front view of the oxide thicknesses surface. The final thicknesses X in an arbitrary crystal directions are obtained using the available oxidation models and the results from previous plots (cf. Fig. 8). Simulations are performed for the dry thermal oxidation of 4H-SiC (0001) Si-face (n-type, on-axis) at $T = 1100\text{ }^{\circ}\text{C}$ for 15 min.

faces are approximately $X_m = 5.74\text{ nm}$, $X_a = 2.67\text{ nm}$, $X_{Si} = 1.42\text{ nm}$, and $X_C = 11.16\text{ nm}$. These results are in agreement with [1,6] for the dry thermal oxidation of 4H-SiC.

6. Conclusions

We investigated the anisotropy of 4H- and 6H-SiC oxidation processes with regard to surface orientations. By carefully studying geometrical aspects of the hexagonal crystal structure we have proposed an interpolation method to compute oxidation growth rate constants in one-, two-, and three-dimensional problems. The interpolation method includes well known anisotropy of the oxidation of the Si- and the C-face, as well as the anisotropic behavior of the m- and the a-face. In the basic crystal plane x - y , which intersects with the origin of the unit cell, six maxima and six minima are computed, corresponding to the crystal symmetry in the shape of a star.

Using the proposed interpolation method we have calculated linear growth rates for the wet thermal oxidation of 6H-SiC at $T = 1100\text{ }^{\circ}\text{C}$ and initial growth rates for the dry thermal oxidation of 4H-SiC at $T = 1100\text{ }^{\circ}\text{C}$. With results from the interpolation we have additionally performed one-, two-, and three-dimensional simulations using the Massoud oxidation model.

The presented results of thermal oxidation of SiC are in good agreement with experimental findings from the literature. We can also show that the proposed nonlinear interpolation method fits the geometry dependence of 4H- and 6H-SiC oxidation very well. Moreover, with the proposed method, we are now able to simulate three-dimensional dry and wet oxidation of SiC, where the only limiting factor is the set of fixed growth rates, which are usually obtained from measurements.

Acknowledgment

The authors wish to thank Y. Hijikata for providing experimental data. The financial support by the *Austrian Federal Ministry of Science, Research and Economy* and the *National Foundation for Research, Technology and Development* is gratefully acknowledged.

References

- [1] Goto D, Hijikata Y, Yagi S, Yaguchi H. Differences in SiC thermal oxidation process between crystalline surface orientations observed by in-situ spectroscopic ellipsometry. *J Appl Phys* 2015;117(9):095306.
- [2] Hijikata Y. Physics and technology of silicon carbide devices. InTech, Croatia; 2013.
- [3] Gupta SK, Akhtar J. Thermal oxidation of silicon carbide (SiC)-experimentally observed facts. InTech, China; 2011.
- [4] Weitzel CE, Palmour JW, Carter CH, Moore K, Nordquist KJ, Allen S, et al. Silicon carbide high-power devices. *IEEE Trans Electron Dev* 1996;43(10):1732–41.
- [5] Casady J, Johnson RW. Status of silicon carbide (SiC) as a wide-bandgap semiconductor for high-temperature applications: a review. *Solid-State Electron* 1996;39(10):1409–22.
- [6] Song Y, Dhar S, Feldman LC, Chung G, Williams JR. Modified deal grove model for the thermal oxidation of silicon carbide. *J Appl Phys* 2004;95(9):4953–7.
- [7] Harris C, Afanas'ev V. SiO₂ as an insulator for SiC devices. *Microelectron Eng* 1997;36(1–4):167–74.
- [8] Vickridge IC, Ganem JJ, Battistig G, Szilagyi E. Oxygen isotopic tracing study of the dry thermal oxidation of 6H-SiC. *Nucl Instrum Methods Phys Res Sect B: Beam Interact Mater Atoms* 2000;161:462–6.
- [9] Knaup JM, Deák P, Frauenheim T, Gali A, Hajnal Z, Choyke WJ. Theoretical study of the mechanism of dry oxidation of 4H-SiC. *Phys Rev B* 2005;71(23):235321.
- [10] Schürmann M, Dreiner S, Berges U, Westphal C. Structure of the interface between ultrathin SiO₂ films and 4H-SiC (0001). *Phys Rev B* 2006;74(3):035309.
- [11] Fiorenza P, Raineri V. Reliability of thermally oxidized SiO₂/4H-SiC by conductive atomic force microscopy. *Appl Phys Lett* 2006;88(21):2112.
- [12] Yamamoto T, Hijikata Y, Yaguchi H, Yoshida S. Oxygen-partial-pressure dependence of SiC oxidation rate studied by in situ spectroscopic ellipsometry. In: *Proc Mater Sci Forum*. p. 667–70.
- [13] Ahn JJ, Jo YD, Kim SC, Lee JH, Koo SM. Crystallographic plane-orientation dependent atomic force microscopy-based local oxidation of silicon carbide. *Nanoscale Res Lett* 2011;6(1):1–5.
- [14] Christiansen K, Helbig R. Anisotropic oxidation of 6H-SiC. *J Appl Phys* 1996;79(6):3276–81.
- [15] Deal BE, Grove A. General relationship for the thermal oxidation of silicon. *J Appl Phys* 1965;36(12):3770–8.
- [16] Massoud HZ, Plummer JD, Irene EA. Thermal oxidation of silicon in dry oxygen accurate determination of the kinetic rate constants. *J Electrochem Soc* 1985;132(7):1745–53.
- [17] Yamamoto T, Hijikata Y, Yaguchi H, Yoshida S. Growth rate enhancement of (0001)-face silicon-carbide oxidation in thin oxide regime. *Japan J Appl Phys* 2007;46(8L):L770.
- [18] Yamamoto T, Hijikata Y, Yaguchi H, Yoshida S. Oxide growth rate enhancement of silicon carbide (0001) Si-faces in thin oxide regime. *Japan J Appl Phys* 2008;47(10R):7803.
- [19] Arrhenius S. Über die Dissociationswärme und den Einfluss der Temperatur auf den Dissociationsgrad der Elektrolyte. Germany: Wilhelm Engelmann; 1889.
- [20] Segel IH et al. Enzyme kinetics. New York: Wiley; 1975.
- [21] Chang R. Physical chemistry for the biosciences. California: University Science Books; 2005.
- [22] Tokura N, Hara K, Miyajima T, Fuma H, Hara K. Current-voltage and capacitance-voltage characteristics of metal/oxide/6H-silicon carbide structure. *Japan J Appl Phys* 1995;34(10R):5567.
- [23] Šimonka V, Nawratil G, Hössinger A, Weinbub J, Selberherr S. Direction dependent three-dimensional silicon carbide oxidation growth rate calculations. In: *Proceedings of 2016 joint international EUROSOL workshop and international conference on ultimate integration on silicon*. p. 226–9.
- [24] Hijikata Y, Yaguchi H, Yoshida S. A kinetic model of silicon carbide oxidation based on the interfacial silicon and carbon emission phenomenon. *Appl Phys Exp* 2009;2(2):021203.
- [25] Ogawa S, Takakuwa Y. Rate-limiting reactions of growth and decomposition kinetics of very thin oxides on Si (001) surfaces studied by reflection high-energy electron diffraction combined with auger electron spectroscopy. *Japan J Appl Phys* 2006;45(9R):7063.
- [26] Opila EJ. Oxidation kinetics of chemically vapor-deposited silicon carbide in wet oxygen. *J Am Ceram Soc* 1994;77(3):730–6.
- [27] Hijikata Y, Yaguchi H, Yoshida S, Takata Y, Kobayashi K, Nohira H, et al. Characterization of oxide films on 4H-SiC epitaxial (0001) faces by high-energy-resolution photoemission spectroscopy: comparison between wet and dry oxidation. *J Appl Phys* 2006;100(5):053710.
- [28] Hosoi T, Nagai D, Shimura T, Watanabe H. Exact evaluation of interface-reaction-limited growth in dry and wet thermal oxidation of 4H-SiC (0001) Si-face surfaces. *Japan J Appl Phys* 2015;54(9):098002.



Framework to model neutral particle flux in convex high aspect ratio structures using one-dimensional radiosity



Paul Manstetten^{a,*}, Lado Filipovic^b, Andreas Hössinger^c, Josef Weinbub^{a,b}, Siegfried Selberherr^b

^a Christian Doppler Laboratory for High Performance TCAD, Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Wien, Austria

^b Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, 1040 Wien, Austria

^c Silvaco Europe Ltd., Compass Point, St Ives, Cambridge PE27 5JL, United Kingdom

ARTICLE INFO

Article history:

Available online 20 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

High aspect ratio
Neutral particle flux
Etching
Radiosity
View factor

ABSTRACT

We present a computationally efficient framework to compute the neutral flux in high aspect ratio structures during three-dimensional plasma etching simulations. The framework is based on a one-dimensional radiosity approach and is applicable to simulations of convex rotationally symmetric holes and convex symmetric trenches with a constant cross-section. The framework is intended to replace the full three-dimensional simulation step required to calculate the neutral flux during plasma etching simulations. Especially for high aspect ratio structures, the computational effort, required to perform the full three-dimensional simulation of the neutral flux at the desired spatial resolution, conflicts with practical simulation time constraints. Our results are in agreement with those obtained by three-dimensional Monte Carlo based ray tracing simulations for various aspect ratios and convex geometries. With this framework we present a comprehensive analysis of the influence of the geometrical properties of high aspect ratio structures as well as of the particle sticking probability on the neutral particle flux.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

High aspect ratio structures are essential for the fabrication of various semiconductor devices, where the aspect ratio (AR) of the structure is defined as *depth/diameter* in case of cylinders and as *depth/width* in case of trenches. One particular example is negative-AND (NAND) flash cell fabrication [1], where three-dimensional multi-layer designs (3D-NAND) involve vertical holes which require aspect ratios above 40. Significant pressure on control of the fabrication process as well as on modeling and simulation techniques originates from these high aspect ratio structures.

One process to fabricate high aspect ratio structures is ion-enhanced chemical etching (IECE) [2]. In this process, the surface is exposed to reactive atoms and molecules from the plasma, which chemically react with the surface to form a volatile product. However, not only volatile products are created in this reaction, but also non-volatile by-products which hinder subsequent surface reactions and therefore decrease the etch rate. This chemical sealing is frequently desired on the vertical sidewall of high aspect ratio structures. To maintain a high etch rate at the bottom region of a structure, the surface is additionally bombarded with vertically

accelerated ions, with the purpose of removing the non-volatile by-products on exposed areas. This makes a highly anisotropic chemical etching possible, supporting the fabrication of high aspect ratio structures.

To simulate an IECE process, a common approach is to model the reactive atoms and molecules of the plasma as electrically neutral particles that diffuse into the domain. In contrast, the accelerated ions are modeled as a directed source. A general simulation sequence for a single time step is to (a) calculate the local neutral particle flux and the local ion flux adsorbed on the surface, (b) model the local surface reaction using the obtained flux rates, and (c) calculate the new surface positions.

Common approaches for three-dimensional flux calculation are Monte Carlo ray tracing [3] and radiosity based [4] methods. When applying these methods to high aspect ratio structures, the computational costs for the neutral flux calculation dominates the simulation. The local neutral flux originating from multiple reflections becomes the dominant component towards the bottom of the structures; this multiplies the computational effort by the number of considered reflection events, compared to the costs for the computation of the direct flux.¹ Also, the flux rates can easily vary by orders of magnitude along the structure depth; this increases the

* Corresponding author.

E-mail address: manstetten@iue.tuwien.ac.at (P. Manstetten).

¹ The flux which originates from direct visibility of the source area.

number of particles necessary to obtain an acceptable signal-to-noise ratio when using a ray tracing approach. For spatial resolutions typically desired for practical simulation cases, this leads to high computational costs for the full three-dimensional computation of the local neutral flux using Monte Carlo ray tracing or radiosity based methods.

We suggest to use a one-dimensional approximation for the calculation of the local neutral flux inside high aspect ratio structures. Our approach, initially introduced in [5], is radiosity based and is applicable to simulations of convex rotationally symmetric holes and convex symmetric trenches with a constant cross-section.

The adsorption of the neutral particles is modeled with a sticking probability s as a locally constant parameter of the surface. All sources and reflections are treated as ideal diffusive, which is a common assumption for neutral particles [6]. Molecular flow (ballistic transport) is assumed for the neutral particles. The sum of these assumptions allows for the computation of the neutral flux distribution using a radiosity approach, which was originally used in the context of heat transfer [7] and later adopted in computer graphics to compute global illumination [8].

The surface of the structure is discretized into elements along the line of symmetry. Assuming a constant flux and a constant sticking probability s over each surface element, we reformulate the discrete radiosity equation to obtain a receiving perspective, which allows for fully adsorbing surface elements.

We establish a general formulation to compute the view factor between two elements of a convex rotationally symmetric hole, based on a formula for the view factor between two coaxial disks of unequal radius. The view factors between two elements of a convex symmetric trench with a constant cross-section is derived using the crossed-strings method [7].

The framework is validated using a three-dimensional Monte Carlo ray tracing based simulator [9] by comparing results for different aspect ratios and sticking probabilities. Furthermore, we study the influence of geometric variations along the wall, as well as the variations of the particle sticking probability, on the flux distribution.

Kokkoris et al. [6] also proposed a framework to approximate the neutral flux in long trenches and holes by exploiting symmetry properties of the structures: The three-dimensional problem is reduced to a line integral and the Nyström method [10] is used for discretization, where a special numerical treatment is needed to avoid singularities. Spikes and oscillations of the solution near corners of the structure were reported, when the resolution is not refined (compared to the resolution required by the Nyström method) at these critical spots. Assumptions for the neutral flux, which are the same for our framework, are the ideal diffuse sources/reflections, the locally constant sticking probability, and molecular flow (ballistic transport without considering inter-particle collisions) of the neutral particles.

In the following sections we first define the simulation domain and introduce the surface model (Section 2). Then, we derive the receiving perspective for the discrete radiosity equation (Section 3) and describe the computation of the view factors for holes and trenches (Section 4). Finally, we present the results of the validation and the effects of geometric variations on the wall (Section 5).

2. Simulation domain

For cylindrical holes, the simulation domain is a rotationally symmetric closed convex surface. For trenches, the simulation domain is a trench with a closed convex symmetric cross-section. The neutral flux source is modeled by closing the structures at the top. This leads to a disk-shaped source and a strip-shaped source for holes and trenches, respectively. Fig. 1a and b illustrates the

cross-sections of domains with vertical walls and with a kink at one half of the depth, respectively.

The surface adsorption is modeled using a locally constant sticking probability s . The received flux R is split according to s into an adsorbed flux A and a re-emitted flux RE as depicted in Fig. 1c. Source areas additionally emit flux E independent of R .

For the remainder of this work, a sticking probability $s_s = 1$ is used for source areas which to not have any reflections originating from these artificial areas; the bottom is modeled as a fully adsorbing area with a sticking probability $s_b = 1$. A constant sticking probability s_w is used for the walls of the structures. These choices represent a reasonable approximation to the prevalent conditions for the neutral particles in an IECE environment.

3. Radiosity equation

Our assumptions, particularly that all sources and surfaces are ideal diffuse and that the transport of the neutral particles is ballistic, allows for the use of a radiosity formulation.

By assuming a constant flux and a constant sticking probability over each surface element, the problem can be formulated using the discrete radiosity equation: for a surface element i the equation reads

$$B_i = E_i + (1 - \alpha_i) \sum_j (F_{ji} B_j), \quad (1)$$

where B is the radiosity (sum of emitted and reflected energies), E is the self-emitted energy, α is the absorptance, and F_{ji} is the view factor (proportion of the radiated energy, which leaves element j and is received by element i). We adapt (1) to our problem by substituting the absorptance α with the sticking probability s and identifying the adsorbed flux as the adsorbed energy A . The radiosity B is then related to the adsorbed energy A by

$$A_i = (B_i - E_i) \frac{s_i}{1 - s_i}. \quad (2)$$

Since we are also interested in the adsorbed flux at the fully adsorbing areas, (1) and (2) are not applicable because $\lim_{s_i \rightarrow 1} A_i = \infty$. For this reason we use the following formulation for the received flux R :

$$R_i = \sum_j (E_j F_{ji}) + \sum_j ((1 - s_j) R_j F_{ji}), \quad (3)$$

where the relation to the adsorbed flux is

$$A_i = R_i s_i. \quad (4)$$

Rewritten in matrix notation and resolved for the vector of the received flux R we obtain

$$\mathbf{F}^T \cdot E + \text{diag}(1 - s) \mathbf{F}^T \cdot R = R, \\ (\mathbf{I} - \text{diag}(1 - s) \mathbf{F}^T) \cdot R = \mathbf{F}^T \cdot E, \quad (5)$$

with the vector of emitted flux E , a vector of sticking probabilities s , and a matrix of view factors \mathbf{F} (where F_{ij} corresponds to the view factor $i \rightarrow j$).

We approximate the solution of the resulting diagonally-dominant linear system of Eq. (5) using the Jacobi method. Each iteration of the Jacobi method can be imagined as a concurrent diffuse re-emission of each element to all other elements. The adsorbed flux A is obtained by multiplying the entries in the solution for R with the corresponding sticking probability s of the element (4). The relation $\|A\| - \|E\| = 0$, which holds for closed surfaces, can be used to test the implementation and to define a stopping criterion for the Jacobi iterations.

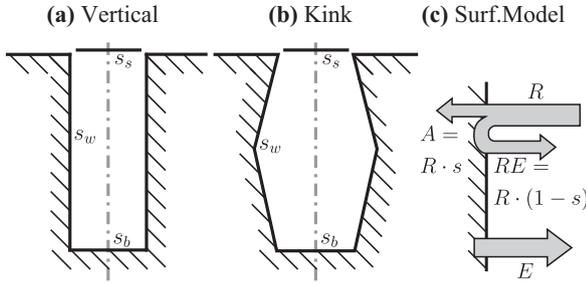


Fig. 1. Cross-sections of simulation domains with vertical walls (a) and with a kink at one half of the depth (b). s_s , s_w , and s_b designate the sticking probabilities for the source, the wall, and bottom region, respectively. (c) Illustrates the surface model showing the relation between the received flux R , the adsorbed flux A , and the emitted flux RE ; source areas emit flux E independent of the received flux R .

4. View factors

Our approach is based on a discretization of the surface into discrete surface elements along the structure's line of symmetry. Fig. 2 shows the cross-section of a convex structure and the shape of the resulting surface elements. Two vertical ranges are indicated in Fig. 2b and the resulting surface elements a and b are shown for a trench (Fig. 2a) and a hole (Fig. 2c). The elements are formed from two strips for the trench and take the form of a sliced cone for the hole.

To assemble the matrix \mathbf{F} we need to evaluate the view factors between all possible pairs of surface elements.

4.1. Trench view factors

The view factor between two segments of a symmetric convex trench with a constant cross section, as depicted in Fig. 2a, is derived using the crossed-strings method [7]. This method computes the view factor between two surfaces with a constant cross section and infinite length utilizing a two-dimensional reformulation of the problem. For two mutually completely visible strips of infinite length the view factor is [7]

$$F_{1-2} = \frac{(d_1 + d_2) - (s_1 + s_2)}{2 \cdot a_1}, \quad (6)$$

where d_1 and d_2 denote the lengths of the diagonals when connecting the cross-section of the two strips to form a convex quadrilateral, s_1 and s_2 denote the lengths of the sides of that quadrilateral which connects the strips, and a_1 denotes the length of the side of the quadrilateral which represents the emitting strip.

Fig. 3a is an isometric view of the four strips from Fig. 2a. The view factors from the top right strip a_r towards the other three

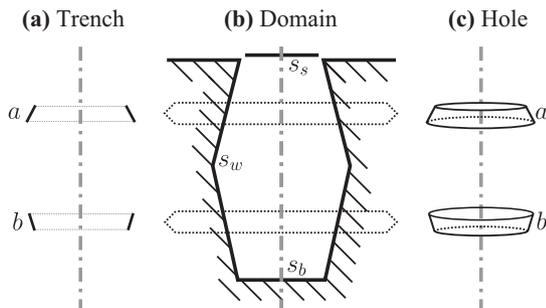


Fig. 2. Two surface elements, which result when discretizing the domain (b) are displayed: (a) is the side view of two surface elements a and b , which result from a trench discretization and (c) is the isometric view of two surface elements a and b , which result from a hole discretization.

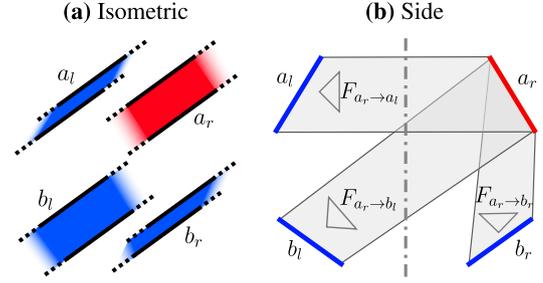


Fig. 3. Isometric (a) and side view (b) on the four infinite strips which correspond to the surface elements a and b from Fig. 2a. In (b), the view factors from the top right strip a_r towards the other three strips are visualized.

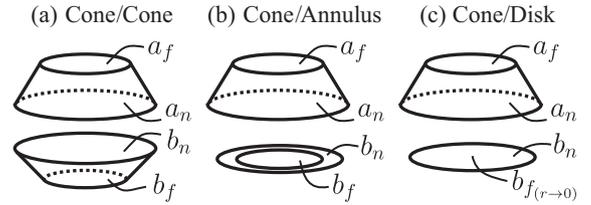


Fig. 4. Three possible pairs of segments as they result from discretizing the hole. For each pair, the near apertures a_n and b_n , and the far apertures a_f and b_f are denoted. (a) Two cone-like segments. (b) Cone and annulus. (c) Cone and disk. The far aperture is treated as an infinitely small element.

strips is visualized in Fig. 3b. The view factor between the two segments a and b is

$$F_{a-b} = F_{a_r-b_r} + F_{a_r-b_l}, \quad (7)$$

where the subscripts denote the side of the strip according to Fig. 3b. a_l can be neglected, as the cross section is symmetric. The view factor of an element to itself is

$$F_{a-a} = F_{a_r-a_l}, \quad (8)$$

where again the other direction can be neglected due to symmetry. Eq. (6) is used to compute the view factors between individual strips in (7) and (8).

4.2. Hole view factors

We derive a general formulation to compute the view factors between two segments of a rotationally symmetric convex hole as depicted in Fig. 2c. It is based on the view factor between two coaxial disks of unequal radii r_1 and r_2 at a distance z defined by

$$F_{1-2} = \frac{1}{2} \left(X - \sqrt{X^2 - 4(R_1/R_2)^2} \right), \quad (9)$$

where $R_i = r_i/z$ and $X = 1 + (1 + R_2^2)/R_1^2$ [11]. Using this relation and the reciprocity theorem of view factors

$$S_1 \cdot F_{1-2} = S_2 \cdot F_{2-1}, \quad (10)$$

where S is the element area, we derive a general formulation for the view factor between the inner wall surfaces of two coaxial cone-like segments whose surfaces are mutually completely visible. Fig. 4a shows two segments a and b in such a configuration and denotes the four coaxial disks which represent the apertures of the two elements.

In our formulation, the final goal to compute the view factor between two elements a and b is divided into multiple inexpensive view factor computations between coaxial disks. First, the difference of the view factors from b_f towards the two disks of a is computed, and the reciprocity theorem (10) is applied to obtain F_{ab_f} (red indicates sending and blue receiving areas):

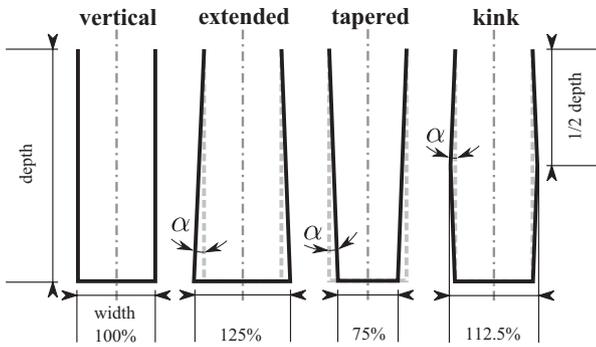


Fig. 5. Cross sections of the geometric variations of the wall for holes and trenches (shown for AR = 3). Starting from a vertical wall, the bottom width is increased by 25% (extended) and reduced by 25% (tapered). Finally, the width at 1/2 of the total depth is increased by 12.5% to form a kink. The resulting angle α , which is identical for all three variations, is depicted.

$$\widehat{F_{b_j a}} = \widehat{F_{b_j a_n}} - \widehat{F_{b_j a_f}} \Rightarrow \frac{S_{b_j}}{S_a} \cdot F_{b_j a} = \widehat{F_{a b_j}} \quad (11)$$

The same is done for b_n to obtain $F_{a b_n}$:

$$\widehat{F_{b_n a}} = \widehat{F_{b_n a_n}} - \widehat{F_{b_n a_f}} \Rightarrow \frac{S_{b_n}}{S_a} \cdot F_{b_n a} = \widehat{F_{a b_n}} \quad (12)$$

Finally F_{ab} is obtained by subtracting $F_{a b_j}$ from $F_{a b_n}$:

$$\widehat{F_{a b_n}} - \widehat{F_{a b_j}} = \widehat{F_{a b}} \quad (13)$$

The view factor of an element to itself F_{aa} is computed by subtracting the flux which leaves through the two apertures from unity:

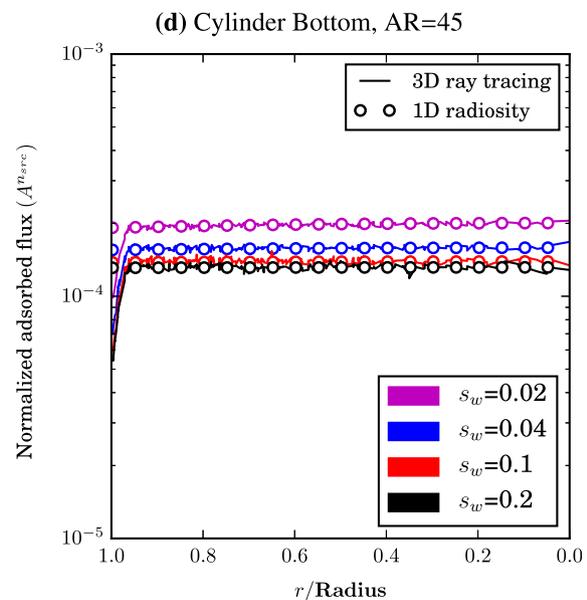
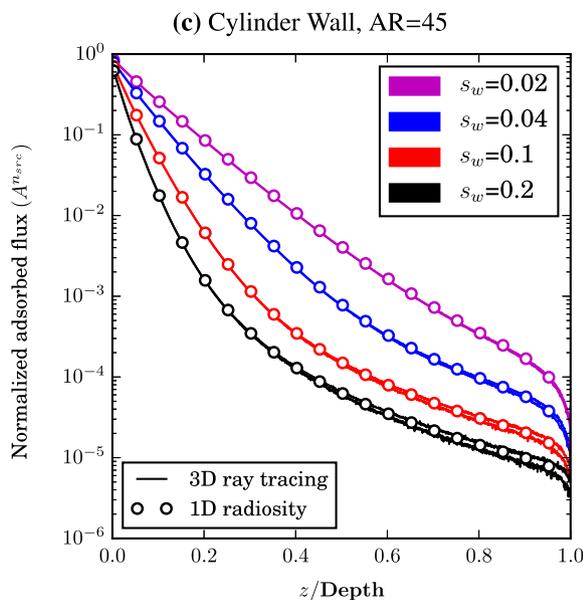
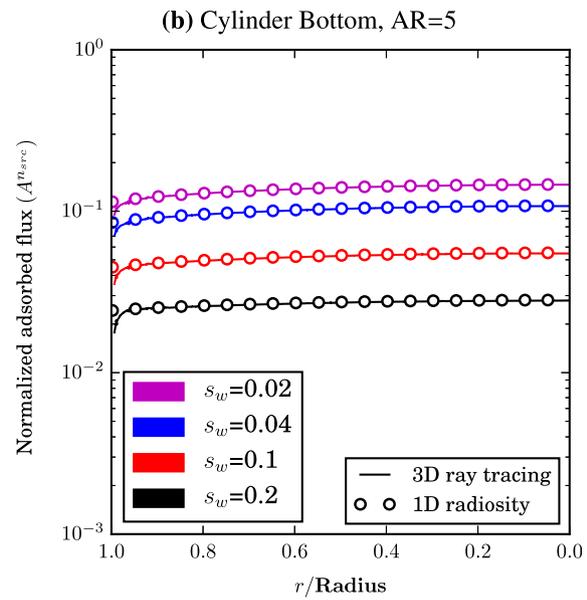
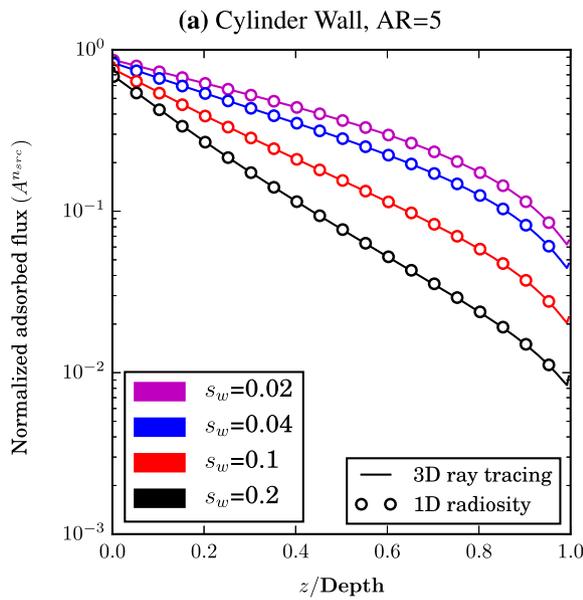


Fig. 6. Normalized flux distributions along the wall and at the bottom for cylinders of aspect ratio 5 (a, b) and aspect ratio 45 (c, d). Our one-dimensional radiosity approach (circles) is compared to a three-dimensional ray tracing simulator (lines). The sticking probability of the wall s_w is varied between 0.02 and 0.2. The deviations between ray tracing and radiosity towards the wall-bottom interface are due to the resolution of the ray tracing simulator. In (c) the ray tracing results are plotted using the minimum and maximum along the cylinder radius, particularly visible for $s_w = 0.2$.

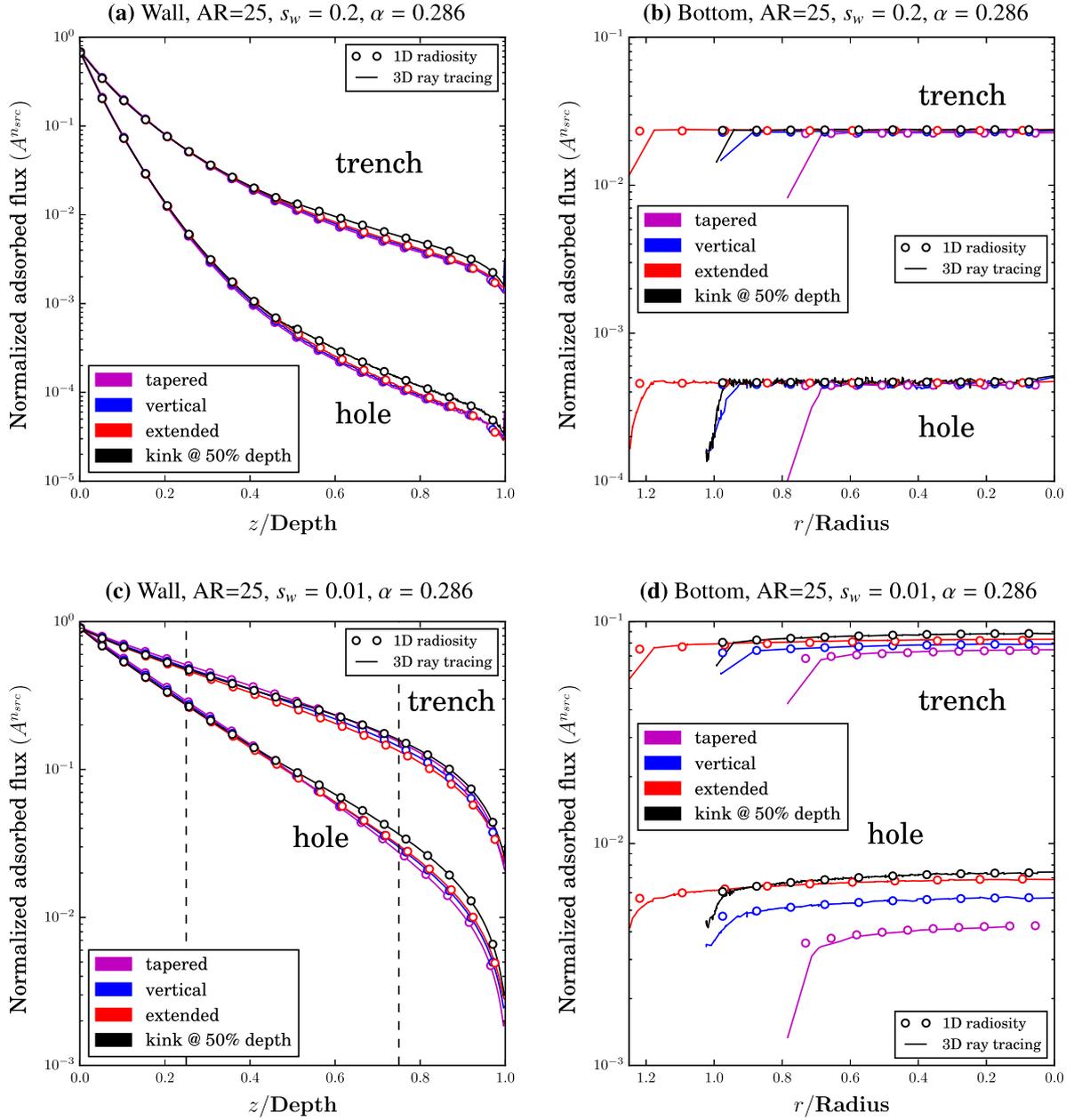


Fig. 7. Normalized flux distributions along the wall and at the bottom of a hole and a trench of $AR = 25$ for sticking probabilities $s_w = 0.2$ (a, b) and $s_w = 0.01$ (c, d). The geometry of the structures is varied (according to Fig. 5). The results for structures with vertical sidewalls are plotted as a reference. Lines represent the results of the reference ray tracing simulator [9]. The deviations between ray tracing and radiosity towards the wall-bottom interface are due to the limited grid resolution of the ray tracing simulator. The flux distributions at the bottom span the interval $[0.75, 0]$ for the tapered structures and $[1.25, 0]$ for the extended structures. In (c), the vertical dashed lines mark positions at 25% and 75% of the total depth as a reference for the results in Fig. 8.

$$F_{aa} = 1 - F_{aa_n} - F_{aa_f}. \quad (14)$$

If an element is an annulus or a disk (see Fig. 4b and c, respectively), the general formulation still applies. For a disk, the far aperture is treated as an infinitely small element.

5. Results

To provide a good qualitative comparison we normalize the results to only depend on the aspect ratio of the structure and the sticking probability. The adsorbed flux A is divided by the area of the element ($A_i^n = A_i/S_i$) and normalized to the flux which a surface of the same sticking probability would absorb, if it is fully planar-exposed to the source ($A_i^{n_{src}} = A_i^n/E_{i_{src}}^n \cdot S_i$).

The sticking probabilities of the source areas at the top s_s and the bottom of the structures s_b are modeled as fully adsorbing in all of the following results.

5.1. Validation: cylindrical holes

To evaluate the quality of our one-dimensional radiosity model, we analyze different simulation setups of cylinders, where we vary the sticking probability of the wall between $s_w = 0.02$ and $s_w = 0.2$. Fig. 6a and b compares the flux distributions for structures where $AR = 5$ obtained using the proposed one-dimensional radiosity approach with results generated with a reference Monte Carlo ray tracing tool [9]. Similarly, Fig. 6c and d compares the flux distributions for structures where $AR = 45$. The results show a good agreement, aside from the deviation at the wall/bottom interface,

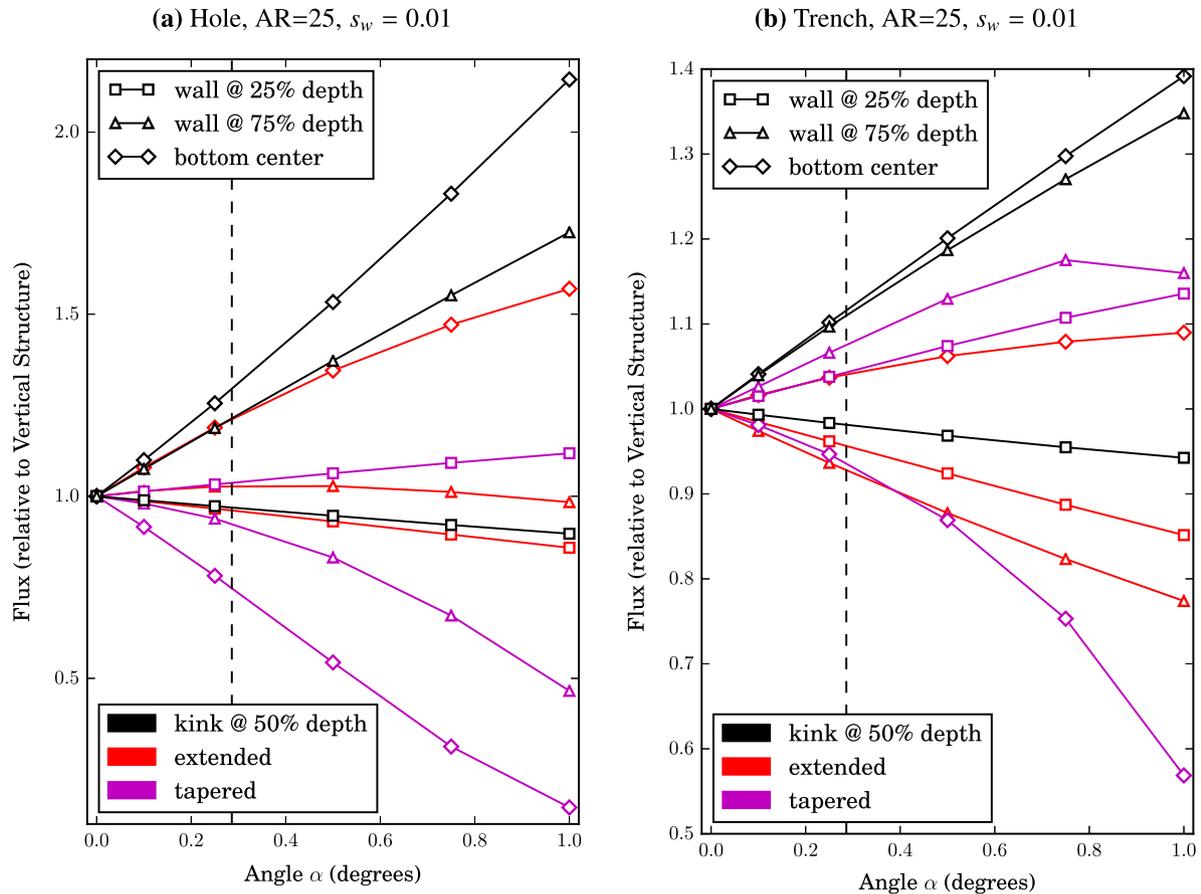


Fig. 8. Flux in a convex hole (a) and trench (b) (with tapered, extended, and kinked sidewalls) relative to the flux in a hole and a trench with vertical walls for an aspect ratio 25 and a wall sticking probability $s_w = 0.01$. The flux at the bottom center and at two points on the wall (at 25% and 75% of the total depth) is plotted over the angle α (introduced in Fig. 5). The angle α corresponds to the taper angle (tapered), the extension angle (extended), and to the angle which is formed by a kink at one half of the total depth of the structure. The vertical dashed lines indicate an angle $\alpha = 0.286^\circ$ as a visual reference to the results in Fig. 7c.

caused by the discretization which is used in the ray tracing simulation. In Fig. 6c two flux distributions are plotted for the ray tracing results along the wall; they represent the minimum and maximum along the cylinder radius. The separation of the flux distributions, particularly visible for $s_w = 0.2$ (Fig. 6c), and the visible noise in Fig. 6d, reflect the stochastic nature of the ray tracing approach.

5.2. Validation: convex structures

To validate our method for convex structures, several geometric variations including an extended, tapered, and kinked sidewall, visualized in Fig. 5, are applied to a hole and a trench of aspect ratio 25. Good agreement is achieved when comparing to the results obtained with a reference Monte Carlo ray tracing simulator.

Furthermore, the results allow to study the influence of the geometrical properties of high aspect ratio structures as well as of the particle sticking probability on the neutral particle flux. Fig. 7 compares the resulting flux distributions along the wall and at the bottom for sticking probabilities $s_w = 0.2$ and $s_w = 0.01$. For a sticking probability $s_w = 0.2$, Fig. 7a and b shows small variations along the wall and at the bottom for both, holes and trenches. Solely the presence of the kink clearly increases the flux on the bottom half of the wall.

When decreasing the sticking probability to $s_w = 0.01$, Fig. 7c indicates stronger deviations along the entire wall for all geometries. Fig. 7d reveals a variation of about $\pm 25\%$ and $\pm 10\%$ for the bottom flux in a hole and a trench, respectively.

To summarize, low sticking probabilities increase the influence of geometric variations on the flux distributions along the wall, and especially at the bottom of high aspect ratio structures.

5.3. Variation of wall geometries

Using our framework, the influence of the wall geometry on the flux distributions is studied in more detail using a hole and a trench of aspect ratio 25 with a wall sticking probability $s_w = 0.01$. For $\alpha = 0.286^\circ$, this reassembles the configuration used to produce the results in Fig. 7c and d. Fig. 8 compares the flux at the bottom center and two points on the wall (at 25% and 75% of the depth) when additionally varying α (depicted in Fig. 5) from 0° to 1° .

For the kinked structures, the angular dependence of the flux at all three positions can be approximated with a linear relation to the flux in a structure with vertical sidewalls. The results for the extended and especially the tapered structures reveal the generally non-linear relation already for small angles. Fig. 8a and b shows that the kinked sidewall leads to higher flux rates at the bottom, compared to the extended configuration; the tapered configuration reduces the bottom flux.

When interpreting the results, it must be considered that the fully adsorbing bottom area changes its size when tapering or extending the structure. For $\alpha = 1^\circ$, the bottom width/diameter is reduced to 13% and extended to 187%, for the tapered and extended structures, respectively. This is likely one reason why

the bottom flux is, at all angles, higher for the kinked structures, when comparing with the extended structures.

6. Summary and outlook

We provide an approximation of the local neutral flux in three-dimensional plasma etching simulations of high aspect ratio holes and trenches, using a one-dimensional radiosity approach. The radiosity equation is reformulated into a *receiving* perspective, which allows to model fully adsorbing surface elements. We compute all relevant view factors for holes by establishing an inexpensive general formulation for the view factor between coaxial cone-like segments. Comparing the results for various convex configurations using a rigorous three-dimensional Monte Carlo ray tracing simulation good agreement is noted and the applicability of our model for practical situations is confirmed.

We study the influence of geometric variations on the wall as well as the sticking probability on the flux distributions. The results indicate a strong influence for low sticking probabilities which are typical in IECE simulations of high aspect ratio structures. The influence is studied in more detail for holes and trenches using a sticking probability $s_w = 0.01$ and an aspect ratio of 25. The results provide a compact overview on the magnitude of the flux deviation at different positions in the structure, when comparing to the idealized shape.

Our framework is based on a computationally inexpensive and straightforwardly implementable method to compute the neutral flux distributions inside convex symmetric holes and convex symmetric trenches of constant cross-section. It can be used as a drop-in replacement for the neural flux computation during three-dimensional IECE simulations of high aspect ratio structures to significantly reduce simulation times in practical simulation

cases – or as a stand-alone tool which provides fast results for general investigations.

Acknowledgments

The financial support by the *Austrian Federal Ministry of Science, Research and Economy* and the *National Foundation for Research, Technology and Development* is gratefully acknowledged.

References

- [1] Dimitrakis P. Charge-trapping non-volatile memories. Basic and advanced devices, vol. 1. Springer; 2015.
- [2] Francis JPC, Chen F. Lecture notes on principles of plasma processing. Springer; 2003.
- [3] Ertl O, Selberherr S. Three-dimensional level set based bosch process simulations using ray tracing for flux calculation. *Microelectron Eng* 2010;87(1):20–9.
- [4] Ikeda T, Saito H, Kawai F, Hamada K, Ohmine T, Takada H, et al. Development of SF₆/O₂/Si plasma etching topography simulation model using new flux estimation method. In: Proceedings of the international conference on simulation of semiconductor processes and devices (SISPAD). p. 115–8.
- [5] Manstetten P, Filipovic L, Weinbub J, Hössinger A, Selberherr S. Using one-dimensional radiosity to model neutral particle flux in high aspect ratio holes. In: Joint international EUROSOL workshop and international conference on ultimate integration on silicon (EUROSOL-ULIS). IEEE; 2016. p. 120–3.
- [6] Kokkoris G, Boudouvis AG, Gogolides E. Integrated framework for the flux calculation of neutral species inside trenches and holes during plasma etching. *J Vac Sci Technol A* 2006;24(6):2008–20.
- [7] Modest MF. Radiative heat transfer. Academic Press; 2013.
- [8] Philip D, Kavita B, Philippe B, Peter S. Advanced global illumination. AK Peters Ltd; 2006.
- [9] Ertl O, Filipovic L, Weinbub J, Vienna TS; 2015. <<https://github.com/viennats/viennats-dev>>.
- [10] Nyström EJ. Über die praktische Auflösung von Integralgleichungen mit Anwendungen auf Randwertaufgaben. *Acta Math* 1930;54(1):185–204.
- [11] Howell JR, Menguc MP, Siegel R. Thermal radiation heat transfer. CRC Press; 2010.



Tuning the tunneling probability by mechanical stress in Schottky barrier based reconfigurable nanowire transistors



Tim Baldauf^{a,*}, André Heinzig^a, Jens Trommer^b, Thomas Mikolajick^{a,b}, Walter Michael Weber^{a,b}

^aCenter for Advancing Electronics Dresden (CFAED) and Chair of Nanoelectronic Materials, TU Dresden, Germany

^bNaMLab gGmbH, Dresden, Germany

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

Silicon nanowire
Reconfigurable logic
CMOS
RFET
SBFET
Tunneling
Schottky junction
Stress
Strain
Symmetry
Deformation potential
Self-limited oxidation
Simulation
TCAD

ABSTRACT

Mechanical stress is an established and important tool of the semiconductor industry to improve the performance of modern transistors. It is well understood for the enhancement of carrier mobility but rather unexplored for the control of the tunneling probability for injection dominated research devices based on tunneling phenomena, such as tunnel FETs, resonant tunnel FETs and reconfigurable Schottky FETs. In this work, the effect of stress on the tunneling probability and overall transistor characteristics is studied by three-dimensional device simulations in the example of reconfigurable silicon nanowire Schottky barrier transistors using two independently gated Schottky junctions. To this end, four different stress sources are investigated. The effects of mechanical stress on the average effective tunneling mass and on the multi-valley band structure applying the deformation potential theory are being considered. The transfer characteristics of strained transistors in n- and p-configuration and corresponding charge carrier tunneling are analyzed with respect to the current ratio between electron and hole conduction. For the implementation of these devices into complementary circuits, the mandatory current ratio of unity can be achieved by appropriate mechanical stress either by nanowire oxidation or the application of a stressed top layer.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Reconfigurable nanowire (NW) transistors (RFETs) provide an increased functionality of highly integrated circuits beyond classical device scaling [1–3]. With its two independently gated Si-NiSi₂ Schottky junctions (SJ) at source and drain side the RFET is able to work as either a n-FET or a p-FET device using the same physical structure as defined by a programming voltage [4] (Fig. 1). The flexible programming feature at runtime enables the synthesis of logic circuits and gates with lower transistor count and reduced critical paths compared to CMOS based circuits. For example a logic gate cell containing only six RFETs can be switched between full-swing complementary NAND and NOR functionality even giving an equal delay for both functions [5]. Compact XOR functions in a transmission gate configuration have been proposed [6] and demonstrated [7] recently with only four RFETs. The same cell with a different wiring serves as a 3

input majority (MAJ) gate. Since arithmetic operations can be realized efficiently with XOR and MAJ gates novel opportunities for circuit design and design automation arise [6]. One-bit adders with half of the device count compared to CMOS could be shown recently [8]. Nevertheless, the drain currents of unstrained n- and p-RFETs with Si-NiSi₂ junctions (barrier for electrons ~0.66 eV, for holes ~0.46 eV) are not per-se symmetric and thus do not satisfy the requirements for complementary circuit operation using a device with identical geometry as n- and p-type transistor. Work-function/ band-offset tuning at the metal contacts or through doping is a difficult task in terms of variability at the nanometer-scale. In this work we employ mechanical stress to modify the band structure of the semiconductor nanowire channel yielding an effective mechanism to precisely adjust the symmetry between n- and p-RFET without the need of doping or altering the electrode material composition [9]. Process and device simulations were carried out with Sentaurus from Synopsys (K-2015.06) to analyze the NW induced stress profiles and the resulting transfer characteristics of n- and p-programmed RFETs [10]. Effective masses are determined by band structure

* Corresponding author.

E-mail address: Tim.Baldauf@namlab.com (T. Baldauf).

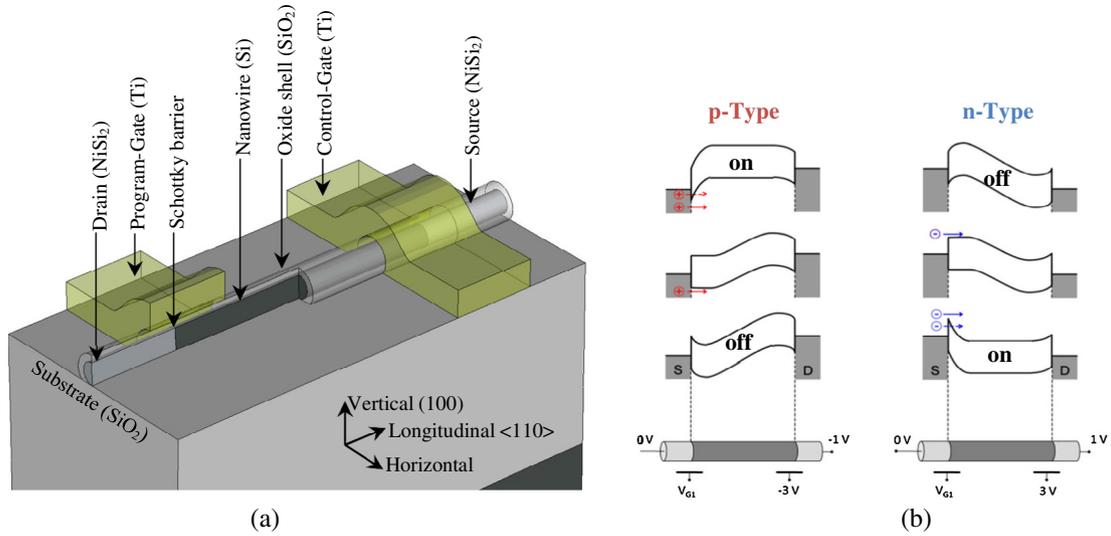


Fig. 1. Schematic view of a reconfigurable silicon NW-RFET with two independently gated Schottky junctions at source and drain, (a) RFET structure (b) operation states.

calculations based on the empirical pseudopotential method [11]. The induced mechanical stress was introduced by four different approaches as illustrated in Fig. 2. First, the self-limiting oxidation represents a reliable and uncomplicated way to form strong radial compressive stress and will be described in detail. Moreover, the simulated results could be verified by experimental data [12]. Nevertheless, a high gate oxide thickness can be detrimental to device performance, thus additional stressor vehicles are investigated: epitaxial stress from the silicidation of source/drain contacts, use of metal gate contacts with an intrinsic compressive stress and finally a tensile stressed top layer deposited on top of a geometrically optimized device.

The achieved results can also be applied to other type of devices encompassing tunneling through a barrier in the on-state, such as any Schottky FETs and in certain transport mechanisms of resonant tunneling transistors [13].

2. Mechanical stress of oxidized silicon NW

In the process simulation a 220 nm long and nearly 20 nm thick undoped silicon NW with a $\langle 110 \rangle$ channel direction and six facets (two times $\langle 100 \rangle$, four times $\langle 111 \rangle$) was oxidized at 875 °C with 10 slm O₂ capturing the experimental structure reported in Ref. [12]. Note that the oxide reaction rate for $\langle 111 \rangle$ surfaces are 30–100% higher than for $\langle 100 \rangle$ resulting in an oval NW cross section. The appearing oxide has approximately twice the thickness versus the consumed silicon leading to a strong volume expansion of the silicon oxide shell and consequently giving a radial compressive stress in the silicon NW (Fig. 3). In addition, the stress modulated reaction rate at the Si-SiO₂ interfaces and the stress dependent oxygen diffusion in the tensile oxide shell cause a self-limitation of oxidation shown by the slightly saturating gate oxide thickness with increasing time in accordance with [14]. Both stress effects as

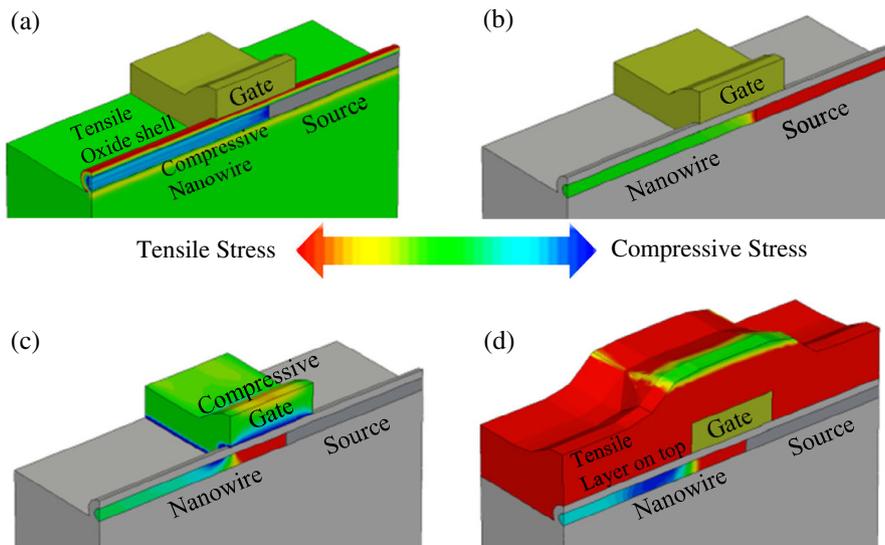


Fig. 2. Simulated stress profiles of silicon nanowire junctions induced by several stress sources: (a) compressive stress from oxide shell, (b) tensile stress from the silicidation of source/drain contacts, (c) compressively stressed metal gates and (d) tensile stressed layer on top of the device. (Stress scale is normalized to the maximum value for each case individually).

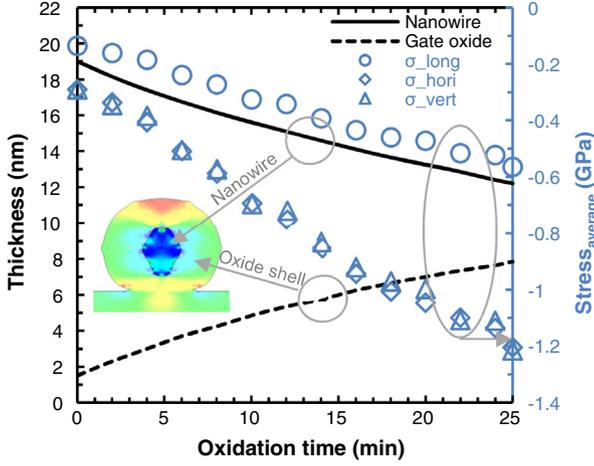


Fig. 3. Simulated average thickness of silicon NW t_{si} , gate oxide thickness t_{ox} (left) and average stress values near the Schottky junction (right) versus oxidation time (875 °C, $O_2 = 10$ slm). Initial values were $t_{si} = 19$ nm and $t_{ox} = 1.5$ nm (native oxide).

well as the surface dependent oxidation rate are considered by the process simulator. The shown stress values of the three basic directions (110) longitudinal – channel direction, (001) vertical – wafer orientation, $(1-10)$ horizontal – wafer flat) are simulated with the viscoelastic Maxwell model and are averaged over the cross-section near the Schottky interface. They increase roughly linearly with the simulated oxidation time and are in a ratio of 2:1 between the radial and the longitudinal direction.

3. Stressed Schottky barrier devices

Unstrained NW-RFETs with $Si-Ni_xSi_{1-x}$ Schottky junctions are reported to have strongly asymmetric transfer characteristics for n- and p-type operation [4,7]. The disparities mainly arise from dissimilar tunneling probabilities of electrons and holes. The simulations of an unstrained structure with 12 nm NW thickness and 8 nm gate oxide shell (after 25 min oxidation) show a ratio of 0.13 between the currents for the ntype and the ptype device (referred to as n/p ratio in the following). This is in accordance to the results reported in Ref. [4]. By tuning the mechanical stress the ratio could be enhanced to 1.20 nearly matching the experimental results with a ratio of 1.05 [12] (Fig. 4). The device simulation is based on drift-diffusion transport. Modified local-density approximation is used for quantum-mechanical confined carrier distributions occurring at Si-SiO₂ interface near the SJ. Also mobility models like high-field degradation, surface scattering and stress dependent effective transport masses are considered even though no significant influence for short RFET devices emanates from them due to the limitation of drain current by barrier tunneling. The carrier injection at the SJ is computed with a nonlocal tunneling model based on the approach presented in [15]. The included tunneling probability based on Wentzel-Kramers-Brillouin (WKB) approximation can be written for example for the electrons as

$$\Gamma_e(r) = \exp \left[-\frac{2}{\hbar} \int_0^r \sqrt{2\overline{m}_{t,e}^* \left(\frac{\Phi_{B,e} - W_{NiSi} + E_c(x)}{q} \right)} dx \right], \quad (1.1)$$

where \hbar , $\overline{m}_{t,e}^*$, $\Phi_{B,e}$, W_{NiSi} and $E_c(x)$ are the reduced Planck constant, the average effective tunneling mass as well as the barrier height for electrons, the NiSi₂ workfunction and the conduction band edge, respectively. There are two important parameters controlling the tunneling probability which are depending on the mechanical stress. One parameter is the band offset from the Fermi level of the injecting contact to the corresponding effective band edges E_c

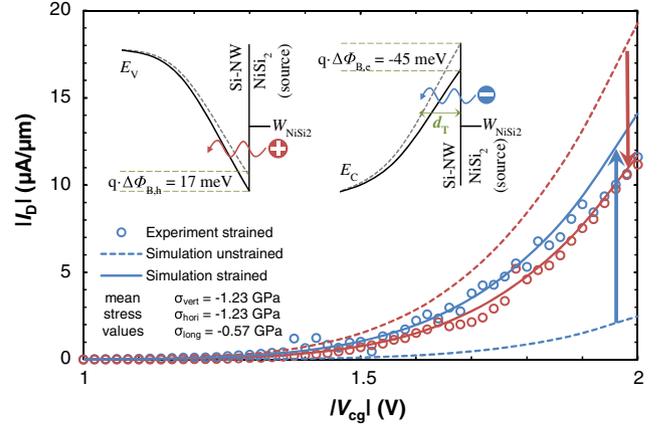


Fig. 4. Simulated and measured drain current of a nanowire device as a function of the control gate voltage V_{cg} normalized to the NW diameter. The program gate voltage V_{pg} was 2 V to achieve ntype characteristics (blue) and -2 V to achieve ptype characteristics (red). The drain voltage V_{ds} was set to 2 V for n-type and -2 V for p-type, respectively. Simulations were performed for the unstrained case (dashed) and the case including stress after 25 min oxidation (solid). The inset shows a schematic view of strained and unstrained electron barrier (left) and hole barrier (right) at the source junction. In the insets the vertical axis is the energy, the horizontal is the longitudinal distance to the metallurgic interface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and E_v . The other parameter is given by the average effective tunneling mass of electrons $\overline{m}_{t,e}^*$ and holes $\overline{m}_{t,h}^*$, respectively. The impact of both parameters is considered in the simulations.

For determining the barrier heights $\Phi_{B,e}$ and $\Phi_{B,h}$ the NiSi₂ workfunction was fixed at ~ 4.73 eV related to the vacuum level and the affinity of the corresponding band edges was calculated with the deformation potential theory of a multi-valley band structure model [16]. Non-hydrostatic mechanical strain distorts the strong symmetry of the silicon lattice and results in an energetic split between the subbands. Consequently, electrons and holes are re-distributed between lower or higher energy levels. The weighted contribution of each sub-band is lumped into a single effective band shift ΔE_c and ΔE_v according to

$$\frac{\Delta E_c}{k_B \cdot T_{300K}} = -\ln \left[\frac{1}{n_c} \cdot \sum_{i=1}^{n_c} \exp \left(\frac{-\Delta E_{c,i}}{k_B \cdot T_{300K}} \right) \right] \quad (1.2)$$

and

$$\frac{\Delta E_v}{k_B \cdot T_{300K}} = -\ln \left[\frac{1}{n_v} \cdot \sum_{i=1}^{n_v} \exp \left(\frac{\Delta E_{v,i}}{k_B \cdot T_{300K}} \right) \right]. \quad (1.3)$$

In Eqs. (1.1) and (1.2) k_B is the Boltzmann constant and n the number of sub-bands. This corresponds to an ideal Schottky barrier height, neglecting for simplicity Fermi level pinning e.g. originating from interface states such as metal induced gap states (MIGS). Note that transport and tunneling is not individually calculated for each sub-band, the simulator only considers the band edges. The application of compressive radial stress lowers E_c and thus the electron barrier and the barrier width. As a consequence the tunneling probability increases. Conversely, the hole barrier increases and the hole tunneling probability decreases (see inset of Fig. 4). This trend is consistent with band calculations with density functional theory [17]. The barrier heights for electrons and holes show a nearly linear behavior with oxidation time. This originates from the linear stress increase with oxidation time and the direct dependence of the deformation potential (Fig. 5(a)).

The effective tunneling masses of electrons and holes have an exponential and direct influence on the tunneling probability into

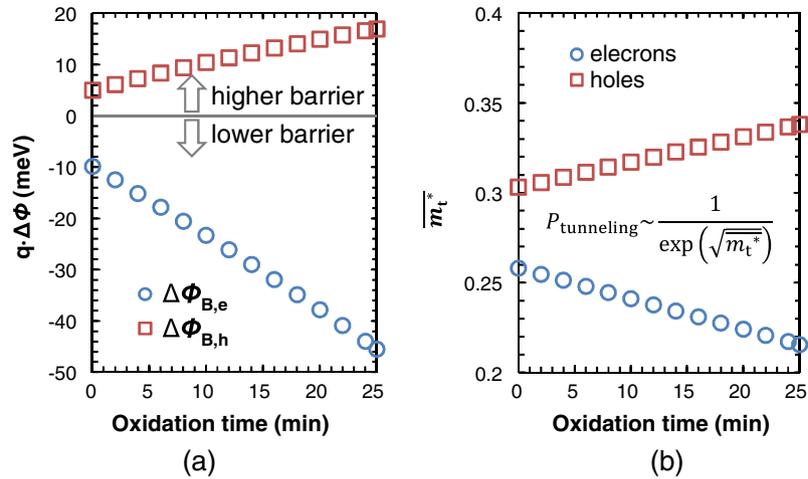


Fig. 5. Stress dependent parameters as function of oxidation time (a native oxide of 1.5 nm was assumed): (a) change of barrier height and (b) effective electron and hole mass averaged over the corresponding subbands.

or out of the band and are derived from band bending around the conduction band minimum and valence band maximum where the most carriers are located. Therefore, the band structure of the strained silicon was simulated with the empirical pseudopotential method. Mechanical stress not only modulates the energy level of the sub-bands as described by the deformation potential it also effects the band bending and thus the effective masses near the sub-band edges. However, this effect is much stronger for the valence bands than for the conduction bands. The effective tunneling masses of electrons and holes were calculated considering the harmonic average of the effective (tunneling) mass of each sub-band oriented perpendicular to the SJ while the sub-bands are weighed by their energy levels and density of states (DOS). Due to the parabolic behavior and a negligible shear stress the DOS remains quite constant for all conduction subbands. Hence, the significant stress dependent change of the average effective tunneling mass mainly arises from the energetic re-population of charge carriers similar to the behavior of the barrier height. As an example, in the case of the radial compressive stress from oxidation applied to the $\langle 110 \rangle$ channel direction the light electron sub-bands in $\langle 001 \rangle$ direction having an effective mass of $0.19 \cdot m_0$ are lowered while the heavier sub-bands in the (110) plane with an effective mass of $0.32 \cdot m_0$ are lifted. This increases the tunneling probability into or out of the $\langle 001 \rangle$ -sub-bands compared to the $\langle 100 \rangle$ - or $\langle 010 \rangle$ -sub-bands and hence decreases the average effective electron tunneling mass (Fig. 5(b)). In contrast, the heavy hole band is lifted while the light hole band is lowered resulting in an increased average effective tunneling mass for holes. Moreover, mechanical stress changes the DOS effective mass of the non-parabolic valence bands [18]. The resulting anisotropy of the effective masses is well characterized for the two main hole bands at band edge and at room temperature (26 meV). Fig. 6 shows the effective hole mass in reciprocal space using spherical coordinates to visualize the differences between an unstrained silicon NW and a compressively strained NW by a 25 min oxidation step. The DOS effective mass is calculated by the integration over all directions in k-space away from the Γ -point (similar to Eq. (3) of [16]). The DOS effective masses at the energy level around the band edges are shown in Fig. 6. Their values are comparable to the ones obtained in Ref. [18].

Using the stress-modulation from oxidation for NW RFETs the transport characteristics of the gated SB should always be considered in connection with the geometry of the device. Consequently, t_{si} and t_{ox} are determined by the oxidation time and have a direct influence on the gate control. The effect is shown in Fig. 7 by the n/p ratio of an unstrained device. A thinner gate oxide achieved

after short oxidation times allows a stronger band modulation in the NW and promotes the tunneling through the higher electron barrier. Therefore, the n/p ratio is increased. With the progressive oxidation the ratio as determined by pure geometric and electrostatic considerations decreases to 13%. This can be compensated with additional mechanical stress shown by the n/p ratio of the strained device. For this combination of a NW with 19 nm initial diameter and 1.5 nm initial native oxide we simulated a nearly ideal n/p ratio of 100% after 22 min oxidation time comparable to the experimental result of 105%. Nanowires with a thinner initial t_{si} have a smaller volume to comply with the oxide growth and thus are more stressed after the same oxidation time. To decouple the geometry from stress effects the device was simulated with and without stress dependent models to calculate a relative drain current change $\Delta I_{D, \text{strain}}$ according to:

$$\Delta I_{D, \text{strain}} = \frac{I_{D, \text{strained}} - I_{D, \text{unstrained}}}{I_{D, \text{unstrained}}} \quad (1.4)$$

for n- and p-RFET as a function of oxidation time. The stress induced current amplification of the n-type transistors behaves supra-linear and rises up to approximately 400%. In contrast, p-type RFET operation shows a degradation of up to -50% .

Both effects, barrier modulation and effective mass tuning, applied individually show comparable influences on the n/p ratio and the corresponding drain current modulation $\Delta I_{D, \text{strain}}$ (Fig. 8, for 22 min oxidized NW). The impact of the barrier height is slightly higher than the one of effective tunneling mass most likely due to the combination of the Schottky barrier height determined by the NiSi_2 workfunction and the choice of silicon crystal orientation. Thus with a different crystal orientation the stress depended modulation of the average effective tunneling mass could be weaker whereby the influence of the barrier height would become dominant. It is also visible that the combination of both effects has a higher/lower impact for the electrons/holes of the n-type/p-type than expected from the linear superposition of the effects. This means that barrier height and effective tunneling mass need to be considered at the same time as the effects on the tunneling probability amplify or diminish each-other under the influence of mechanical stress.

Furthermore, the influence of individual stress directions related to the defined NW orientation was analyzed by separating the stress components for $\langle 110 \rangle$ - as well as $\langle 100 \rangle$ -NWs, which are also interesting in relation to top-down processes. Fig. 9 shows the drain current modulation of the n- and p-type RFET for modified radial and longitudinal stress. The current of a $\langle 110 \rangle$ -NW is slightly more sensitive to the radial components than to the longitudinal one. It

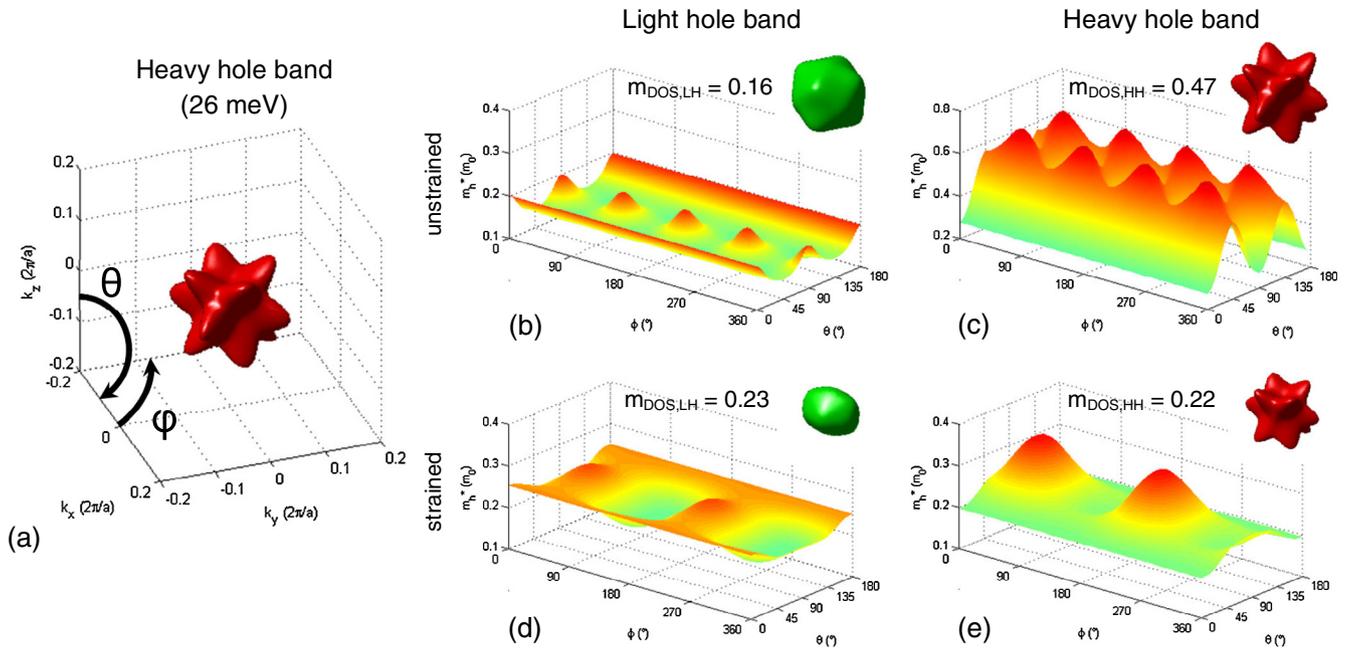


Fig. 6. Simulated valence bands and hole effective masses. (a) isoenergy surface of heavy hole band at room temperature thermal energy (26 meV). (b) and (c) show the effective masses of the heavy hole band (HH) and the light hole band (LH) at the band edge in unstrained silicon as a function of reciprocal space using spherical coordinates. (d) and (e) show the same for strained silicon (NWs oxidized for 25 min).

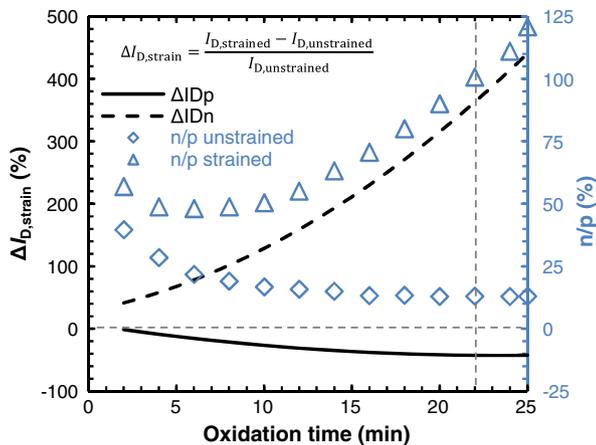


Fig. 7. Stress induced modulation of the drain current $\Delta I_{D, strain}$ (lines) and n/p ratio of strained and unstrained RFET structures (symbols) as a function of oxidation time. Refer to Fig 3 for the thicknesses of the silicon NW the gate oxide thickness and the stress values corresponding to the oxidation time.

can also be observed that the n/p ratio moves in favor of the n-RFET for less compressive longitudinal stress. In other words, a tensile longitudinal stress has the same benefit to adjust the drain currents of n- and p-configuration as compressive radial stress. However, this does not apply to $\langle 1\ 0\ 0 \rangle$ -NWs. Although the electron tunneling current behaves similar, the oxide stress increases the hole tunneling current as well. As the effects for n- and p- are not counteracting as for $\langle 1\ 1\ 0 \rangle$, the adjustment of the n/p ratio is not possible. At last, for stress engineering it is crucial to apply a stress profile to the device where sub-bands with low effective masses are energetically preferred, increasing the tunneling probability.

4. Other stress sources for silicon NWs

As mentioned above, a thick gate oxide with high radial stress reduces the gate control. Hence, other possible stress sources have

been investigated. As first alternative the epitaxial stress from the silicidation of the source and drain contacts should be considered. In our experiments it could be observed that such a thin NW structure promotes the formation of a very sharp silicon rich and cubic NiSi₂ phase to silicon interface. It is known, that the lattice constant of NiSi₂ along $\langle 1\ 1\ 1 \rangle$ is slightly smaller than that of silicon and hence an intrinsic tensile stress in the silicide region is expected [19]. Results on the exact stress values are currently not available from experiments. Therefore, an estimated intrinsic tensile stress of 500 MPa was assumed. In the process simulation it could be seen that this stress expands only through the first 5 nm distance within the silicon NW beyond the metallurgical junction and relaxes strongly which results in relatively low average stress values within the tunneling region (Table 1). Thus, no large changes in effective tunneling mass and band deformation occur leading to a slightly increased n/p ratio of 19% for a comparable structure with 12 nm NW thickness and 8 nm gate oxide shell (after 25 min oxidation).

A further source of mechanical stress could be the top lying and omega shaped gate contact. Depending on the process conditions a tensile or compressive stress can be imprinted [20]. Both variants, with +1.0 GPa and -1.0 GPa intrinsic gate stress, have been examined. With the compressive gate material we could achieve an improved n/p ratio of 31%. The enhanced ratio results from the increased effective tunneling mass of holes which only degrades the p-type drain current and thus the device performance.

The most promising alternative is a stressed top layer above the complete gate stack structure. Silicon-nitride (Si₃N₄) is typically used for stress enhanced planar MOSFETs in industrial production and can be deposited either with a compressive intrinsic stress up to 3.5 GPa or a tensile intrinsic stress up to 1.7 GPa [21,22]. That makes this stress source universally applicable for use in other nano-scale transistor types having an energy barrier for tunneling in the on-state and also for application to different crystal orientations. First the effect of a tensile stressed top layer was analyzed for the NW structure oxidized for 25 min. This already gave an improved the n/p ratio of 25%. Further stress penetration into the

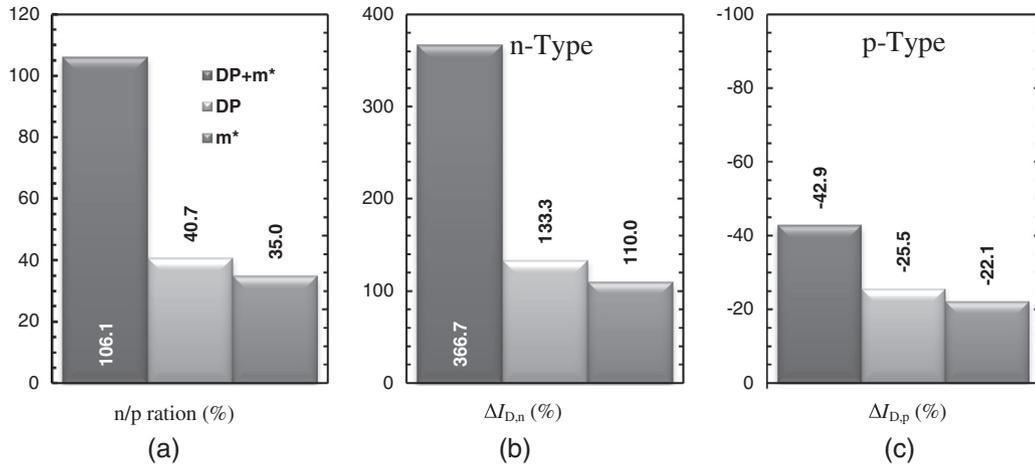


Fig. 8. (a) n/p ratio and drain current modulation ($\Delta I_{D, \text{strain}}$) for (b) n-type and (c) p-type devices caused by a 22 min oxidation in a RFET structures. The three bars show the dependency of the stress modulated effective tunneling mass only (m^*), the deformation potential theory only (DP) and the combination of both methods (DP + m^*).

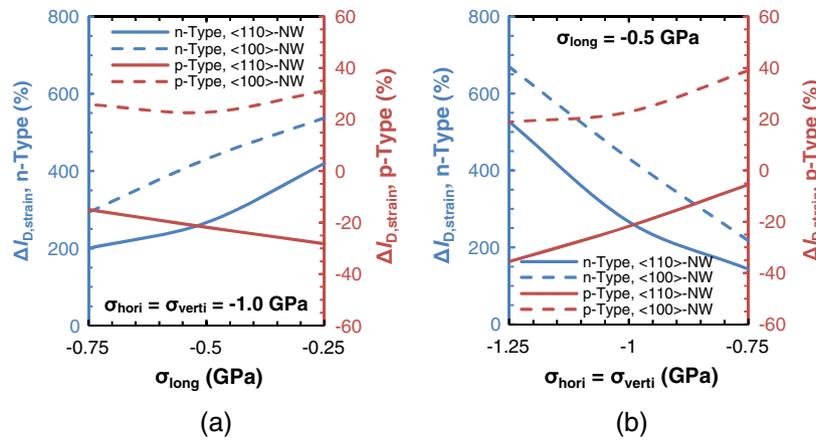


Fig. 9. Drain current modulation $\Delta I_{D, \text{strain}}$ of n- and p-type RFET depending on individual stress directions for $\langle 110 \rangle$ - and $\langle 100 \rangle$ -NWs. Base stress profile was given by $\sigma_{\text{hori}} = \sigma_{\text{verti}} = -1.0$ GPa and $\sigma_{\text{long}} = -0.5$ GPa (similar to stress from oxidation). (a) Variation of longitudinal stress with constant radial stress. (b) Variation of radial stress with constant longitudinal stress.

Table 1

Overview of the simulated imprinted stress values, stress dependent effective masses and change of barrier height as well as n/p ratio (unstrained **0.13**) for the four investigated stress sources.

	25 min oxidation (compressive)	NiSi ₂ (tensile)	Metal-Gate (compressive)	Overlayer (tensile)	Overlayer (tensile) with optimized structure
σ_{long} (GPa)	-0.57	0.12	0.59	0.37	0.59
σ_{hori} (GPa)	-1.23	-0.09	-0.24	0.08	0.03
σ_{vert} (GPa)	-1.23	-0.08	0.41	-0.30	-0.31
$m_{t,e}^*/m_{t,h}^*$	0.22/0.34	0.25/0.31	0.28/0.52	0.21/0.27	0.21/0.30
$q \cdot \Delta \Phi_{B,e} / q \cdot \Delta \Phi_{B,h}$ (meV)	-45/17	0/-1	8/-13	-8/-8	-11/-12
n/p	1.20	0.19	0.31	0.25	1.09

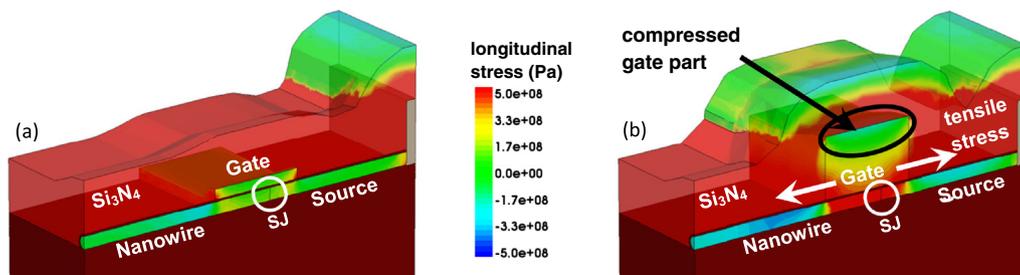


Fig. 10. 3D slice of the longitudinal stress profile of a gated SJ with tensile stressed (1.7 GPa) Si₃N₄ top layer: (a) 10 nm thick metal gate region. Compressive stress from the interface between top layer and gate compensates the tensile stress originating from the sides of the gate resulting in an unstrained SJ. (b) 60 nm thick metal gate region. The compressive stress relaxes over the gate thickness and the SJ is mainly tensile strained in longitudinal direction. The effect is explored in the case of the NW RFET.

nanowire was improved by choosing a thinner gate oxide thickness of only 2 nm. However, it was also necessary to adjust the thickness of the gate contact lying between the stressed top layer and the SJ because the gate material directly at the interface to the tensile silicon-nitride becomes compressively stressed. This eliminates the tensile stress originating from the sides of the gate region. Thus, for specific gate thicknesses the stress above the SJ can vanish (Fig. 10). This calls for a thicker gate layer to relax the compressive component of the top side interface and to develop the tensile portion over the SJ. For the improved RFET structure with the tensile stressed top layer we observed a n/p ratio of 1.09%.

5. Conclusion

Mechanical stress provides a flexible and dopant-free method to adjust drain currents in energy barrier based transistors. For reconfigurable FETs it allows to precisely tune the ratio between the current of the n and the p-type operation which is an important requirement to realize complementary circuits and to fully exploit fine-grain reconfigurability of circuits. A stress profile generated by thermal oxidation of the silicon NW was used to describe the stress dependent multi-valley band structure and the average effective tunneling mass as primary modifiable parameter of the current injection through the Schottky junctions. Furthermore, we could show the individual influence of both parameters as well as the amplifying impact of their combination. Moreover, the influence of different stress directions on silicon NW structures with $\langle 100 \rangle$ - and $\langle 110 \rangle$ -orientation was investigated. With the oxidized NW RFET we achieved a symmetric transfer characteristic between n- and pRFET, highly matching previous reported experimental results. As alternative to a thick gate oxide with inherent disadvantages for the electrostatics of the device we further examined the mechanical stress formed by the silicidation of the source and drain contacts, intrinsically stressed metal-gates and the application of a stressed silicon-nitride overlayer. The tensile overlay in combination with an improved structure seems to be a promising approach for current VLSI technologies providing the applicability of this method in conventional top-down process flows. In summary, mechanical stress is a versatile tool to tune electron and hole conduction in devices based on different barrier tunneling mechanisms such as RFETs with potential use in resonant tunneling transistors and band-to-band tunneling devices (TFETs) [23].

Acknowledgment

Parts of this work are supported by “Deutsche Forschungsgemeinschaft (DFG)” in the project ReProNano (MI 1247/6-2 and WE 4853/1-2) and the Cluster of Excellence ‘CfAED’.

References

- [1] Lee CW, Yu CG, Park JT, Colinge JP. Device design guidelines for nano-scale MuGFETs. *Solid-State Electron* 2007;51(3):505–10.
- [2] Weber WM, Heinzig A, Trommer J, Martin D, Grube M, Mikolajick T. Reconfigurable nanowire electronics—a review. *Solid-State Electron* 2014;102:12–24.
- [3] Zhang J, Tang X, Gaillardon PE, De Micheli G. Configurable circuits featuring dual-threshold-voltage design with three-independent-gate silicon nanowire FETs. *IEEE Trans Circ Syst I* 2014;61(10):2851–61.
- [4] Heinzig A, Slesazek S, Kreupl F, Mikolajick T, Weber WM. Reconfigurable silicon nanowire transistors. *Nano Lett* 2012;12(1):119–24.
- [5] Trommer J, Heinzig A, Slesazek S, Mikolajick T, Weber WM. Elementary aspects for circuit implementation of reconfigurable nanowire transistors. *IEEE Electron Device Lett* 2014;35(1):141–3.
- [6] Amarú L, Gaillardon PE, De Micheli G. Biconditional binary decision diagrams: a novel canonical logic representation form. *IEEE J Emerg Select Topics Circ Syst* 2014;4(4):487–500.
- [7] De Marchi M, Sacchetto D, Zhang J, Frache S, Gaillardon PE, Leblebici Y, De Micheli G. Top-down fabrication of gate-all-around vertically stacked silicon nanowire FETs with controllable polarity. *IEEE Trans Nanotechnol* 2014;13(6):1029–38.
- [8] Trommer J, Heinzig A, Baldauf T, Slesazek S, Mikolajick T, Weber WM. Functionality-enhanced logic gate design enabled by symmetrical reconfigurable silicon nanowire transistors. *IEEE Trans Nanotechnol* 2015;14(4):689–98.
- [9] Baldauf T, Heinzig A, Trommer J, Mikolajick T, Weber WM. Stress-dependent performance optimization of reconfigurable silicon nanowire transistors. *IEEE Electron Device Lett* 2015;36(10):991–3.
- [10] Synopsys. Sentaurus™, User Guide. Version K-2015.06.
- [11] Sverdlov V. Strain-induced effects in advanced MOSFETs. Springer Science & Business Media; 2011.
- [12] Heinzig A, Mikolajick T, Trommer J, Grimm D, Weber WM. Dually active silicon nanowire transistors and circuits with equal electron and hole transport. *Nano Lett* 2013;13(9):4176–81.
- [13] Seabaugh AC, Zhang Q. Low-voltage tunnel transistors for beyond CMOS logic. *Proc IEEE* 2010;98(12):2095–110.
- [14] Liu H, Biegelsen DK, Ponce FA, Johnson NM, Pease RFW. Self-limiting oxidation for fabricating sub-5 nm silicon nanowires. *Appl Phys Lett* 1994;64(11):1383–5.
- [15] Jeong M, Solomon PM, Laux SE, Wong HS, Chidambarrao D. Comparison of raised and Schottky source/drain MOSFETs using a novel tunneling contact model. In: *IEEE International Electron Devices Meeting, IEDM Technical Digest*. pp. 733–736.
- [16] Fischetti MV, Laux SE. Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys. *J Appl Phys* 1996;80(4):2234–52.
- [17] Niquet YM, Delerue C, Krzeminski C. Effects of strain on the carrier mobility in silicon nanowires. *Nano Lett* 2012;12(7):3545–50.
- [18] Guillaume T, Mouis M. Calculations of hole mass in [110]-uniaxially strained silicon for the stress-engineering of p-MOS transistors. *Solid-State Electron* 2006;50(4):701–8.
- [19] Steegen A, Maex K. Silicide-induced stress in Si: origin and consequences for MOS technologies. *Mater Sci Eng: R: Rep* 2002;38(1):1–53.
- [20] Kang CY, Choi R, Song SC, Choi K, Ju BS, Hussain MM, Kirsch P. A Novel Electrode-Induced Strain Engineering for High Performance SOI FinFET utilizing Si Channel for Both N and PMOSFETs. *IEEE International Electron Devices Meeting, IEDM* 2006:1–4.
- [21] Flachowsky S, Wei A, Herrmann T, Illgen R, Horstmann M, Richter R, Stenzel R. Gate length scaling trends of drive current enhancement in CMOSFETs with dual stress overlayers and embedded-SiGe. *Mater Sci Eng, B* 2008;154:98–101.
- [22] Thompson SE, Sun G, Choi YS, Nishida T. Uniaxial-process-induced strained-Si: extending the CMOS roadmap. *IEEE Trans Electron Devices* 2006;53(5):1010–20.
- [23] Zhao QT, Hartmann JM, Mantl S. An improved Si tunnel field effect transistor with a buried strained source. *IEEE Electron Device Lett* 2011;32(11):1480–2.



Reconfigurable field effect transistor for advanced CMOS: Advantages and limitations



C. Navarro*, S. Barraud, S. Martinie, J. Lacord, M.-A. Jaud, M. Vinet

CEA-LETI, Minatec Campus, F-38054 Grenoble, France

ARTICLE INFO

Article history:

Available online 21 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

FDSOI
Reconfigurable FET
Polarity gate
Reversible FET
RFET
Schottky barrier

ABSTRACT

Reconfigurable FETs (RFETs) are optimized in planar Fully Depleted (FD) SOI. Their basics, electrostatics and performance are studied and compared with standard 28 nm FDSOI and other RFETs results in the literature. The main challenge for future broad adoption is analyzed and commented. Finally, some tips to improve the performance such as the asymmetric silicidation at source/drain are discussed.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

For more than 40 years CMOS technology has followed an incessant scaling, i.e. transistors shrinking, tracking the Moore's Law [1]. However, more recently, this trend has been slowed and is currently threatened for sub-10 nm [2]. The industry is thus researching new device structures for future electronics that enable continued CMOS scaling. Many different competitors have been proposed: Trigate FETs [3], Tunneling FETs [4] or stacked nanowires [5] for example. Reconfigurable FETs, RFETs [6], stand as an interesting option to reduce the number of devices in future circuits thanks to their reprogrammable operation.

RFETs feature metallic (typically nickel silicides as NiSi [6]) source and drain (S/D) regions, where both type of carriers are able to move freely. Together with a control gate (CG), responsible of handling the current flow as in MOSFETs, one or more additional polarity gates (PG) are employed to deal with the S/D-body Schottky barriers (SB). The goal of these PGs is to suppress the ambipolar current in SB MOSFETs during the OFF state. This configuration allows to in-situ switch from N to P-like FETs by selecting which carrier is injected by tunneling. The RFET main advantages are: (i) no doping required for S/D regions (no random dopant fluctuations), (ii) fewer fabrication steps and lower thermal budget (no S/D implantation, epitaxy or dopant activation), (iii) reversible operation and (iv) possible reduction in the number of devices for

similar logic functions. On the other hand, the presence of the PGs make RFETs to present larger layout designs (up to 3 times bigger), more complex routing and increased capacitances. Furthermore, the signaling associated to these polarity gates requires extra circuitry in the design. Finally, the silicide-silicon junction quality needs to be well-controlled to avoid defects leading to fermi-level pinning [7], thus to unexpected silicide effective workfunctions and current asymmetries between N/P RFETs.

In this paper we optimized different planar Fully Depleted (FD) SOI RFET structures (Fig. 1a–c), we select the best among them and we realized a native benchmark with FDSOI MOSFETs (Fig. 1d) using the same 28 nm technology. The following section describes the basics in the operation of reconfigurable FETs, in particular, for the three top-gates (3G) structure. In Section 3, we discuss about the contrasts between the different RFET structures in Fig. 1a–c. We also document the reason of choosing the 3G-RFET as the best candidate to compare with typical FDSOI MOSFETs. Section 4 is dedicated to the native benchmark with respect to FDSOI MOSFET in terms of isolated device and logic inverter performance. In Section 5, we compare our results with those available in the literature and comment on the differences between them. Some tips and ideas to improve the RFET performance are given in Section 6. Finally, the last section is devoted to the conclusions.

2. Reconfigurable FET basics

The metallic S/D regions ensure the availability of both holes and electrons when required. The current in RFETs, as in SB

* Corresponding author.

E-mail address: carlosnm@ugr.es (C. Navarro).

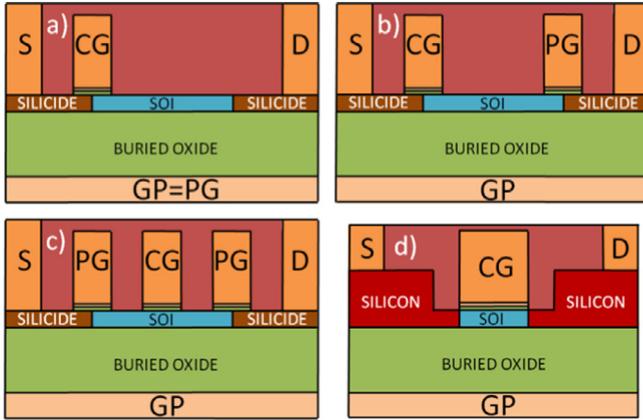


Fig. 1. (a) 1-Gate, (b) 2-Gates and (c) 3-Gates planar RFET structures compared to (d) 28 nm FDSOI featuring same gate-stack and films thicknesses. S/D regions are metallic (silicides) in RFETs while highly doped Silicon is used in FDSOI. CG = control gate, PG = polarity gate and GP = ground plane.

MOSFETs [8,9], is due to both thermionic and field emission, that is, the contribution of carriers with sufficient energy to surmount the Schottky barrier or able to tunnel through it, respectively. Since the S/D metal workfunction is not exactly aligned with the Silicon energy of conduction (valence) bands, the injection of electrons (holes) in the channel by thermionic emission is very limited. Thus, RFETs require the modulation of the SB thickness to enhance the current by tunneling, especially at the source side.

In SB MOSFETs the S/D workfunction is clearly located closer to one of the Silicon conduction or valence energy bands, hence favoring the injection of one of the carriers over the other. However, since N/P RFETs are identical (no individual aspect ratio or conduction booster may be applied) they must feature symmetric N/P current by default. Thus, the electron/hole currents need to be adjusted by tuning the conduction properties and especially the electron and hole SB height. Typically, this implies the use of S/D workfunctions close to the Silicon mid-gap ($\Phi_{S/D} \approx \chi_{Si} + E_g/2 \approx 4.61$ eV). The different effective masses and carriers mobilities between electron and holes may motivate slight workfunction deviations from mid-gap to achieve full current symmetry. Setting the silicides workfunction close to mid-gap will be finally responsible of the very low RFET output currents and limited performance as will be discussed later.

In case of the 3G-RFET (Fig. 1c), the modulation of the Schottky barriers is managed by the two lateral PGs, as seen in Fig. 2 for low drain voltage ($V_{DS} \approx 0$ V). As observed, the energy barrier

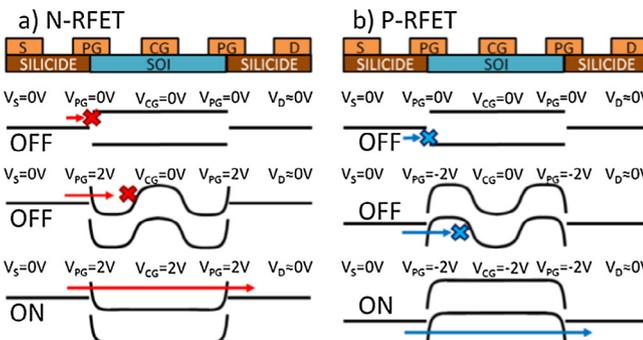


Fig. 2. Simplified front-channel energy bands diagram for the 3G RFET at $V_{DS} \approx 0$ V. (a) N-type ($V_{PG} = +2$ V) and (b) P-type operation ($V_{PG} = -2$ V) in OFF-state and ON-state. The PGs select the carrier that is allowed to tunnel while the CG controls the current flow (ON/OFF) from source to drain as in typical MOSFETs.

thickness, related to the carrier tunneling transmission, can be controlled thanks to the PG voltage applied with respect to S/D. Positives PG biases enable electrons to tunnel from the metallic source to the conduction band (Fig. 2a) while negative voltages do the same for holes toward the valence band (Fig. 2b). This carrier injection is more efficient closer to the front-channel where the electrostatic control induced by the polarity gates permits a sharper energy band bending. Meanwhile, the control gate (CG) is responsible of regulating the current flow by inducing (OFF state) or not (ON state) an energy barrier in the middle of the channel. The operation of the two gate (2G) RFET is similar but the CG is also responsible of governing one of the Schottky barriers which may increase the OFF-state current. In planar RFETs, the use of the ground plane (GP) is only mandatory in case of the one top-gate (1G) RFET structure, where it may act as PG or CG depending on the role of the top-gate.

3. Planar FDSOI multi-gate RFETs

All three RFET structures from Fig. 1a–c were at first simulated using Synopsys TCAD [10] in 2D to discern the best aspirant to compare with FDSOI MOSFETs. The Wentzel Kramers Brillouin (WKB) model was employed as an approximation for the quantum tunneling. This model has proven to be consistent with experimental results as shown in [6]. RFET structures feature 25 nm BOX thickness, 10 nm SOI layer and equivalent oxide thickness of 1.4 nm. All top gates length and their spacing are fixed to 30 nm. The top-gates overlapping with the silicide is set to 15 nm. The final silicon film length depends on the number of top-gates, being 45/60/120 nm for the 1G/2G/3G planar RFETs, respectively. All top-gates and S/D workfunctions are initially settled to align at Silicon mid-gap (4.61 eV). The PG are fixed to ± 2 V to establish the N or P behavior. A constant $V_{DS} = \pm 1$ V is applied while the control gate is swept from -2 to $+2$ V. The GP remains always grounded for the 2G and 3G-RFETs but for the 1G-RFET it operates as PG. The width is fixed to 1 μm in all structures. Reducing the width in planar RFETs yields lower currents due to the lower area at the SB junction where tunneling occurs. In a real circuit, the width scaling may also impact the design since the routing becomes more complex.

The comparison of the transfer characteristics is illustrated in Fig. 3. As expected, the devices operate either in N or P mode depending on the PG biasing conditions. The main electrostatic results are summarized in Table 1. The N-RFET ON current is systematically higher than in P-RFET. The lower hole effective mass ($m_n^* = 0.3 m_0$ and $m_p^* = 0.2 m_0$, chosen as in [11]) does not compensate the lower hole mobility when the SB heights are symmetric (mid-gap), thus larger silicides workfunction (>4.61 eV) are required to achieve N/P current symmetry. Three main contrasts

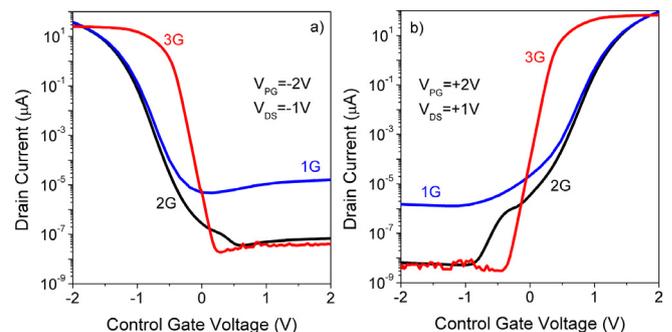


Fig. 3. Simulated $I_D(V_{CG})$ transfer characteristics for different RFET structures with 1,2 and 3 top-gates (Fig. 1a–c). $V_{PG} = \pm 2$ V, $V_{DS} = \pm 1$ V, $V_{BG} = 0$ V (2G and 3G), $V_{BG} = V_{PG}$ (1G). $\Phi_{S/D} = \Phi_{CG/PG} = 4.61$ eV. $L_{CG/PG} = 30$ nm, $EOT = 1.4$ nm, $T_{SOI} = 10$ nm and $T_{BOX} = 25$ nm. $W = 1$ μm .

Table 1

Preliminary simulations of planar RFET for different number of top gates. $V_{BG} = 0\text{ V}$ (2G and 3G), and $V_{DS} = \pm 1\text{ V}$. $V_{PG} = +2\text{ V}$ (N-RFET) and $V_{PG} = -2\text{ V}$ (P-RFET). $V_T = V_{CG}$ ($I_{DS} = 0.1 \cdot W/L_{SJ} \mu\text{A}/\mu\text{m}$), $I_{ON} = I_{DS}$ ($V_{CG} = V_T \pm 0.65\text{ V}$) and $I_{OFF} = I_{DS}$ ($V_{CG} = V_T \mp 0.35\text{ V}$). SS extracted at $0.01 \cdot I_{DS}$ ($V_{CG} = V_T$). DIBL calculated as $\Delta V_T/\Delta V_{DS}$ for $V_{DS} = 1\text{ V}$ and 0.05 V . $\Phi_{S/D} = \Phi_{CG/PG} = 4.61\text{ eV}$, $L_{CG/PG} = 30\text{ nm}$, $EOT \approx 1.4\text{ nm}$, $T_{SOI} = 10\text{ nm}$ and $T_{BOX} = 25\text{ nm}$.

		SS mV/dec	DIBL mV/V	$I_{ON} \mu\text{A}/\mu\text{m}$	$I_{OFF} \text{nA}/\mu\text{m}$	$I_{ON}/I_{OFF} \cdot 10^5 \text{ A/A}$
1G	P	145.4	–	24.3	12.4	0.016
	N	163.7	–	41.0	10.6	0.039
2G	P	142.9	–	25.7	10.0	0.025
	N	154.5	–	46.6	8.4	0.055
3G	P	79.2	37.5	18.6	0.009	1.91
	N	78.5	39.6	42.5	0.005	8.96

can be discerned between the RFETs architectures: (i) the OFF state current, is higher in the 1G-RFET. The CG is not able to adequately suppress the back-channel tunneling induced by the PG (ground plane). (ii) The maximum current is slightly lower for the 3G due to the longer structure. (iii) The electrostatic control is boosted in the 3G structure being the unique device with sub 100 mV/dec SS (Subthreshold Swing) and larger I_{ON}/I_{OFF} ratio. On the other hand, the 3G architecture presents the larger layout design and more complex routing among all architectures considered.

Based on the electrostatic and current ratio from the preliminary results in Table 1, the 3G-RFET was finally selected as the structure to be further optimized and benchmarked against 28 nm FDSOI MOSFETs. The gates length, gates spacing, silicon film thickness, PG overlapping with S/D and back-gate bias, among other parameters, were tuned to see the impact on the electrostatics. For example the influence of the control gate length, L_{CG} , in the current characteristic is shown in Fig. 4a for a N-type 3G-RFET. As the CG is shortened, the electrostatic control over the channel is gradually lost which is reflected by the increase in the SS. On the other hand, making the CG larger slightly improves the SS but neg-

atively affects the ON current simply because the device is longer and the lateral electric field decreases. The influence of the spacing between the PG at the source side and the CG is also illustrated, Fig. 4b. Longer spacers enable the development of an undesired energy barrier throughout the uncover gate region between the PG and the CG. This additional resistance significantly reduces the current. In fact, the 3G-RFET can be considered as a tunneling Schottky Barrier, which is the main physical mechanism limiting the carrier transport, in series with channel resistance plus spacer resistance.

The final optimized 3G RFET (Fig. 5a) presents $\Phi_{S/D} = 4.66\text{ eV}$ to achieve N/P current symmetry, mid-gap workfunction top-gates $\Phi_{CG/PG} = 4.61\text{ eV}$, N-type GP, $EOT = 1.55\text{ nm}$, $T_{SOI} = 6.5\text{ nm}$, and $T_{BOX} = 25\text{ nm}$. A $1\text{ }\mu\text{m}$ silicon substrate, $N_D = 2 \cdot 10^{18}\text{ cm}^{-3}$, is considered. The gate length is reduced down to $L_G = 20\text{ nm}$ while their spacing is shortened to $L_{SP} = 15\text{ nm}$. The overlapping of PGs with the body decreases to 10 nm shrinking the 3G-RFET to $L_{SOI} = 70\text{ nm}$. The carriers mobility fits experimental 28 nm FDSOI results at similar gate length [12] (constant mobility model with $227/60\text{ cm}^2\text{ V}^{-1}\text{ s}^{-1}$ for electrons/holes).

On its side, the FDSOI MOSFET structure (Fig. 5b) used to benchmark the 3G-RFETs exhibits the same vertical film architecture but a 15 nm epitaxy is carried out to raise the S/D regions and reduce the series resistance. The spacing between the CG and the doped S/D regions is fixed to $L_{SP} = 9\text{ nm}$. Two control gate lengths are probed, 27 nm and 60 nm , adjusting the constant carrier mobility for each length independently with experimental 28 nm FDSOI results [12] as for the 3G-RFET.

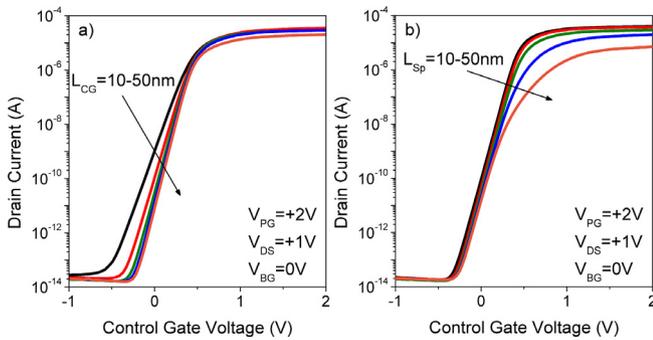


Fig. 4. $I_D(V_{CG})$ curves for different (a) control gate length and (b) spacing between the polarity gate at the source side and the control gate (the other spacer at drain side remains fixed at 30 nm) for a N-type 3G-RFET. $W = 1\text{ }\mu\text{m}$.

4. Optimized 3G-RFET vs. FDSOI

The comparison of the $I_D(V_{CG})$ curves between the optimized 3G reconfigurable FET and the FDSOI MOSFETs is depicted in Fig. 6. Table 2 shows the main electrostatic and current parameters at minimum and equivalent effective gate length for the 28 nm FDSOI. It can be observed how the RFET presents better electro-

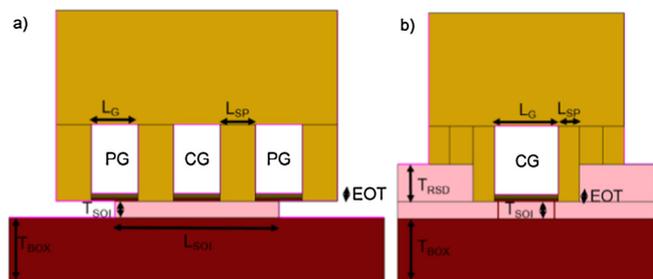


Fig. 5. (a) Optimized 3G-RFET and (b) 27 nm length FDSOI 2D structures used during TCAD simulations. The $1\text{ }\mu\text{m}$ silicon substrate is not represented.

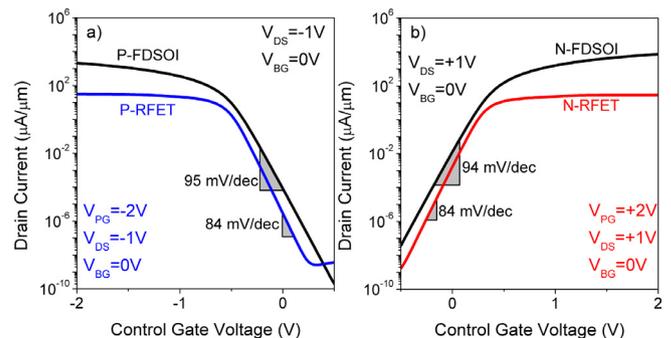


Fig. 6. (a) P- and (b) N-type current comparison between optimized 3G-RFET and FDSOI devices. RFET $L_{SOI} = 70\text{ nm}$. FDSOI $L_{SOI} = L_G = 27\text{ nm}$. $V_{BG} = 0\text{ V}$. $W = 1\text{ }\mu\text{m}$.

Table 2
Optimized 3G-RFET vs. FDSOI benchmark in planar 28-FDSOI. $V_{BG} = 0$ V and $V_{DS} = \pm 1$ V. $V_{PG} = +2$ V (N-RFET) and $V_{PG} = -2$ V (P-RFET). $V_T = V_{CG}$ ($I_{DS} = 0.1 \cdot W/L_{Si}$ $\mu\text{A}/\mu\text{m}$), $I_{ON} = I_{DS}$ ($V_{CG} = V_T \pm 0.65$ V) and $I_{OFF} = I_{DS}$ ($V_{CG} = V_T \mp 0.35$ V). SS extracted at $0.01 \cdot I_{DS}$ ($V_{CG} = V_T$). DIBL calculated as $\Delta V_T/\Delta V_{DS}$ for $V_{DS} = 1$ V and 0.05 V.

		SS mV/dec	DIBL mV/V	I_{ON} $\mu\text{A}/\mu\text{m}$	I_{OFF} nA/ μm	$I_{ON}/I_{OFF} \cdot 10^5$ A/A
3G RFET	P	84	98	24.5	0.19	1.27
	N	84	94	22.2	0.20	1.11
27 nm FDSOI	P	95	148	447	1.19	3.77
	N	94	136	1160	0.88	13.1
60 nm FDSOI	P	71	37	309	0.03	117
	N	70	30	1020	0.02	518

static results, SS and DIBL, than the 27 nm long FDSOI. However, when the gate length is expanded to 60 nm, the FDSOI electrostatic is much better than for the reversible device. This fact suggests that the worse FDSOI electrostatic control at 27 nm length is coming from Short-Channel Effects [13] rather than any other factor.

The maximum drain currents at fixed supply voltage (V_{DD}) are compared in Fig. 7a. The reversible transistor values remain very low but symmetric, being from 70 to 105 times lower compared to the 27 nm length FDSOI devices. This reduction in the drain current is related to several aspects: (i) the low efficiency of the carrier injection mechanism by tunneling (at the silicide/silicon junction), (ii) to the additional resistance induced by the spacers, uncover gate regions between PG and CG contacts, and (iii) to the reduced lateral electric field. Fig. 8 illustrates the top-interface (≈ 0.5 nm away from the front-gate oxide) energy bands for similar biasing conditions and effective gate length. The lateral electric field, i.e. slope of the energy bands, is weaker in RFETs which implies lower carrier accelerations and less drain current. On the other hand, the ON vs. OFF current figure of merit is depicted in Fig. 7b for several

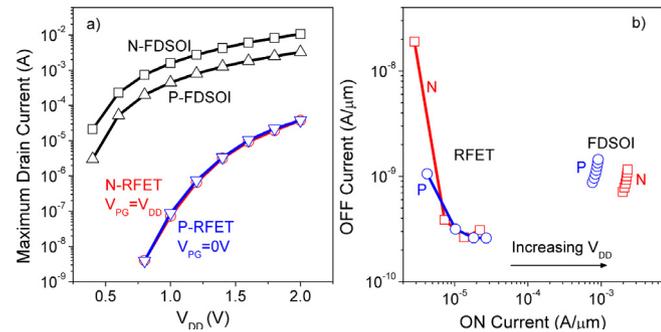


Fig. 7. (a) Maximum drain current and (b) I_{ON} vs. I_{OFF} figure of merit comparison for 3G-RFET and FDSOI. $V_T = V_{CG}$ ($I_{DS} = 0.1 \cdot W/L_{Si}$ $\mu\text{A}/\mu\text{m}$), $I_{ON} = I_{DS}$ ($V_{CG} = V_T \pm 0.65$ V) and $I_{OFF} = I_{DS}$ ($V_{CG} = V_T \mp 0.35$ V). RFET $L_{SOI} = 70$ nm. FDSOI $L_{SOI} = L_G = 27$ nm. $V_{BG} = 0$ V. $W = 1$ μm .

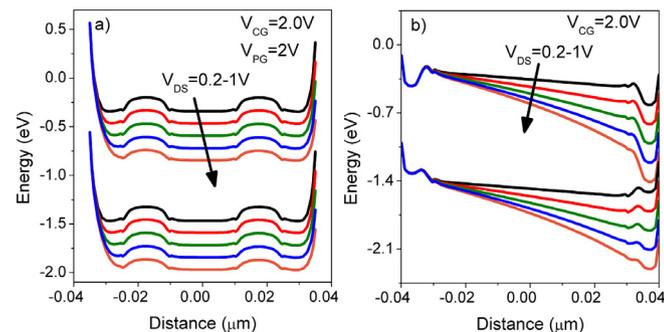


Fig. 8. Horizontal (≈ 0.5 nm from top-interface) N-MOS conduction and valence energy bands for (a) RFET and (b) FDSOI at similar length. $V_{BG} = 0$ V.

power supply voltages. Notice how in RFETs increasing V_{DD} firstly reduces the I_{OFF} unlike in MOSFETs, where it monotonically increases. This artifact comes due to the way the ON and OFF currents are extracted based on fixed biasing ($\pm 2/3$ V and $\mp 1/3$ V, respectively) from the constant current threshold voltage [14]. At low V_{DD} , the RFET V_T is in the linear region due to the limited current and I_{OFF} is very high. As V_{DD} rises, the OFF current gradually moves to the exponential region and decreases drastically. For higher V_{DD} I_{OFF} increases again as usual.

Concerning the capacitance analysis, the simpler case with $V_{DS} = V_{BG} = 0$ V is addressed due to the complex charge distribution in the reconfigurable FET. In such case, the control gate capacitances are compared in Fig. 9a for isolated devices. It can be noticed how the FDSOI CG capacitance is a bit larger simply due to the thinner lateral spacers (L_{SP}) in the structure. However, and despite the thicker spacers, under normal conditions when a reversible transistor changes its polarity, the overall capacitance is still higher in RFETs due to the need of driving also the capacitance associated to the two additional polarity gates, Fig. 9b.

Mixed-mode 2D simulations [10] are now considered to study the logic inverter response. An external load capacitance of 0.3 fF is connected at the inverter output to account for possible parasitic as in [15]. No other capacitances or resistances due to interconnections have been accounted for. The voltage-transfer characteristics (VTC) of single stage inverters made with FDSOI and RFETs are depicted in Fig. 10a. Both designs show the typical inverter operation, the RFET configuration nonetheless presents a degraded VTC. Due to particular transport mechanism, the current variation in RFET is strongly different compared to FDSOI. The drain current at high (N-RFET) or low (P-RFET) VCG is determined by the tunneling barrier thickness rather than by the control gate-induced energy barrier. In fact, the first derivative of the $I_D(V_D)$ has a super-linear variation and this effect strongly influences the voltage-transfer function of the RFET inverter. The transient response to a square signal of 500 MHz is represented in Fig. 10b. The FDSOI inverter response clearly outraces the RFET. The rise, t_{LH} , and fall,

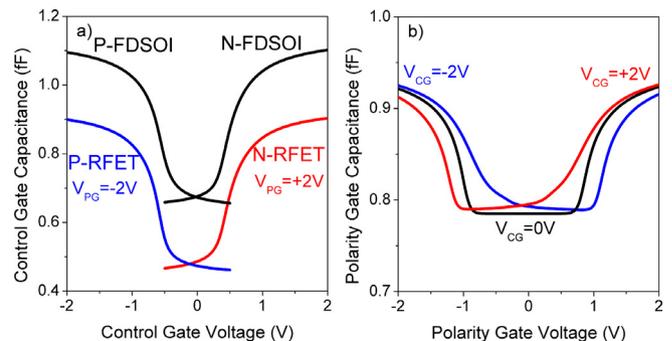


Fig. 9. Capacitance benchmark for (a) control gate in isolated RFET and 27 nm FDSOI. (b) Single 3G RFET polarity-gate capacitance at different control gate biases. $V_{DS} = 0$ V and $V_{BG} = 0$ V. FDSOI/RFET $W_P = W_N = 1$ μm .

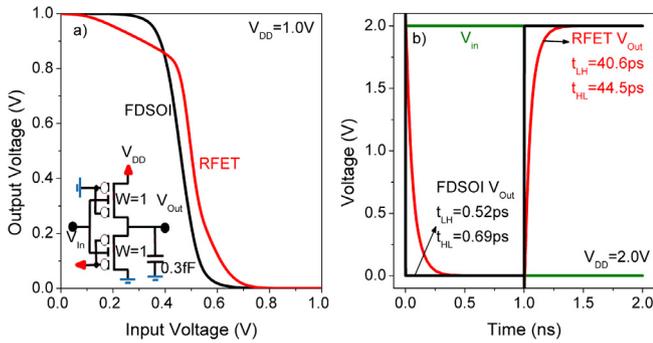


Fig. 10. Mixed-mode (a) voltage-transfer characteristics. (b) Transient response to square signal, $f = 500$ MHz. Inset: RFET inverter with 0.3 fF output capacitance. FDSOI $W_p = 2W_n$. RFET $W_p = W_n$.

t_{HL} , times are calculated as the crossing points at 50% of the supply voltage, V_{DD} . These rise and fall times can be averaged to extract the inverter propagation delay, τ_p , illustrated in Fig. 11a. Symmetric S/D silicided RFETs (circles, \circ) feature, at least ($V_{DD} = 2$ V), 70 times larger delay than FDSOI (squares, \square). The different RFET and FDSOI inverter delays we extract contrast with those employed in [16] to compute the overall circuit delay in complex designs with the logical effort method [17]. The logical effort enables the calculation of the circuit delay normalized to the inverter delay for a given technology. In this work we show that the delay is significantly higher in RFETs with respect to FDSOI, so, assuming both inverter delays identical, as done in [16], leads to misleading results.

Fig. 11b shows the static inverter current against the delay for different supply biases. The larger delays in reversible FETs are obtained at lower static current. FDSOI inverters are much faster than RFETs at the expense of larger static power. This however, does not directly imply reduced power consumption, the main consumption factors are related to the dynamic and the direct path power as will be commented later. The longer delays in RFET are induced by the lower drain current, Fig. 7a, yielding high equivalent resistances [18]:

$$\tau_p = \ln(2) \cdot C_L \cdot R \quad (1)$$

with C_L being the total inverter load capacitance. This load capacitance may be extracted from Eq. (1) through the propagation delay and the equivalent resistance R . The resistance was approximated using the maximum currents ($R = V_{DS}/I_{DS}$, with I_{DS} from Fig. 7a). Values of $C_L = 2.2$ and 2.0 fF were obtained for FDSOI and RFET inverters, respectively. These capacitance values are in line with the simulations results in Fig. 9. Once C_L is determined, the dynamic energy per transition, E , can be obtained by using [18]:

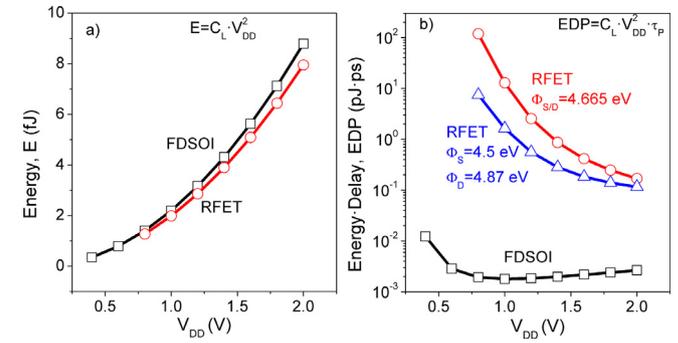


Fig. 12. (a) Energy per operation and (b) energy delay product comparison between 27 nm FDSOI and 3G-RFETs. FDSOI $W_p = 2W_n$. RFET $W_p = W_n = 1 \mu\text{m}$.

$$E = C_L \cdot V_{DD}^2 \quad (2)$$

The energy is shown in Fig. 12a. Very similar energies are extracted for FDSOI and RFETs as a result of having comparable output load capacitances. As observed, the energy can be made arbitrarily low by reducing the supply voltage. From this perspective, the optimum voltage to run the circuit would be the lowest possible that still ensures functionality. This comes at expense of the delay, Fig. 11a. A more relevant metric combining the measure of consumption and performance is the energy delay product (EDP) [18], represented in Fig. 12b:

$$EDP = E \cdot \tau_p = C_L \cdot V_{DD}^2 \cdot \tau_p \quad (3)$$

The optimum V_{DD} to run the inverter is therefore the point for which the EDP becomes minimum. It turns out to be ≈ 1 V for FDSOI while for RFETs the optimum voltage is larger than the maximum V_{DD} considered, 2 V. In any case, we find a much lower EDP for FDSOI technology than for any RFET configuration reflecting the clear superiority of FDSOI.

Speaking of the inverter power, Fig. 13 depicts the three main consumption components, P_{Sta} (static), P_{Dym} (dynamic) and P_{Dp} (power associated to the direct path current while N- and P-FETs are active) together with the total consumption, P_{Tot} . FDSOI MOS-FETs consumes from 10% to 18% more power than RFETs due to the larger static current (Fig. 11b) and especially the slightly larger energy per transition (Fig. 12a). However, this result needs to be carefully considered since the RFET scenario is optimum. The PG capacitances are not driven reducing the total output capacitance, C_L , hence the energy per transition, Eq. (2), and the dynamic and total consumption.

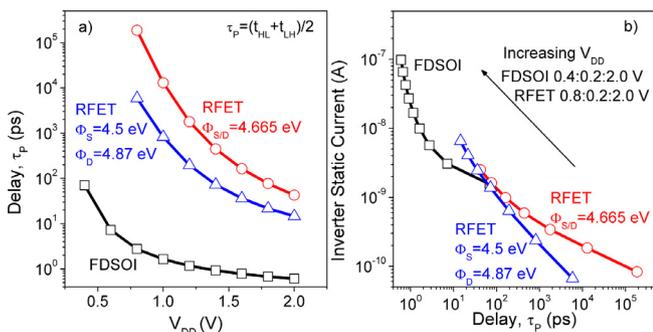


Fig. 11. (a) Inverter delay and (b) inverters static current vs. delay inverter comparison between 27 nm FDSOI (squares, \square) and 3G-RFETs with symmetric, $\Phi_s = \Phi_D$, (circles, \circ) and asymmetric, $\Phi_s \neq \Phi_D$, (triangles, Δ) S/D silicidations. FDSOI $W_p = 2W_n$. RFET $W_p = W_n = 1 \mu\text{m}$.

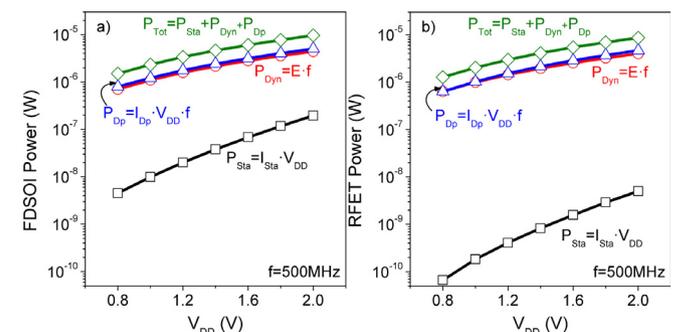


Fig. 13. Power consumption comparison between (a) 27 nm FDSOI and (b) 3G-RFET at $f = 500$ MHz. $P_{Sta} = I_{Sta} \cdot V_{DD}$ (I_{Sta} = static current), $P_{Dym} = E \cdot f$, $P_{Dp} = I_{Dp} \cdot V_{DD} \cdot f$ (I_{Dp} = time integrated direct path current) [18]. FDSOI $W_p = 2W_n$. RFET $W_p = W_n = 1 \mu\text{m}$.

Table 3
RFET simulated results from the literature. All devices but this work are nanowire shape. The symmetric silicides workfunctions, $\Phi_{S/D}$, are extracted from the SB heights assuming $\chi_{Si} = 4.05$ eV and $E_g = 1.12$ eV. The maximum currents are normalized to the nanowire perimeter.

–	$\Phi_{S/D}$ eV	$m_n^* m_0$	$m_p^* m_0$	I_{n-max} $\mu\text{A}/\mu\text{m}$	I_{p-max} $\mu\text{A}/\mu\text{m}$	V_D V	V_{PG} V	V_{CG} V
[19]	4.40	–	–	636.0	636.0	1.2	1.2	1.2
[20]	4.43	–	–	2.1	–	2.6	1.5	2.5
[11]	4.64	0.30	0.20	8.0	8.0	2.0	2.0	2.0
[15]	4.46	0.19	0.16	636.0	636.0	1.2	1.2	1.2
[16]	4.71	0.19	0.16	159.0	159.0	1.5	1.5	1.5
This work	4.66	0.30	0.20	37.5	37.5	2.0	2.0	2.0

5. State-of-the-art

The limited RFET drain current represents the main drawback regarding the performance difference with FDSOI. In this section, we compare the current from all reconfigurable FETs simulations and experimental results published so far. Table 3 gathers all TCAD results with the maximum currents for N/P RFETs, together with the effective masses and S/D workfunction employed when specified. In other simulation works, RFETs seem to provide very decent and symmetric current levels that could balance the performance with typical MOSFETs. Two basic discrepancies appear between the results obtained in this work and the best values in Table 3 from EPFL [15,19]: (i) the huge maximum currents, over 600 $\mu\text{A}/\mu\text{m}$ (almost 20 times more), and (ii) the current symmetry given the different S/D silicides workfunction.

Initially, the high current level could be explained by the lighter effective masses. Fig. 14a illustrates the electron/hole maximum drain current dependence with the effective masses for a planar 3G-RFET structure. Lower masses improve the tunneling efficiency, thus enhancing the current. Nevertheless, the reduction in the effective mass only yields a factor 2–3 in the current. A not very fine mesh, with an arbitrary 50% enhancement in the tunneling transmission (parameters gc/gv [10]) and the field enhancement induced by the improved confinement due to the 3D nanowire structure (shorter screening length than in planar FDSOI) turn to

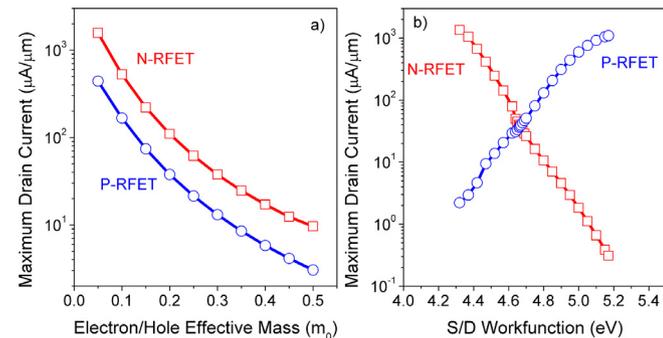


Fig. 14. a) Maximum N/P drain current as a function of the (a) effective mass ($\Phi_{S/D} \approx 4.66$ eV) and (b) the S/D silicides workfunction ($m_n^* = 0.3 m_0$ and $m_p^* = 0.2 m_0$) for the optimized 3G-RFET. $V_{DS} = \pm 2$ V and $V_{BG} = 0$ V. $W = 1$ μm .

Table 4

RFET Experimental results from the literature. All devices are nanowire shape but [21] that is FDSOI tri-gate with thick BOX (145 nm).

–	$\Phi_{S/D}$ eV	I_{n-max} $\mu\text{A}/\mu\text{m}$	I_{p-max} $\mu\text{A}/\mu\text{m}$	V_D V	V_{PG} V	V_{CG} V
[6]	4.40	636.0	636.0	1.2	1.2	1.2
[21]	4.43	2.1	–	2.6	1.5	2.5
[22]	4.64	8.0	8.0	2.0	2.0	2.0
[23]	4.46	636.0	636.0	1.2	1.2	1.2
[15]	4.71	159.0	159.0	1.5	1.5	1.5
[24]	4.66	37.5	37.5	2.0	2.0	2.0

be the responsible of this current difference. This scenario reflects a rather too optimistic case with beneficial strain for both carriers simultaneously, which has yet to be proven experimentally [25], and induce an unrealistic current gain of around 5 to 15 times that combines with the effective mass impact. In any case, the use of ideal models during simulations, not considering parasitic effects near the interface between metal and semiconductor, may motivate too favorable currents with respect to experimental results.

Attending to the identical N/P currents in [15,19], they are obtained for $\Phi_{S/D}$ close to 4.4 eV, which leads to extremely asymmetric SB heights, 0.35/0.77 eV for electrons and holes, respectively. Fig. 14b illustrates the maximum drain current for N/P 3G-RFETs as a function of the silicides workfunction. Thus, the selected barriers in [15,19] should typically lead to NFET currents much larger than in PFET. Identical currents are extracted at $\Phi_{S/D} \approx 4.66$ eV yielding barriers height of 0.61/0.51 eV. The also strong currents from [16] are obtained thanks to the lower effective mass, improved electrostatic when using a thinner EOT and the shorter length. It is worth noting the strong impact the Schottky barrier height has on the current, being able to modulate it by more than three orders of magnitude. Finally, simulations show how by choosing mid-gap S/D workfunctions, the maximum N/P R-FET current drops by a factor 30, Fig. 14a. Hence, the N/P current symmetry goal imposes the fundamental performance boundary in RFETs.

On the other hand, experimental results are summarized in Table 4. Currents are consistently below 30 $\mu\text{A}/\mu\text{m}$ validating our weak simulated currents. Only devices from [15] show high drain currents over 290 $\mu\text{A}/\mu\text{m}$, partly induced by the large polarity gate biases used and by the fact that the current is not symmetric which benefits one of the N/P branches.

6. RFET enhancement tips

This section provides some ideas to improve the operation of RFETs. Three main advises are commented now in order to increase the drain current:

1. Lower the effective mass to improve tunneling efficiency by applying strain boost techniques.
2. Use compound materials to reduce the semiconductor energy band-gap, thus the electron/hole SB heights.

Table 53G-RFET simulations to improve the current by reducing the energy band-gap, using asymmetric S/D workfunctions and lowering the effective mass. $V_{DS} = \pm 1.0$ V.

Eg eV	χ_0 eV	Φ_S eV	Φ_D eV	$m_n^* m_0$	$m_p^* m_0$	I_{n-max} $\mu\text{A}/\mu\text{m}$	I_{p-max} $\mu\text{A}/\mu\text{m}$
1.12	4.05	4.66	4.66	0.30	0.20	28.1	30.8
0.99	4.035	4.66	4.66	0.30	0.20	23.3	234.6
1.12	4.05	4.50	4.87	0.30	0.20	130.8	142.7
1.12	4.05	4.66	4.66	0.19	0.16	92.6	51.4
1.12	4.05	4.50	4.87	0.19	0.16	277.1	182.9
0.99	4.035	4.50	4.87	0.19	0.16	235.8	606.8

3. Use different silicides at S/D to obtain asymmetric workfunctions, each of them favoring the tunneling of one type of carrier (typically $\Phi < E_g/2$ eV for electrons and $\Phi > E_g/2$ eV for holes).

Table 5 compares the current at each case independently and with all enhancing factors at the same time. The selected values for band-gap, workfunctions and effective masses are not arbitrary. The 0.99 eV band-gap corresponds to SiGe with 30% Ge ($\chi_0 \approx 4.035$ eV) [26], the silicides workfunctions to PtSi ($\Phi \approx 4.87$ eV) and TiSi₂ ($\Phi \approx 4.5$ eV) [27] and the lower effective masses are commonly used as default values [15,16]. All of these modifications independently enhance the current.

Using SiGe slightly deteriorates (18%) the electron current since the electron affinity is smaller which leads to an electron SB height a bit larger (≈ 0.015 eV higher). This could be solved by readjusting the S/D workfunction to mid-gap though. On the other hand, SiGe is extremely beneficial for the P-RFET current, the lower affinity combines with the smaller band-gap and reduce the hole SB height (≈ 0.135 eV less), hence huge current improvements (760%) are achieved. The asymmetric boost of SiGe demonstrates that using different band-gap materials can be interesting to balance the current between N/P RFETs.

The asymmetric silicidation at source and drain is even more interesting since it may be beneficial for both N and P configurations. The simulated workfunctions were selected based on Fig. 14b to maintain similar electron/hole current, Fig. 15a. The enhancement is close to a factor 4.7. Even more dissimilar silicidations could be employed but the transfer characteristics is rapidly degraded as shown in Fig. 15b. Increasing (reducing) the workfunction for holes (electrons) boosts the maximum currents at the expense of the off current and electrostatic control due to the increase in the thermionic emission. As the S/D metallic fermi level approaches the valence or conduction energy band, less energy is required to surmount the energy barrier and more carriers are injected in the body from the source. The direct consequence of using this asymmetric silicidation is observed in Figs. 11 and 12 (triangles, Δ) where the delay and EDP are reduced with respect to the symmetric RFET configuration (circles, \circ) by a factor

between 3 and 30 times. The problem is still the degraded electrostatics and the possible implementation in actual and more complex circuits.

Finally, lightweight effective masses provide a gain of around 1.6–3, being stronger for N-RFET since the drop in the mass is larger. By combining all the boosters, the current may rise up to 20 times (as in the case for P-RFETs). This impressive enhancement is however not sufficient to compensate the large delay difference with MOSFETs, especially if we account for the degradation of the electrostatics and inverter VTC characteristics.

7. Conclusions

Reconfigurable FETs provide, in theory, an interesting scenario to build reprogrammable logic. They represent a powerful tool to reduce the number of devices, the critical paths and allowing XOR based circuits, which are inefficient in CMOS. However, in practice, their poor performance compared to typical MOSFETs threatens their broad adoption in future applications. The main RFET asset, which is the N/P operation switch in the same device, becomes also their fundamental drawback. The simultaneous enhancement of both electron and hole currents is very challenging and a tradeoff appears between N- and P-RFETs performance.

Aspects like the increasing surface due to the extra gates and the associated circuitry to control the polarity gates, the rising routing complexity and the control of the SB quality also jeopardize the possible benefits of reducing the total number of transistors in a circuit design. In any case, the global RFET performance is not expected to equal the one from MOSFETs, thus limiting RFETs for embedded or not very demanding applications.

Acknowledgements

Project supported by the French ANR via Carnot funding.

References

- [1] Moore GE. Cramming more components onto integrated circuits. *Proc IEEE* 1998;86(1):82–5.
- [2] Iwai H. Materials and structures for future nano CMOS. In: *Nanotechnology Materials and Devices Conference (NMDC)*, 2011 IEEE. p. 14–8.
- [3] Barraud S, Coquand R, Casse M, Koyama M, Hartmann JM, Maffini-Alvaro V, et al. Performance of omega-shaped-gate silicon nanowire MOSFET with diameter down to 8 nm. *IEEE Electron Device Lett* 2012;33(11):1526–8.
- [4] Villalon A, Royer CL, Cass M, Cooper D, Hartmann JM, Allain F, et al. Experimental investigation of the tunneling injection boosters for enhanced ion ET-SOI tunnel FET. *IEEE Trans Electron Devices* 2013;60(12):4079–84.
- [5] Gaben L, Barraud S, Jaud MA, Martinie S, Rozeau O, Lacord J, et al. Stacked-nanowire and FinFET transistors: guidelines for the 7 nm node. In: *2015 International conference on Solid State Devices and Materials (SSDM)*.
- [6] Heinzig A, Slesazek S, Kreupl F, Mikolajick T, Weber WM. Reconfigurable silicon nanowire transistors. *Nano Lett* 2012;12(1):119–24.
- [7] Rowe JE, Christman SB, Margaritondo G. Metal-induced surface states during Schottky-barrier formation on Si, Ge, and GaAs. *Phys Rev Lett* 1975;35:1471–5.
- [8] Lepselter MP, Sze SM. SB-IGFET: an insulated-gate field-effect transistor using Schottky barrier contacts for source and drain. *Proc IEEE* 1968;56(8):1400–2.
- [9] Larson JM, Snyder JP. Overview and status of metal S/D Schottky-barrier MOSFET technology. *IEEE Trans Electron Devices* 2006;53(5):1048–58.
- [10] Synopsys I. Sentaurus device user guide, i.2014-09 ed.; 2014.

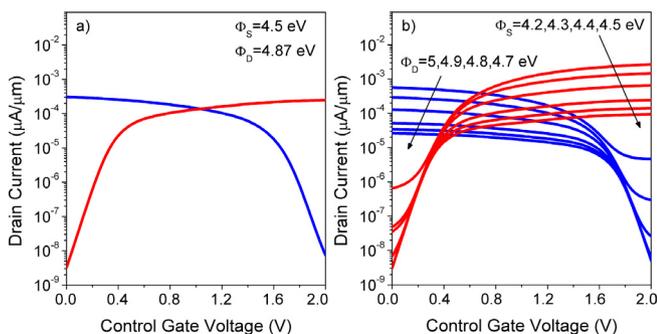


Fig. 15. Transfer characteristics for asymmetric silicidated RFET ($\Phi_S \neq \Phi_D$) with (a) similar output current and (b) similar electron/hole SB heights. $V_{DS} = \pm 2$ V. $W = 1 \mu\text{m}$.

- [11] Trommer J, Heinzig A, Slesazek S, Mikolajick T, Weber WM. Elementary aspects for circuit implementation of reconfigurable nanowire transistors. *IEEE Electron Device Lett* 2014;35(1):141–3.
- [12] DeSalvo B, Morin P, Pala M, Ghibaudo G, Rozeau O, Liu Q, et al. A mobility enhancement strategy for sub-14 nm power-efficient FDSOI technologies. In: 2014 IEEE international electron devices meeting. p. 7.2.1–4.
- [13] Veeraraghavan S, Fossum JG. Short-channel effects in SOI MOSFETs. *IEEE Trans Electron Devices* 1989;36(3):522–8.
- [14] Schroder D. *Semiconductor Material and Device Characterization*. John Wiley; 2006.
- [15] Marchi MD, Sacchetto D, Zhang J, Frache S, Gaillardon PE, Leblebici Y, et al. Top-down fabrication of gate-all-around vertically stacked silicon nanowire FETs with controllable polarity. *IEEE Trans Nanotechnol* 2014;13(6):1029–38.
- [16] Trommer J, Heinzig A, Baldauf T, Slesazek S, Mikolajick T, Weber WM. Functionality-enhanced logic gate design enabled by symmetrical reconfigurable silicon nanowire transistors. *IEEE Trans Nanotechnol* 2015;14(4):689–98.
- [17] Sutherland I, Sproull B, Harris D. *Logical effort: designing fast CMOS circuits*. Morgan Kaufmann; 1999.
- [18] Rabaey J, Chandrakasan A, Nikolic B. *Digital integrated circuits. A design perspective*. Prentice Hall; 2003.
- [19] Zhang J, Tang X, Gaillardon PE, Micheli GD. Configurable circuits featuring dual-threshold-voltage design with three-independent-gate silicon nanowire FETs. *IEEE Trans Circuits Syst I: Regular Pap* 2014;61(10):2851–61.
- [20] Marchi MD, Zhang J, Frache S, Sacchetto D, Gaillardon PE, Leblebici Y, et al. Configurable logic gates using polarity-controlled silicon nanowire gate-all-around FETs. *IEEE Electron Device Lett* 2014;35(8):880–2.
- [21] Schwalke U, Krauss T, Wessely F. CMOS without doping on SOI: multi-gate Si-nanowire transistors for logic and memory applications. *ECS J Solid State Sci Technol* 2013;2(6):Q88–93.
- [22] Heinzig A, Mikolajick T, Trommer J, Grimm D, Weber WM. Dually active silicon nanowire transistors and circuits with equal electron and hole transport. *Nano Lett* 2013;13(9):4176–81.
- [23] Zhang J, Marchi MD, Sacchetto D, Gaillardon PE, Leblebici Y, Micheli GD. Polarity-controllable silicon nanowire transistors with dual threshold voltages. *IEEE Trans Electron Devices* 2014;61(11):3654–60.
- [24] Weber WM, Trommer J, Grube M, Heinzig A, Knig M, Mikolajick T. Reconfigurable silicon nanowire devices and circuits: opportunities and challenges. In: 2014 Design, Automation Test in Europe conference exhibition (DATE). p. 1–6.
- [25] Baldauf T, Heinzig A, Mikolajick T, Weber WM, Trommer J. Strain-engineering for improved tunneling in reconfigurable silicon nanowire transistors. In: 2016 Joint international EUROSOL workshop and international conference on Ultimate Integration on Silicon (EUROSOL-ULIS). p. 1–4.
- [26] Kasper E. Properties of strained and relaxed silicon Germanium. *INSPEC*; 1995.
- [27] Drummond TJ. Work functions of the transition metals and metal silicides. *J Appl Phys FT* 1999.



Simulation study of a novel 3D SPAD pixel in an advanced FD-SOI technology



M.M. Vignetti ^{a,*}, F. Calmon ^a, P. Lesieur ^a, A. Savoy-Navarro ^b

^a Institut des Nanotechnologies de Lyon, Université de Lyon, France

^b Laboratoire d'AstroParticule et Cosmologie, Université Paris-Diderot, Paris, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

3D pixel
APD
SPAD
SOI
FDSOI
Geiger-mode
Avalanche diode
TCAD simulations
Back-side illumination
Avalanche triggering probability
Dark count rate
Photon detection probability
Time-of-flight applications

ABSTRACT

In this paper, a novel SPAD architecture implemented in a Fully-Depleted Silicon-On-Insulator (SOI) CMOS technology is presented. Thanks to its intrinsic vertical 3D structure, the proposed solution is expected to allow further scaling of the pixel size while ensuring high fill factors. Moreover the pixel and the detector electronics can benefit of the well-known advantages brought by SOI technology with respect to bulk CMOS, such as higher speed and lower power consumption. TCAD simulations based on realistic process parameters and dedicated post-processing analysis are carried out in order to optimize and validate the avalanche diode architecture for an optimal electric field distribution in the device but also to extract the main parameters of the SPAD, such as the breakdown voltage, the avalanche triggering probability, the dark count rate and the photon detection probability. A comparison between the efficiency in back-side and front-side approaches is carried out with a particular focus on time-of-flight applications.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

There are manifold areas of research and commercial applications where Single-Photon Avalanche Diodes (SPADs, or Geiger-mode avalanche diodes) have been extensively studied and successfully implemented for the detection of weak optical signals in the visible and near-infrared spectrum range [1]. Moreover, during the last few years, there has been an active research on the implementation of SPADs in the field of High Energy Physics and Medical Physics for the detection of ionizing particles [2], opening new promising fields of application for these devices.

A SPAD consists of a p-n junction which is reverse biased above the breakdown voltage V_{bd} by an excess bias V_{ex} . When an electron-hole pair (EHP) is generated in the space charge region (SCR) of the junction (e.g. by an incoming photon, ionizing particle), a self-sustained charge multiplication process by impact ion-

ization could be triggering according to a certain probability (depending on the applied reverse bias), giving rise to a macroscopic electric current. Such a high current usually translates into a high current density which could cause permanent damage in the device. For this reason, every SPAD pixel needs suitable “quenching” electronics responsible for interrupting the multiplication process right after the avalanche build-up by promptly lowering the reverse bias of the junction below the breakdown threshold.

Then, after a certain dead-time (hold-off time) during which the pixel is “blind” to any incoming photon (or particle), the electronics restores the p-n junction to the initial bias (reset phase). The whole quench/recharge cycle provides the information that an event has occurred [1]. In this work, a novel SPAD architecture consisting of a 3D pixel with associated quenching electronics is presented and studied by means of TCAD simulations and dedicated numerical methods for post-processing of the output data obtained from simulation. The pixel is conceived for an advanced 28 nm Fully-Depleted Silicon-On-Insulator (FD-SOI) CMOS technology and it is suitable for the detection of ionizing particles as well as

* Corresponding author.

E-mail address: matteo.vignetti@insa-lyon.fr (M.M. Vignetti).

¹ Early Stage Researcher supported by the 2012-FP7-ITN, n° 317446, INFIERI EU program.

light (visible and near-infrared range) in backside illumination (BSI) mode. Thanks to its vertical 3D structure, the proposed solution is expected to allow very small pixels (down to a few μm^2) while enabling high fill factor. Moreover, the pixel read-out electronics as well as the whole detector electronics can benefit of the well-known advantages brought by SOI technology such as higher speed and lower power consumption with respect to bulk CMOS. To the best of the authors' knowledge, this is the first time that a 3D monolithic SPAD integrated in a standard FDSOI CMOS process is proposed and studied.

2. 3D pixel concept for SPAD in SOI technology

Compared to SPAD architectures conceived so far in SOI technology [3–5], the pixel proposed in this work has the great advantage to provide a monolithic 3D structure without the need of dedicated 3D integration techniques. The avalanche diode is indeed defined beneath the Buried Oxide (BOX) while the quenching electronics is sitting on top of it, in the SOI layer as depicted in Fig. 1.

The pixel has been designed according to the features of an advanced Fully-Depleted SOI technology, by exploiting the available implantations and diffusions that are normally meant to provide different back-biasing strategies for the transistors. The diode sensitive region is defined in the Space Charge Region (SCR) of a p-well/deep n-well junction. Premature Edge Breakdown (PEB) risk is prevented thanks to a guard-ring placed around the sensitive area. A low doped p-type region can indeed be obtained thanks to the retrograde doping of the deep n-well in the epitaxial p-type substrate. Such a region is responsible for smoothing down the electric field at the junction edge that otherwise would be too intense to allow Geiger-mode operation. The diode can be connected to its associated electronics thanks to back-gate contacts featured by the adopted FDSOI technology, but originally meant to enable a “tunable” threshold voltage for the transistors in the SOI. Two different biasing options are possible for the avalanche diode, depending on whether the output node is the anode (diode TOP) or the cathode (diode BOTTOM) as shown in Fig. 2. It is important to point out that the p-well placed below the BOX (the avalanche diode's anode) acts as a back-gate for the transistors' channel in the SOI, i.e. the threshold voltage is affected by the bias chosen for the anode. It is therefore recommended to bias the p-well at ground in order to prevent any threshold variation for the transistors. In the “diode BOTTOM” configuration (chosen for the TCAD study discussed in Section 3), the output is sensed at the cathode, i.e. the deep n-well in Fig. 1, while the anode (according to the above discussion) is biased at ground avoiding any transistor's threshold variation. However, under this configuration, the output voltage falls within the range from V_{bd} to $V_{bd} + V_{ex}$ which is not compatible with standard digital voltage levels, i.e. 0V to V_{dd} . This requires an additional DC decoupling capacitor between the output of the avalanche diode and the pixel electronics. In the “diode TOP” configuration, the p-well (anode) bias is varying since it is the output node, swinging within the voltage range from 0V to V_{ex} which is compatible with standard voltage levels if $V_{ex}^{max} = V_{dd}$. During every avalanche event the transistors in the SOI layer would thus experience higher or lower threshold voltages with respect to the “quiescence state” ones, depending on the channel type. For this reason such a configuration requires pixel electronics insensitive to (or conveniently exploiting) back-gating effects.

Based on the considerations made so far, two matrix arrangements can be obtained with the proposed pixel, as shown in Fig. 3. Solution (a) provides shielding of the pixel electronics by grounding the p-well. The output is thus sensed at the cathode which penalizes the fill-factor as every pixel needs an independent

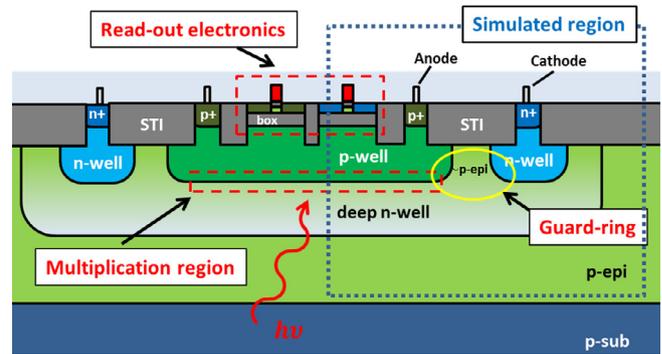


Fig. 1. Schematic representation of the proposed 3D pixel based on an advanced FDSOI technology.

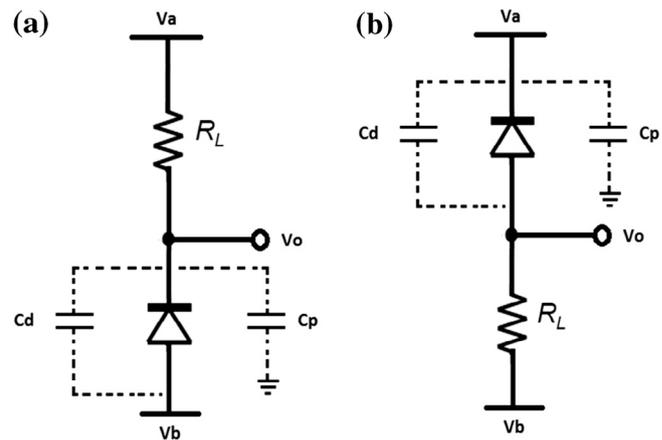


Fig. 2. Schematic circuit (simplified) of (a) “diode BOTTOM” configuration showing all the capacitive contributions (b) “diode TOP” configuration. Observe that C_d and C_p are the p-well/deep-n-well and deep n-well/p-substrate space charge capacitances respectively.

deep n-well. Solution (b) enables higher fill factor (common deep n-well) but the electronics need to be insensitive to back-gating effects. It is worth noticing that the resulting back-side illuminated 3D pixel can dramatically improve the detector fill factor (FF) with respect to traditional SPADs, especially in case of small pixels, where this would be strongly degraded by the surface of the quenching electronics sitting next to the avalanche diode. The fill factor of the proposed 3D pixel is indeed limited by the “dead” areas of the avalanche diode only, due to the guard-ring and the minimum distance between two diodes.

3. Simulation and post-processing methodology

The proposed 3D pixel has been studied by means of TCAD simulations and dedicated post-processing analysis following the approach depicted in Fig. 4. TCAD simulations provided an estimation of the avalanche diode breakdown voltage (V_{bd}) and allowed validating the diode architecture for a correct Geiger-mode operation by studying the electric field distribution all over the pixel. They also provided the ionization coefficients for electrons and holes (α_e, α_h) and the EHP generation rates for the Shockley-Read-Hall and band-to-band tunneling mechanisms (G_{SRH}, G_{B2B}). The post-processing analysis allowed the calculation of the main parameters of a SPAD, such as the avalanche triggering probability P_{tr} , the dark count rate (DCR) and the photon detection probability (PDP). In Sections 3.1 and 3.2, TCAD simulations and post-processing, respectively, are discussed in more details.

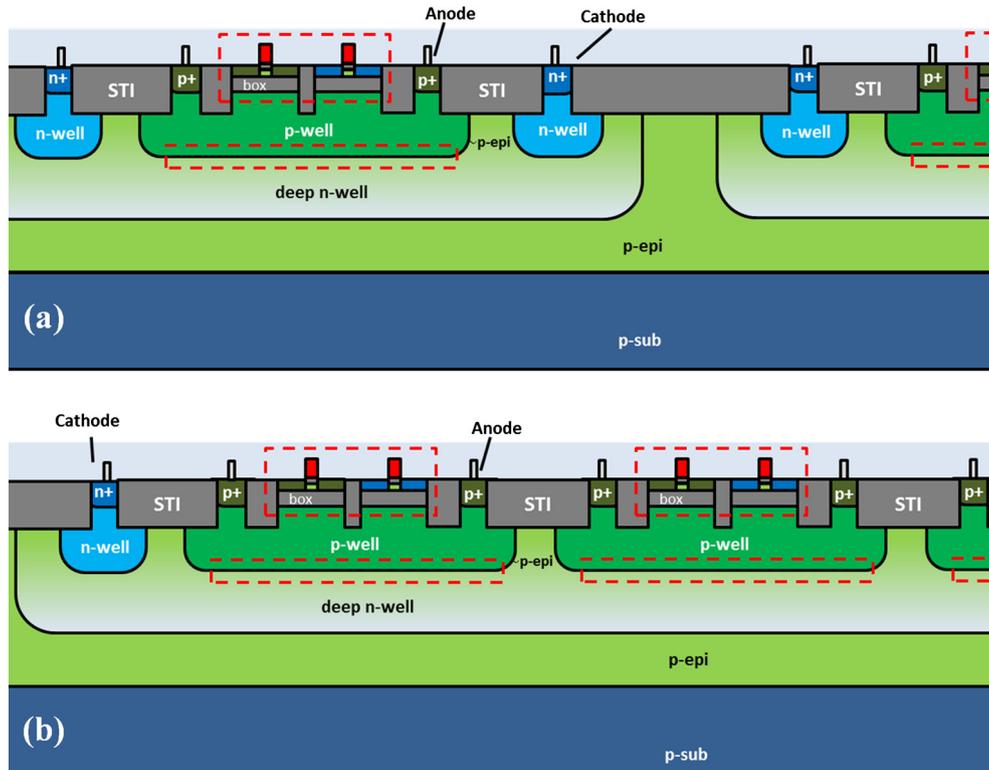


Fig. 3. Possible matrix arrangements for the proposed pixel. Solution (a): grounded anode. Solution (b): common deep n-well.

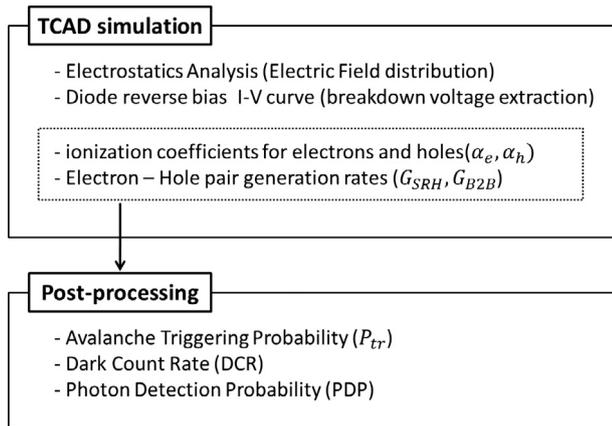


Fig. 4. Schematic representation of the adopted simulation strategy. TCAD results have been used as input data for a custom post-processing analysis.

3.1. TCAD simulations

TCAD simulations of the 3D pixel have been carried out in Synopsys Sentaurus, based on realistic process parameters provided by the foundry of the adopted technology. The pixel has been modeled as a two-dimensional geometry representing the radial cut of an avalanche diode with cylindrical symmetry (Fig. 5). This allowed emulating the geometry of a 3D device while dramatically reducing the overall computational time. Carrier transport in the device has been described by the drift-diffusion equation, accounting for the Fermi-Dirac statistics for the electrons and holes distribution in the semiconductor. The TCAD physical models considered the doping dependence of the carriers' mobility thanks to the Masetti model whose parameters for silicon are based on the experimental data reported in [6]. The avalanche charge multipli-

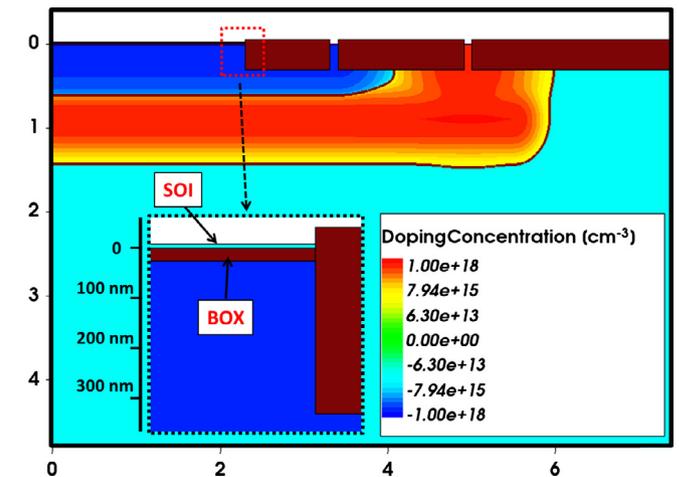


Fig. 5. TCAD geometrical model of the 3D SOI pixel. The insert shows the ultra-thin silicon layer over the buried oxide (Positive values refer to n-type doping. Spatial scales are in μm).

cation process depends on the ionization coefficients for electrons and holes, α_e and α_h respectively, which have been calculated according to the “van Overstraeten – De Man” model based on experimental data reported in [7]. Shockley-Read-Hall (SRH) and band-to-band (B2B) tunneling processes have been considered as the main contributors for the evaluation of the EHP generation-recombination within the avalanche diode sensitive volume.

The SRH model accounts for the presence of deep defect levels in the silicon energy gap, while the B2B one considers the electrons and holes generation enhancement due to the potential barrier thinning along the multiplication region when the diode is reverse biased above the breakdown voltage. The SRH process is strongly dependent on the carriers' lifetime $\tau_{n,p}$ that in turn depends on

many factors such as temperature and doping concentration. Typical values of $\tau_{n,p}$ range from 1 ms to 1 μ s depending on silicon purity, and fall down to about 10 ns only in case of extremely high doping levels where Auger recombination plays a crucial role [8]. For this reason the authors decided to evaluate the SRH generation process under two different scenarios, one accounting for a doping dependent carrier lifetime according to the “Scharfetter model” [9], the other considering constant carrier lifetimes $\tau_n = 10 \mu$ s and $\tau_p = 3 \mu$ s, for electrons and holes respectively [10]. Band-to-Band tunneling has been modeled with the field-enhanced Schenk model neglecting the phonon-assisted tunneling contribution since the electric field peak in the avalanche diode is not expected to exceed the value of $8 \cdot 10^5$ V/cm [11]. In the present study, the authors have chosen to keep the default parameter values provided by the simulation tool for the adopted physical models [10].

3.2. Post-processing

A post-processing analysis based on the parameters extracted by means of TCAD simulations has been performed in order to evaluate the main figures of merit of a SPAD such as the Avalanche Triggering Probability (ATP) P_{tr} , the Dark Count Rate (DCR) and the Photon Detection Probability (PDP).

ATP is the probability that an EHP generated in the multiplication region can successfully trigger an avalanche and can be evaluated by solving the following couple of differential equations [12]:

$$\begin{cases} \frac{dP_e}{dx} = \alpha_e(1 - P_e)(P_e + P_h - P_e P_h) \\ \frac{dP_h}{dx} = -\alpha_h(1 - P_h)(P_e + P_h - P_e P_h) \end{cases} \quad (1)$$

where $\alpha_e(x)$ and $\alpha_h(x)$ are the electrons and holes ionization coefficients (resulting from TCAD simulations), respectively, while $P_e(x)$, $P_h(x)$ are the probabilities for initiation an avalanche by an electron or a hole, respectively, generated at the position x within the space charge region of the avalanche diode. These equations can be integrated with the following boundary conditions:

$$P_h(x = x_p) = 0$$

$$P_e(x = x_n) = 0$$

where x_p and x_n are the boundaries of the diode SCR at the p-side and n-side, respectively. By adopting a numerical method proposed in [13], the ATP at a position x can be conveniently obtained as the joint probability of P_e and P_h :

$$P_{tr}(x) = P_e + P_h - P_e P_h \quad (2)$$

whose average all over the space charge region provides the average avalanche triggering probability:

$$\overline{P_{tr}} = \frac{1}{x_n - x_p} \int_{x_p}^{x_n} P_{tr}(x) dx \quad (3)$$

The knowledge of ATP enables the calculation of the DCR as well as the PDP for the SPAD under study.

Dark Count Rate (DCR) refers to undesired avalanche events which can be triggered by thermal and/or field-assisted generated EHP due to Shockley-Read-Hall (SRH) and band-to-band tunneling generation – recombination processes occurring in the diode space charge regions. The resulting spurious avalanche pulses represent a source of noise for SPAD detectors, depending on the adopted CMOS process and increasing with the sensor area, with temperature and excess bias. Based on the previous discussion, DCR can be calculated according to the following formula:

$$DCR = \int_{x_p}^{x_n} P_{tr}(x) G_{EHP}(x) dx \quad (4)$$

where $G_{EHP}(x)$ is the EHP generation rate as a function of the position x within the diode SCR. Such a parameter is extracted by means of TCAD simulations and, as discussed in Section 3.1, it accounts for both SRH and band-to-band generation phenomena in the diode SCR.

The Photon Detection Probability (PDP) is defined as the probability that a photon of a certain wavelength impinging on the pixel is effectively detected. In order for this to happen, two conditions must be satisfied:

- the photon has to be absorbed within the diode sensitive region, i.e. the n-type and p-type neutral regions and the SPAD multiplication region (i.e. SCR),
- the generated EHP (either the electron or the hole) has to successfully trigger an avalanche process.

The detection process of a single photon can be described with the help of an analytical model based on the work done by Gulnatti et al. [8]. It is convenient to simplify the study in a one-dimensional case, as depicted in Fig. 6, where it is possible to identify six different regions: the back-side Anti-Reflective Coating (ARC), the p-type substrate, a SCR between the deep n-well and the p-substrate, and the SPAD sensitive region consisting of the n/p-type neutral regions and the SPAD multiplication region. It is indeed really unlikely that minority carriers generated in device sensitive region can escape laterally without being collected by one of the two SCR. Only some of those generated really nearby the device periphery will be probably lost through the guarding. These latter minor losses are neglected in this one-dimensional approximation.

A photon with wavelength λ hitting the pixel from the back-side has a chance to be successfully detected only if it is absorbed in the SPAD active region (Fig. 6). In order to reach this area, the photon should not be back-reflected at the ARC – p-substrate interface. Indeed the chance for the photon to cross this optical barrier is defined by the light transmission coefficient, i.e. $T(\lambda) = 1 - R(\lambda) < 1$, due to the refraction index discontinuity. Moreover the photon should not be absorbed in the p-substrate region otherwise it would be irremediably lost as there would be no chance that the generated EHP can reach the SPAD multiplication region. Even if the generated minority electron reached the deep n-well/p-substrate SCR, it would be simply collected at the cathode without triggering any multiplication process. If conversely the photon is absorbed in the SPAD active region, three possible scenarios are possible. Fig. 6(a) represents the case where the photon is absorbed in the multiplication region (i.e. SCR). The local high electric field promptly accelerates the generated EHP towards the SCR ends, which might lead to a self-sustained multiplication process. Under this scenario, an EHP is generated within x and $x + dx$ in the SCR with a probability given by (5):

$$p_{abs}(x, \lambda) dx = e^{-\alpha(\lambda)x} \alpha(\lambda) dx \quad (5)$$

where $\alpha(\lambda)$ is the photon absorption coefficient in silicon. The generated EHP can eventually trigger an avalanche process with a probability $P_{tr}(x)$, according to (2). Therefore, the photon detection probability under this scenario is given by (6):

$$PDP_{(a)} = \int P_{tr}(x) p_{abs}(x) dx \quad (6)$$

where the integral has to be calculated over the entire multiplication region.

Fig. 6(b) represents the case where the photon is absorbed in the p-type neutral region. Under this scenario, the generated EHP can lead to an avalanche event only if the minority electron successfully reaches by diffusion the upper end of the multiplication region where it can eventually trigger an avalanche process with

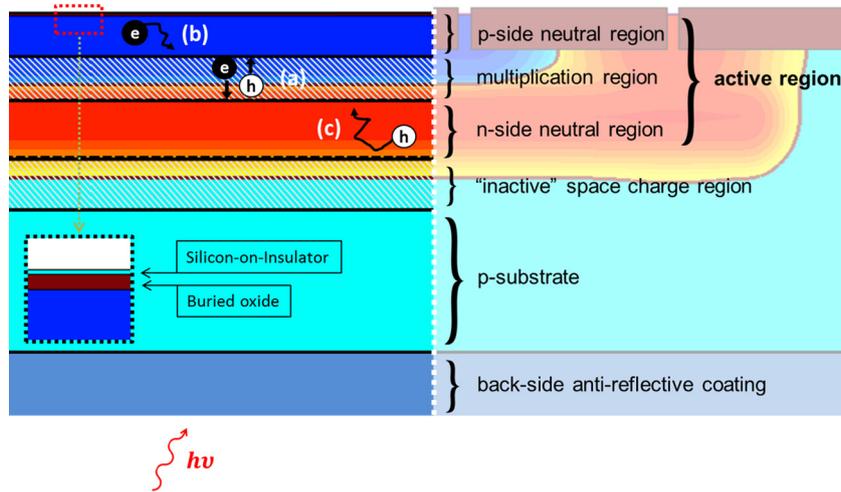


Fig. 6. Schematic representation of a photon absorption in the active region of the proposed 3D pixel. Three main scenarios have been depicted: (a) the photon is absorbed in the multiplication region and the generated EHP are swift away by the high electric field. (b) and (c) The photon is absorbed in the upper p-type/lower n-type neutral region and the generated minority electron/hole randomly diffuses towards the multiplication region where it can eventually fire an avalanche.

a probability $P_{tr}(x_p)$. For this purpose it is convenient to define the minority carrier collection efficiency $\eta_{n-collection}(x)$ as the probability that an electron generated at position x in the p-type neutral region effectively reaches the SPAD multiplication region. By accounting for the probability to generate an EHP, i.e. a minority electron, within x and $x + dx$ in the p-type neutral region, as described in Eq. (5), the photon detection probability under this scenario can be expressed by (7):

$$PDP_{(b)} = P_{tr}(x_p) \int \eta_{n-collection}(x) p_{abs}(x) dx \quad (7)$$

where the integral has to be calculated over the entire p-type neutral region. A similar discussion can be done for scenario (c), where the neutral region is now n-type and the minority carrier is a hole:

$$PDP_{(c)} = P_{tr}(x_n) \int \eta_{p-collection}(x) p_{abs}(x) dx \quad (8)$$

Some considerations are necessary for the calculation of the minority carrier collection efficiency. With respect to scenario (b), for instance, a minority electron generated in the neutral p-type region would move along a random walk in absence of a strong electric field. Such a random walk might end at the upper boundary of the SPAD multiplication region, where the carrier can eventually initiate an avalanche process. The minority electron might also recombine with a hole along its path or at the silicon – BOX interface, but these two latter eventualities turn out to be negligible. Indeed, as discussed in Section 3.1, typical minority carrier lifetimes range from 1 ms to 1 μ s depending on silicon purity, and fall down to about 10 ns only in case of extremely high doping concentration values. Since the diffusion process lasts at most only a couple of nanoseconds, the probability that a generated minority carrier recombines in the neutral region becomes negligible. Moreover the presence of high concentration of defects at the BOX interface would appear as a flux of minority carriers toward the surface, which can be interpreted as the probability to lose a minority carrier due to surface recombination. However, as discussed more in detail in Ref. [8], this effect can be neglected even by adopting unreasonably high values of surface recombination velocity. The silicon dioxide interface imposes thus a (practically) zero flux of carriers, since they cannot escape the insulator. This means that every generated minority electron successfully reaches the multi-

plication region thanks to the oxide barrier at the BOX interface and thanks to the absence of any recombination phenomena within the diffusive time-lapse. For this reason, the minority carrier collection efficiency in the p-type neutral region can be estimated as being 100%. Something different happens in scenario (c). The generated minority hole is, in this case, “confined” between the multiplication region and the “inactive” SCR, both acting as collecting centers. For this reason the carrier collection efficiency is less than 100%. According to [8], this latter can be calculated by solving the drift-diffusion equation for the hole in the n-type neutral region, by assuming an initial injection of minority carriers corresponding to the photon absorption probability density all over the region. In this way, it is possible to extract the probability that a carrier generated within x and $x + dx$ successfully crosses at a given time the lower end of the multiplication region. The collection efficiency is finally obtained by integrating such a probability over the time. A slightly different approach is proposed in this work, by considering only a localized pulse of light generating excess carriers in an n-type semiconductor [14], instead of the entire injected minority carriers distribution based on the photon absorption probability. Moreover, this approach assumes perfect neutrality in the n-type region and thus neglects the drift component in the transport equation. Naturally, as discussed previously, the generation – recombination term of the transport equation is neglected too, since it is really unlikely that minority carriers recombine during the very short diffusive time-lapse. This allowed extracting an analytical solution for the problem which is more convenient for the analysis and discussion of the results, and offers a great improvement in terms of computation complexity. Given a minority hole generated at the position x_{ph} in the neutral region at $t = 0$, the probability to find the carrier at x , after a time $t > 0$ is given by:

$$p_h(x, t) = \frac{1}{\sqrt{4\pi D_p t}} e^{-\frac{(x-x_{ph})^2}{4D_p t}} \quad (9)$$

where D_p is the hole diffusivity in the n-type region. This equation is in the form of a Gaussian normal distribution and describes the way a generated minority carrier diffuses away from the injection point x_{ph} . It is possible to calculate the probability that at time t the carrier crosses the lower edge of the multiplication region, in a form of probability current:

$$\phi_h(x_n, t) = -D_p \left. \frac{dp_h(x, t)}{dx} \right|_{x=x_n} = \frac{2D_p(x_n - x_{ph})}{\sqrt{4^3 \pi D_p^3 t^3}} e^{-\frac{(x_n - x_{ph})^2}{4D_p t}} \quad (10)$$

$$\phi_h(d, t) = \frac{2D_p d}{\sqrt{4^3 \pi D_p^3 t^3}} e^{-\frac{d^2}{4D_p t}}$$

where d has been defined as the distance between the multiplication region lower edge x_n , and the carrier injection point x_{ph} . The probability that the multiplication region successfully collects the hole is thus obtained by integrating the probability current flowing through the multiplication region lower boundary over the time, as follows:

$$\eta_{h\text{-collection}}(d) = P_h(d) = \int_0^\infty \phi_h(d, t) dt = \frac{1}{2} \quad (11)$$

Interpreting Eq. (11) with respect to scenario (c), it is possible to conclude that the carrier will diffuse towards the multiplication region with a 50% probability, wherever the injection point is. For this reason, the minority carrier collection efficiency in the n-type neutral region can be estimated as being 50%. This result should represent a good first order estimation of the collection efficiency. It is however important to point out that in reality a weak electric field is normally present in the neutral regions due to the gradient of the doping concentration. The neglected drift component in the transport equation might therefore lead to a slightly different result in the collection probability.

The overall photon detection probability can be finally calculated by accounting for the three different scenarios/regions:

$$PDP(\lambda) = T(\lambda)(PDP_a(\lambda) + PDP_b(\lambda) + PDP_c(\lambda))$$

For the sake of ease, the present study considers a perfect anti-reflective coating (ARC) providing a reflection coefficient $R = 0$ (i.e. a transmission coefficient $T = 1 - R = 1$).

$$PDP(\lambda) = \int P_{tr}(x) p_{abs}(x, \lambda) dx + P_{tr}(x_p) \int \eta_{n\text{-collection}}(x) p_{abs}(x, \lambda) dx$$

$$+ P_{tr}(x_n) \int \eta_{p\text{-collection}}(x) p_{abs}(x, \lambda) dx$$

$$PDP(\lambda) = e^{-\alpha(\lambda)t_{sub}} (e^{-\alpha(\lambda)W_n} \int_0^{W_{SCR}} P_p(x') e^{-\alpha(\lambda)x'} \alpha(\lambda) dx' + P_{tr}(x_n) \frac{1}{2} (1 - e^{-\alpha(\lambda)W_n}) + P_{tr}(x_p) e^{-\alpha(\lambda)(W_n + W_{SCR})} (1 - e^{-\alpha(\lambda)W_p})) \quad (12)$$

where W_n, W_p, W_{SCR} are the n/p-type and multiplication region widths and t_{sub} is the p-substrate thickness.

4. Results and discussions

The simulation results for the novel 3D SPAD pixel proposed in this work are here presented and discussed. In Section 4.1, the avalanche diode architecture has been first validated for Geiger-mode operation by means of an electrostatic analysis of the adopted geometry, before proceeding with further studies on the device. The main figures of the proposed SPAD are subsequently evaluated: the breakdown voltage (V_{bd}) and the avalanche triggering probability (ATP) in Section 4.2, the dark count rate (DCR) in Section 4.3 and the photon detection probability (PDP) in Section 4.4.

4.1. Electrostatic Analysis of the avalanche diode

An electrostatic analysis on the proposed pixel has been conducted in order to validate the avalanche diode architecture for a correct Geiger-mode operation. Fig. 7 shows the electric field color map of the pixel when the avalanche diode is reverse biased at $V_{rev} = 16.5$ V, with grounded anode and substrate. The avalanche

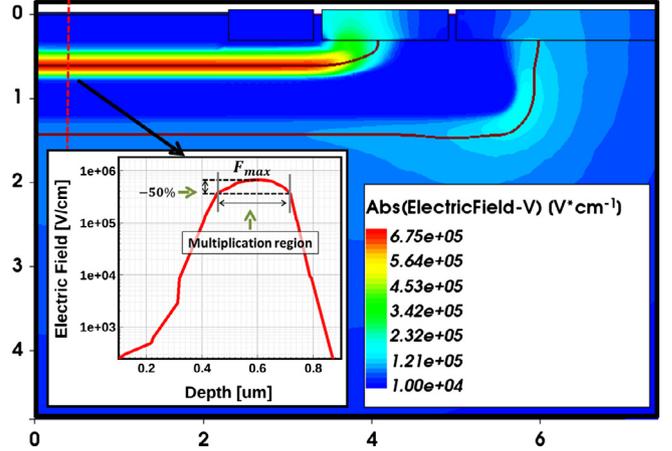


Fig. 7. TCAD simulation: Electric field color map of the pixel, when the avalanche diode is reverse biased at $V_{rev} = 16.5$ V (Spatial scales are in μm).

generation model has been switched off in order to reproduce the electric field distribution as it would be right before the start of the multiplication process, after which the field would be affected by the high amount of charges flowing through the junction. A uniform electric field (red region) is observed in the pixel active region, i.e. all along the horizontal direction in the p-well/deep n-well junction, which ensures a uniform avalanche triggering probability in the sensitive volume. By moving towards the multiplication region periphery, the field drops gradually to lower values, thanks to the retrograde n-type doping in the deep n-well, effectively acting as guard-ring for the device. For this reason, Premature Edge Breakdown (PEB) cannot occur, and Geiger-mode compatibility for the diode architecture can be assessed.

4.2. Breakdown voltage and avalanche triggering probability

The breakdown voltage (V_{bd}) of the avalanche diode in the proposed 3D pixel has been thus extracted in order to define the threshold above which the SPAD can work in the Geiger-mode. For this purpose, the reverse bias $I-V$ curve of avalanche diode directly obtained from TCAD simulation and the average avalanche triggering probability (ATP) as a function of the reverse bias are plotted in Fig. 8 (right and left y-axes respectively). It is worth noticing how both curves led, approximatively, to the same V_{bd} value. The breakdown voltage observed in the $I-V$ curve ($V_{bd} = 13.5$ V) is indeed very close to the voltage corresponding to the threshold for an $ATP > 0\%$ ($V_{th-ATP} \approx 13.4$ V). This corroborates the validity of the adopted post-processing method presented in Section 3.2 for the extraction of the ATP. In order to implement this latter, the ionization coefficient $\alpha_e(x)$ and $\alpha_h(x)$ (but also other parameters discussed later) have been extracted from a cut-line within the device active region, along the vertical direction with respect to the geometry shown in Fig. 5. Provided that the cutline is taken at a horizontal position sufficiently distant from the guard-ring, it is indeed possible to consider the ionization coefficients to be constant along the horizontal direction (1D symmetry). Moreover it is important to point out that the TCAD simulator provides ionization coefficients based on measurements where the electric field is assumed to be constant all over the ionization path [8]. This condition is generally not true for SPAD realized in very deep sub-micrometer CMOS technology due to the typically abrupt and highly doped pn junction defining the avalanche multiplication region, which results in a sharp and triangular shaped electric field along the diode SCR. However the SPAD proposed in the present study could be considered to satisfy (to some extent) such an

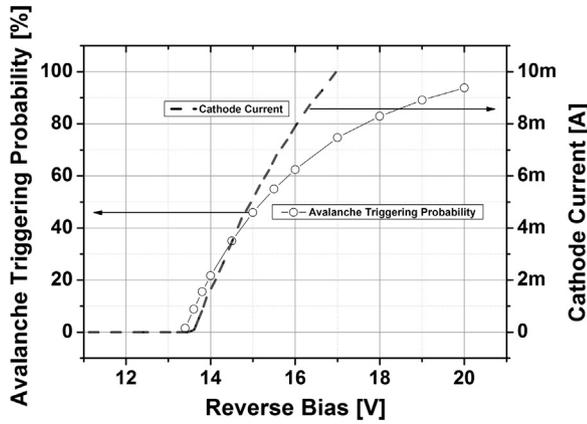


Fig. 8. Symbols, left y-axis: average avalanche triggering probability calculated according to the method proposed in [13]. Dashes, right y-axis: reverse bias $I-V$ curve of the avalanche diode resulting from TCAD simulation. Observe that the SPAD active area diameter is $D = 7 \mu\text{m}$.

assumption thanks to the “linearly graded” p-well / deep n-well junction, leading to a quite smooth electric field distribution within the multiplication region (see insert in Fig. 7).

4.3. Dark count rate

The knowledge of the ATP as a function of the position in the diode multiplication region, i.e. $P_{tr}(x)$, allows the calculation of the DCR as well as the Photon Detection Probability (PDP) for the SPAD under study. According to Eq. (4), DCR calculation requires the extraction of the EHP generation rates as a function of the position x within the diode SCR. This has been done for several excess biases V_{ex} ($V_{ex} = V_{rev} - V_{bd}$) above the breakdown voltage, within the range 0–6.5 V. Fig. 9 shows, for instance, the two main EHP generation mechanisms occurring in the avalanche diode under the same bias adopted for Fig. 7 ($V_{ex} = 3 \text{ V}$). SRH generation seems to be the dominant generation mechanism in the avalanche diode sensitive region. Even if tunneling generation peaks to a value very close to the doping independent SRH generation, the former is very narrowly distributed along the junction, leading to a minor contribution to the overall EHP generation.

Naturally field-enhanced generation process would anyway increase for higher excess bias values, and would likely become the dominant contributors in the DCR. Moreover, as discussed in Section 3.1, the SRH generation has been evaluated based on two different scenarios, i.e. by assuming either doping dependent or constant carrier lifetime for the electrons and holes generated in the SCR. The first scenario however led (*a posteriori*) to a quite high DCR which seemed unrealistic with respect to experimental data found in literature for avalanche diodes having similar breakdown voltages [4]. For this reason only the second scenario has been considered for the DCR calculation reported in Fig. 10 in terms of counts per unit surface, as a function of the applied excess bias. The DCR produced individually by the two EHP generation mechanisms have been plotted too in Fig. 10 in order to highlight the contribution provided by each of them on the total counts. It is possible to distinguish a dominating SRH region for excess bias lower than 4 V and a dominating tunneling region for voltages higher than 5 V, depending on the strongest EHP generation mechanism. In the SRH region, the DCR is indeed relatively moderate and increases with the excess bias according to the ATP enhancement shown in Fig. 8 (symbols, left y-axis). On the other hand, this latter is not really influent in the tunneling region (ATP is >80% and slowly saturates to 100%) where the DCR rises really fast with the excess bias mainly because of field-enhanced carrier generation. It

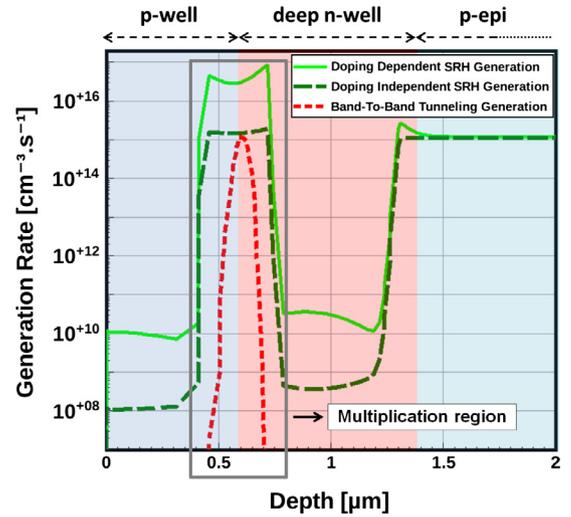


Fig. 9. TCAD simulation: EHP generation mechanisms in the multiplication region of the diode at $V_{rev} = 16.5 \text{ V}$.

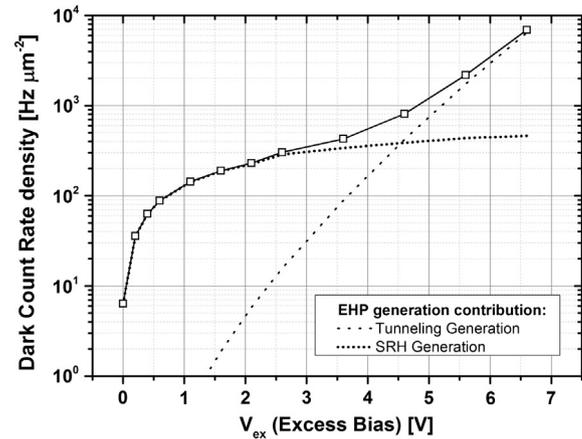


Fig. 10. Total DCR per unit surface for the simulated avalanche diode (and contributions of SRH and band to-band generation mechanisms).

is worth noticing that if different carrier lifetimes for the electrons and holes were adopted in the SRH model, a similar curve would have been obtained in Fig. 10. Assuming for instance that shorter carrier lifetimes with respect to those used in this study have been adopted, a higher DCR would be observed in the SRH region (due to a higher SRH generation) and the corner defining the transition between the SRH and the tunneling regions would be simply shifted towards higher excess bias values. An opposite discussion would stand in case of longer carrier lifetimes. Finally, it is really hard to make a more precise conclusion without any experimental data coming from direct measurements of the device.

Nevertheless, it is interesting to observe that the results shown in Fig. 10 look quite close to what can be found in literature for avalanche diodes having similar breakdown voltages, i.e. $\text{DCR} \sim 150 \text{ Hz}/\mu\text{m}^2$ for $V_{ex} = 1 \text{ V}$ [4].

4.4. Photon detection probability

The Back-Side Illuminated (BSI) Photon Detection Probability (PDP) for the proposed 3D SPAD pixel has been calculated with the help of (12) within the wavelength range [400–1200 nm]. The optical data for the photon absorption coefficient in silicon $\alpha(\lambda)$ have been taken from Ref. [15]. Fig. 11 shows the PDP calcu-

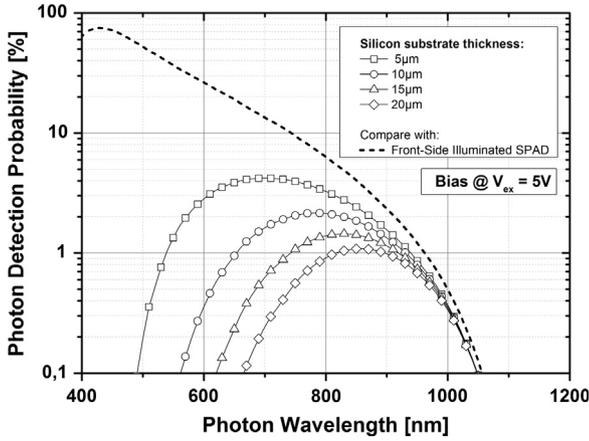


Fig. 11. Photon detection probability for different substrate thicknesses with back-side illumination (and comparison with front-side illumination).

lated when an excess bias of 5 V is applied. Several thickness values of the p-type substrate, down to 5 μm have been considered. Even though this latter value might appear extremely small and quite unrealistic, it is worth noticing that ultra-thinning process down to 4 μm over a 300 mm wafer has been actually reported in literature, i.e. Ref. [16]. The author focused on the feasibility for multi-stack Wafer-on-Wafer (WOW) processes and multi-stacking for Tera-scale high density memory, but the results obtained in his study might be extended to other applications in the near future. It is thus interesting to investigate the BSI performance of a SPAD under this scenario too. According to Fig. 11, the PDP in BSI mode (symbols) looks quite small with respect to what can be obtained in a front-side illuminated mode (dashes), especially in the short wavelength range. At longer wavelengths, conversely, the curves tend to converge towards a common trend. Fig. 11 can be better understood with the help of an approximated form of (12):

$$PDP(\lambda) \approx \underbrace{e^{-\alpha(\lambda)t_{sub}}}_A \underbrace{\alpha(\lambda) \left(\frac{1}{2} P_{tr}(x_n) W_n + P_{tr}(x_p) W_p + \overline{P_{tr}} W_{SCR} \right)}_B \quad (13)$$

where W_n, W_p, W_{SCR} are the n/p-type and multiplication region widths, t_{sub} is the substrate thickness and, according to (3), $\overline{P_{tr}}$ is the average ATP over the multiplication region.

The approximation arises from the fact that the photon penetration depth is much larger than the overall active region width, i.e. $1/\alpha(\lambda) \gg (W_n + W_p + W_{SCR})$ over the considered wavelength range which allows a first-order Taylor approximation of the exponentials in (12). Term “A” in Eq. (13) represents the probability that a photon impinging on the back-side of the pixel, successfully reaches the lower edge of the active region. Term “B” represents conversely the probability that a photon entering the active region is here absorbed and successfully fires an avalanche. It is thus apparent that the PDP degradation on the BSI mode is due to the “A” term, which accounts in fact for the optical losses into the p-substrate. In BSI mode, shorter wavelength photons are indeed mostly absorbed within the first few hundreds of nanometers of the silicon substrate due to their short penetration depth. That is why thinner substrates provide higher PDP within this wavelength range. When the penetration depth approaches the substrate thickness, i.e. $1/\alpha(\lambda) = t_{substrate}$, PDP reaches a maximum. Since $1/\alpha(\lambda)$ increases for longer wavelength, the peak moves to the right for thicker substrates. For longer wavelengths the penetration depth becomes larger which leads the “A” term to be negligible ($A \rightarrow 1$). That explains the convergence of the BSI curves with the

FSI one towards a common trend with respect to the photon wavelength. The “B” term is conversely determined by the electrical and geometrical properties of the SPAD active region. While these latter depends on the avalanche diode architectures and ultimately on the adopted CMOS process, the electrical properties can be controlled by acting on the device excess bias voltage, which affects the avalanche triggering probability, as discussed in Section 3. A little improvement on the PDP can arise from the light reflection at the p-well/BOX interface thanks to the “multi-layer” film defined by the Silicon/BOX/SOI stack (see insert in Figs. 5 and 12). Based on the same considerations made for the approximation of (12) with (13), the overall efficiency can be expressed as follows (14):

$$PDP_{tot}(\lambda) = PDP(1 + R(\lambda)) \quad (14)$$

where $R(\lambda)$ is the reflection coefficient at the silicon/BOX interface, experienced by a photon travelling from the silicon substrate towards the “multilayer” BOX / SOI thin film structure, below the interconnect layer (a dielectric medium). In order to provide a first order estimation of such an improvement, the low-k dielectric medium has been simply modeled with air, but a proper modeling would be necessary for more accurate results. The reflectance under this scenario is thus reported in Fig. 12 [17]. It is worth noticing that within the longer wavelength range, where the BSI PDP would become less dependent on the substrate thickness and would also benefit of reflection enhancement higher than 20%, BSI SPADs can really provide better performances than FSI ones. The scenario considered in Fig. 11 for the FSI case is indeed very “optimistic” since it does not take into account the much lower fill-factor of a FSI pixel (since the electronics would sit next to the avalanche diode) but also the photon reflections due to the interconnect layer of the chip as well as the SOI-BOX thin film structure. All of this translates into an important degradation of the FSI PDP.

The BSI 3D SPAD pixel proposed in this work would be suitable, for instance, for the realization of commercially available time-of-flight (T-o-F) ranging sensors working at $\lambda = 940$ nm. These devices are typically used for accurate measurements of the distance from a target object, in camera assist (ultra-fast autofocus and depth map), user detection for power saving in smartphones or laptops, gesture control, drones, robotics and many other applications [18]. As an interesting case study, Fig. 13 shows the PDP for a BSI SPAD as a function of the substrate thickness, at $\lambda = 940$ nm (T-o-F ranging sensors compatible) when an excess bias of 5V is applied on the avalanche diode. The grey curves propose an interesting comparison with an FSI SPAD, in the ideal case (solid line),

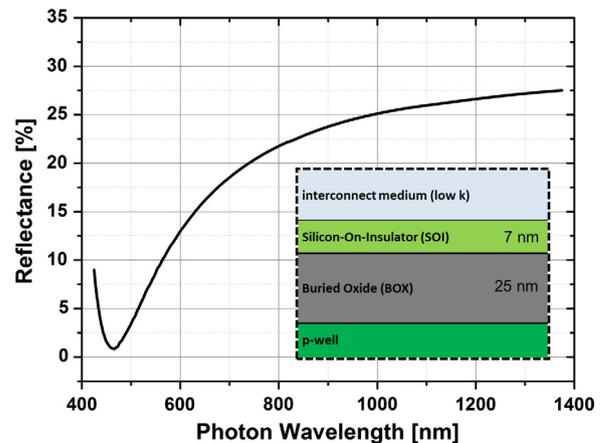


Fig. 12. Reflection coefficient at the Si/BOX interface for a photon travelling from the silicon substrate towards the “multilayer” BOX/SOI thin film structure, below the interconnect layer.

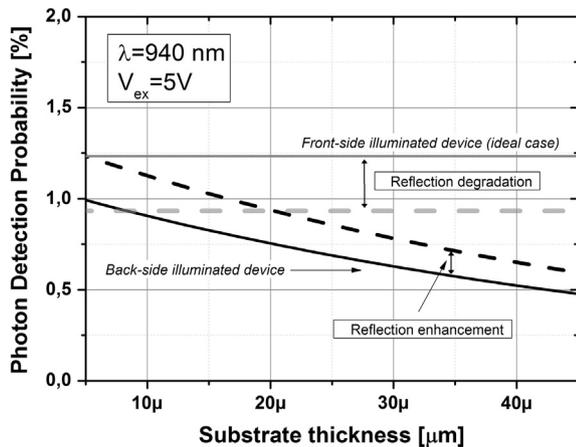


Fig. 13. Photon detection probability for a wavelength of 940 nm (time-of-flight ranging applications) and comparisons between back-side illumination and front-side illumination (with or without reflection).

i.e. like the one considered in Fig. 11, and when only the reflection degradation due to the BOX-SOI multilayer thin film is taken into account (dashes).

As expected, despite of the presence of a silicon substrate introducing optical losses, the BSI efficiency remains within the same percentage order of the “optimistic” FSI one. More interestingly, if reflection effects are taken into account (dashed curves), the BSI SPAD would perform even better than the FSI one, up to a substrate thickness of 20 μm . As discussed previously, FSI PDP degradation would be much larger than what has been estimated in Fig. 13. This would furtherly lower the dashed grey curve and would thus displace the intersection point with the BSI PDP towards thicker substrate, meaning that BSI mode would be definitely preferable for this kind of application.

5. Conclusions

In this work, a novel 3D SPAD pixel architecture consisting of an avalanche diode with associated electronics has been presented and studied by means of TCAD simulations and dedicated numerical methods for post-processing analysis. The pixel has been conceived for advanced Fully-Depleted Silicon-On-Insulator (FDSOI) CMOS technology and it is suitable for the detection of ionizing particles as well as light (visible and near-infrared range) in back-side illumination (BSI) mode. TCAD simulations based on realistic process parameters provided an estimation of the avalanche diode breakdown voltage (V_{bd}) and allowed validating the diode architecture for a correct Geiger-mode operation by studying the electric field distribution all over the pixel. They also provided the parameters required by the post-processing analysis in order to extract

the main figures of merit for a SPAD, such as the avalanche triggering probability P_{tr} , the Dark Count Rate (DCR) and the Photon Detection Probability (PDP). A comparison between the efficiency in back-side and front-side illuminated SPADs showed that the former approach can provide better performance within the long wavelength spectrum range. Based on the obtained results, it is possible to conclude that the BSI 3D SPAD pixel proposed in this work can be a very good candidate single-photon detector for time-of-flight (T-o-F) ranging sensors, and more generally for long wavelength sensing applications.

Acknowledgments

This project has been funded from the European Union’s 7th Framework Program for research, technological development and demonstration under grant agreement n° 317446.

References

- [1] Cova S et al. Avalanche photodiodes and quenching circuits for single-photon detection. *App Opt* 1996;35(12).
- [2] Vilella E et al. A gated single photon avalanche diode array fabricated in a conventional CMOS process for triggered systems. *Sens Actuat A: Phys* October 2012;186:163–8.
- [3] Zou Y, et al. Backside illuminated wafer-to-wafer bonding single photon avalanche diode array. In: 10th conference on Ph.D. Research in Microelectronics and Electronics (PRIME); 2014.
- [4] Lee MJ et al. A first single-photon avalanche diode fabricated in standard SOI CMOS technology with a full characterization of the device. *Opt Exp* 2015;23(10).
- [5] Sun P et al. A flexible ultrathin-body single-photon avalanche diode with dual-side illumination. *IEEE J Sel Top Quantum Electron* 2014;20(6):276–83.
- [6] Masetti G et al. Modeling of carrier mobility against carrier concentration in arsenic, phosphorus, and boron-doped silicon. *IEEE Transact Electron Dev* 1983;ED-30(7):764–9.
- [7] van Overstraeten R, de Man H. Measurement of the ionization rates in diffused silicon p-n junctions. *Solid-State Electron* 1970;13(1):583–608.
- [8] Gulinati et al. Modeling photon detection efficiency and temporal response of single photon avalanche diodes. In: *Proc SPIE 7355, photon counting applications, quantum optics, and quantum information transfer and processing II*, 73550X; May 2009.
- [9] Roulston DJ et al. Modeling and measurement of minority-carrier lifetime versus doping in diffused layers of n + p silicon diodes. *IEEE Transact Electron Dev* 1982;ED-29(2):284–91.
- [10] Sentaurus Device User Guide; 2013.
- [11] Schenk A. Rigorous theory and simplified model of the band-to-band tunneling in silicon. *Solid-State Electron* 1993;36(1):19–34.
- [12] Oldham WG, Samuelson RR, Antognetti P. Triggering phenomena in avalanche diodes. *IEEE Trans Electron Devices* 1972;19(9):1056–60.
- [13] McIntyre RJ. On the avalanche initiation probability of avalanche diodes above the breakdown voltage. *IEEE Transact Electron Dev* 1973;ED-20(7):637–41.
- [14] Sze S, Ng K. *Physics of semiconductor devices*, 3rd ed.; November 2006. ISBN: 978-0-471-14323-9.
- [15] Green MA, Keevers M. Optical properties of intrinsic silicon at 300 K. *Progress Photovol* 1995;3(3):189–92.
- [16] Kim YS. Ultra thinning down to 4- μm using 300-mm wafer proven by 40-nm node 2Gb DRAM for 3D multi-stack WOW applications. *VLSI Technology (VLSI-Technology): Digest of Technical Papers*; June 2014. p. 1–2.
- [17] <http://www.filmetrics.com/reflection-calculator>.
- [18] World smallest Time-of-Flight (ToF) laser ranging sensor - VL53L0X. <http://www2.st.com/resource/en/data_brief/vl53l0x.pdf>.



Reconfigurable ultra-thin film GDNMOS device for ESD protection in 28 nm FD-SOI technology



Sotirios Athanasiou^{a,b,*}, Charles-Alexandre Legrand^a, Sorin Cristoloveanu^b, Philippe Galy^a

^aSTMicroelectronics, 850 rue Jean Monnet, 38920 Crolles, France

^bIMEP, 3 Parvis Louis Néel, CS 50257, 38016 Grenoble Cedex 1, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

CMOS
Electrostatic discharges
ESD
FD-SOI
Gated diode
LBJT
MOSFET
SOI
Thyristor

ABSTRACT

We propose a novel ESD protection device (GDNMOS: Gated Diode merged NMOS) fabricated with 28 nm UTBB FD-SOI high-k metal gate technology. By modifying the combination of the diode and transistor gate stacks, the robustness of the device is optimized, achieving a maximum breakdown voltage (V_{BR}) of 4.9 V. In addition, modifications of the gate length modulate the trigger voltage (V_{T1}) with a minimum value of 3.5 V. Variable electrostatic doping (gate-induced) in diode and transistor body enables reconfigurable operation. A lower doping of the base enhances the bipolar gain, leading to thyristor behavior. This innovative architecture demonstrates excellent capability for high-voltage protection while maintaining a latch-up free behavior.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recent advances in Fully Depleted SOI (FD-SOI) technology [1] continue to shrink the design window [2] of ElectroStatic Discharge (ESD) protection devices, reducing supply voltage V_{DD} and decreasing breakdown voltage V_{BR} . This window is highly dependent on the type of devices to be protected, as well as on the region of operation. Contrary to the typical hybridization techniques, where the thin silicon film and Buried Oxide (BOX) are etched in order to have direct access to the silicon substrate and fabricate bulk-like ESD devices [3], the focus of this work is to use the thin FD-SOI layer.

Previous studies have been conducted on thin film BiMOS device for protection of circuits operating in the typical voltage range for FD-SOI technology [4–6]. In a BiMOS, the lateral bipolar junction transistor (LBJT) is activated by current injection to the base, from a lateral contact, while the MOSFET is operating in sub-threshold region. The BiMOS gain can be modulated electrostatically through the top or bottom gate biasing. Alternatively band

modulation devices [7] such as the Z^2 FET show exceptional behavior as ESD protection elements in FD-SOI technology.

Typical strategy for a standard process flow is to reduce the number of steps required for a particular device, in order to minimize both the variability and the cost. Thus, structures that do not require extra process steps for hybridization or implantation (for example, no customized doping concentration) need to be explored in terms of power and switching performance. Normally, high voltage transistors can sustain several hundred volts, this is not the case with ultra-thin film and BOX (UTT) technology since the breakdown of top and bottom oxides imposes limitations on the power that a device can sustain.

In this work, we present an original structure (GDNMOS) for high voltage protection that combines diode and MOSFET mechanisms, a semi-regular structure that abides by process rules for limiting variability. Electrical measurements prove the flexibility, functionalization and robustness of the GDNMOS. 3D TCAD [8] simulations reveal the details of the operation mechanisms as well as the potential behavioral improvements.

In Section 2, the physical mechanisms involved are described, followed by the device structure and process. Detailed TCAD simulations to investigate the GDNMOS operation during an ESD stress are presented in Section 3. Finally, in Section 4 the fabricated struc-

* Corresponding author at: STMicroelectronics, 850 rue Jean Monnet, 38920 Crolles, France.

E-mail address: athanasiou.sotiris@gmail.com (S. Athanasiou).

tures are investigated by utilizing DC and Transmission Line Pulse (TLP) electrical measurements.

2. Device structure and properties

2.1. General operation mechanisms

GDNMOS is composed of a typical FD-SOI nMOSFET merged with an FD-SOI p-i-n gated diode, see Fig. 1. There is a common (merged) n-type area which acts as the diode cathode and as the MOSFET drain. The ultrathin bodies of the diode and transistor can be electrostatically doped (N⁺ or P⁺, by attracting different types of carriers in the ‘intrinsic’ region below the gate) via the gates or back-plane bias.

During a stress, in a thyristor-like device the trigger voltage (V_{t1}, see Fig. 2) is dependent on the sum of the common base current gain (α) of the two bipolar transistors. In our case the npn LBJT formed within the nMOSFET and pnp LBJT formed between the nMOSFET and the gated diode (Fig. 1). Starting from a typical n-p-n device the common-base current gain α can be defined as [9]:

$$\alpha = \frac{i_C}{i_E}, \alpha_{NPN} = \gamma_E * a_T * \gamma_C \quad (1)$$

where γ_E is the emitter efficiency, a_T is the base transport factor, and γ_C is the collector efficiency:

$$\gamma_E = \alpha_E = \frac{J_n(0)}{J_n(0) + J_p(0)} = \frac{D_{nB}L_{pE}N_{iB}^2N_{DE}}{D_{nB}L_{pE}N_{iB}^2N_{DE} + D_{pE}W_BN_{iE}^2N_{AB}} \quad (2)$$

L and D are the diffusion length and constant for a type of carrier, n_i is the intrinsic carrier density and N is the doping concentration. All of the aforementioned values are indicated with a subscript for a specific area of the BJT: E, Emitter, B, Base, C, Collector.

In order to increase the thyristor efficiency, we have to enhance the emitter efficiency for each BJT. This can be done by shrinking the gate length (by design) or by lowering the acceptor doping in base area (by biasing). Our main purpose is to take advantage of such reconfigurable behavior for emulating a PNP thyristor structure: positively biased P⁺ anode, floating N⁺ drain, P-MOSFET body, grounded N⁺ source (Fig. 1).

2.2. Operation mechanisms under ESD conditions

During standard operation, when the bias applied to an ESD protection device is lower than V_{DD}, the device should exhibit very low leakage current. The low leakage current is attributed to the reverse bias applied to the pn junction between Emitter and Base. Low leakage region is observed by increasing the anode voltage until the point defined by trigger voltage (V_{t1}) and trigger current (I_{t1}) (Fig. 2). When a higher voltage drop occurs, bipolar or thyristor-like effects are activated, and the device starts to conduct current. This behavior is caused by avalanche breakdown that will be discussed later. This operating regime continues until a point indicated as holding point with holding voltage (V_h) and holding

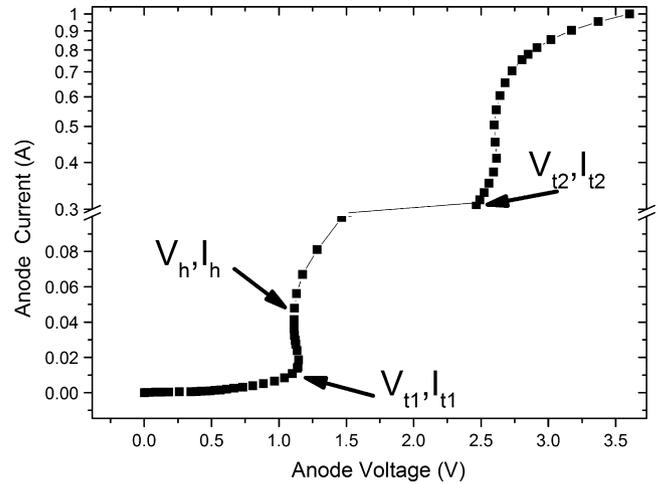


Fig. 2. Typical high current I-V curve of a protection device.

current (I_h). After this holding point, device enters the self-biased bipolar operating region until secondary (thermal) breakdown is reached. At this point, we define a failure voltage (V_{t2}) and a failure current (I_{t2}) beyond which the device behavior is permanently damaged.

2.3. Fabrication and process details

Test devices (Fig. 3) were fabricated with the 28 nm FD-SOI STMicroelectronics process featuring an ultra-thin silicon film of 7 nm, ultra-thin BOX of 25 nm, high-k metal gate stack and p-type backplane (p-BP). The devices were fabricated with different gate stack options (Fig. 3): Standard Gate (SG, EOT = 1.1 nm) and Extended Gate (EG, EOT = 3.4 nm). In DEV1, 2, 3 structures the MOS gate and the diode gate were interconnected, while in DEV4, 5, 6 they were independent. In the mixed gate stack devices, the gate oxide of the diode was always selected to be thick (EG), for increased robustness to the stress applied to anode.

Minimum process-compliant gate dimensions were selected for each gate stack combination. The variation of the gate length (45–150 nm) was due to the difference imposed in fabrication process for EG and SG gate stacks and the minimum distance between polysilicon structures, resulting in different minimum lengths for each case. All devices were fabricated in multi finger topology with total combined finger width W of 100 μm.

3. Simulation results

The 28 nm FD-SOI GDNMOS was meshed in 3D for different gate stack configurations. We consider phonon scattering, Cou-

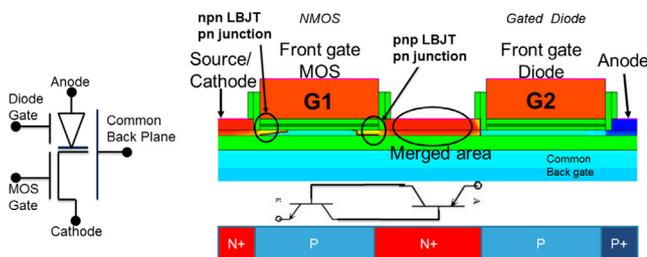


Fig. 1. Schematic (left) and cross-section (right) of GDNMOS structure.

	MOS Gate (G1)	Diode Gate (G2)	Lg1	Lg2	MG-DG2	Bias
DEV1	SG	SG	45nm	45nm	conn	2V
DEV2	SG	EG	30nm	150nm	conn	4V
DEV3	EG	EG	150nm	150nm	conn	4V
DEV4	SG	SG	45nm	45nm	separ	2V
DEV5	SG	EG	30nm	150nm	separ	4V
DEV6	EG	EG	150nm	150nm	separ	4V

Fig. 3. Different device configurations fabricated.

lomb scattering in doped materials, and mobility degradation due to high field saturation and high temperature (for thermodynamic simulations). Other active modules include avalanche model, Auger recombination model as well as dynamic non-local Band-to-Band tunneling model. In all cases Boltzmann statistics were used for electron and hole densities calculations.

For the simulation study, the surge is an Average Current Slope (ACS) stress with a ramp from 0 A to 1 A max current during 100 ns, which is equivalent to the transmission line pulse (TLP) test for human body model (HBM) [10].

3.1. Isothermal simulation

For isothermal simulations, the simulator is solving a system of equations [8] including Poisson:

$$\nabla(\varepsilon\nabla\varphi) = -q(p - n + N_D - N_A) \quad (3)$$

Contact equation:

$$\nabla(\sigma\nabla\Phi_M) = 0 \quad (4)$$

where σ is the metal conductivity and Φ_M the Fermi potential.

Electron and hole equations:

$$\vec{J}_n = -nq\mu_n\nabla\Phi_n \text{ and } \vec{J}_p = -nq\mu_p\nabla\Phi_p \quad (5)$$

where Φ_n and Φ_p are the quasi-Fermi potentials.

As well as circuit equations.

Boundary conditions assume all contacts on semiconductor materials to be Ohmic, subject to charge neutrality and equilibrium.

Fig. 4 shows typical I–V characteristics resulting from ACS simulation. A clear differentiation between the mixed SG-EG gate stacks as well as EG-EG gate stacks is observed as well as the effect of floating versus grounded/biased gate terminals. In the case of floating terminals, the MOSFET energy bands are affected only by the Fermi energy difference between metal (HKMG stack) and the silicon film, and through coupling to the backplane Fermi energy. Naturally the energy bands in silicon are bended slightly downwards and there is no major electrostatic modulation of the LBJT behavior.

For grounded gates, the bands are affected both by the Fermi energy difference and by the potential applied, resulting in differences in carrier concentration under the gate. The reason for applying these initial conditions is due to the fact that in real applications an ESD surge can happen on an integrated circuit even

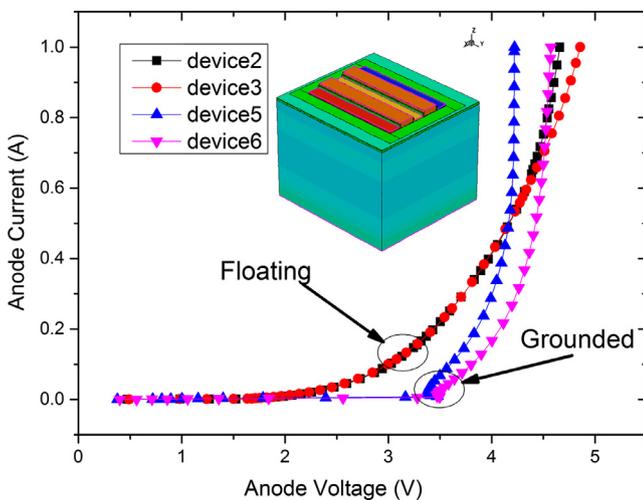


Fig. 4. ACS I–V 3D TCAD simulation for devices 2, 3, 5 and 6.

if there is no power which makes it challenging to apply a constant biasing to the gate terminals. Additionally with grounded and floating gates all the MOSFET effects are suppressed and we observe the real thyristor behavior. At the end of the stress, current density reaches the maximum value with conduction through the whole film (volume inversion, Fig. 5).

Fig. 6 shows the band diagram at the end of an ACS stress for grounded nMOSFET gate and diode gate. There is a significant bending of the bands due to the high voltage drop (more than 4 V). Also, we observe the difference due to using different gate stacks for the nMOSFET with different lengths. This band bending is especially pronounced between the merged area and the nMOSFET gate, leading to increased band to band tunneling (Fig. 7).

Avalanche breakdown serves as a trigger element to initiate the turn on of an ESD device. As the carrier energy approaches the ionization threshold, energy transfers from the carrier to the lattice. This causes impact ionization (II) which results in generation of more carriers and, ultimately, to avalanche breakdown [11]. During the ACS stress of our device (Fig. 8) we observe that maximum II generation is located at the pn junction between the base of the npn bipolar and the merged area. This area is the most potential candidate for creating an avalanche breakdown, and will be discussed later.

3.2. Electrothermal simulation

During an ESD event, energy is injected into the protection device. In the case of UTBB, it is imperative to study the self-heating effects in the device, primarily caused by the poor thermal diffusivity ($\alpha = \frac{\kappa}{\rho c_p}$) [11] (where C_p is the specific heat), of thick silicon dioxide (BOX) compared to bulk silicon. We introduce the lattice temperature equation and modify previous equations (4) and (5) to include the temperature gradients:

Lattice temperature:

$$\frac{\partial}{\partial t} C_l T - \nabla\kappa\nabla T = -\nabla[(P_n T + \Phi_n)J_n + (P_p T + \Phi_p)J_p] - \left(E_c + \frac{3}{2}kT\right)\nabla J_n - \left(E_v - \frac{3}{2}kT\right)\nabla J_p \quad (6)$$

where κ the thermal conductivity, c_l the lattice heat capacitance, P_n and P_p the thermoelectric power for electrons and holes which accounts for Seebeck effect.

Contact equation:

$$\nabla(\sigma\nabla\Phi_M + P\nabla T) = 0 \quad (7)$$

where P is the metal thermoelectric power.

Electron and hole equations:

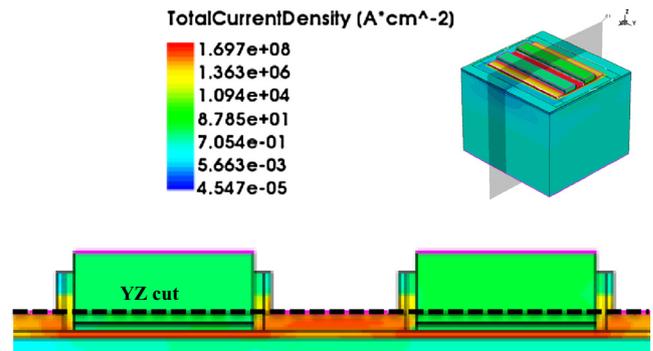


Fig. 5. Current density extracted at the end of ACS stress for device 6, grounded gates. Volume conduction is visible.

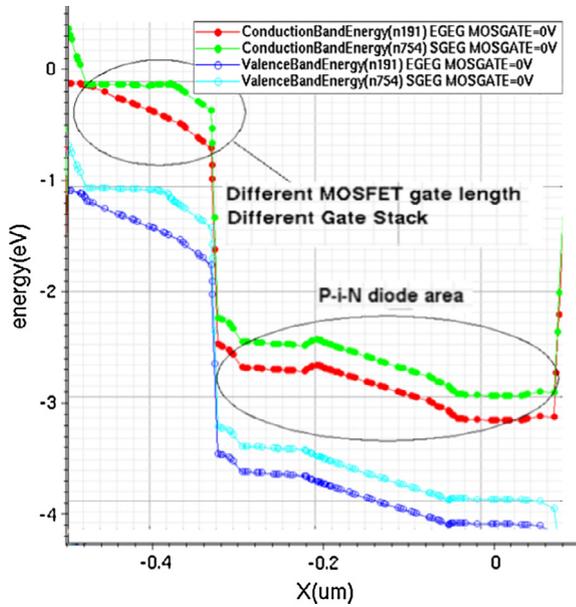


Fig. 6. Energy band diagram (YZ cut) at the end of ACS I–V TCAD simulation for devices 5 and 6.

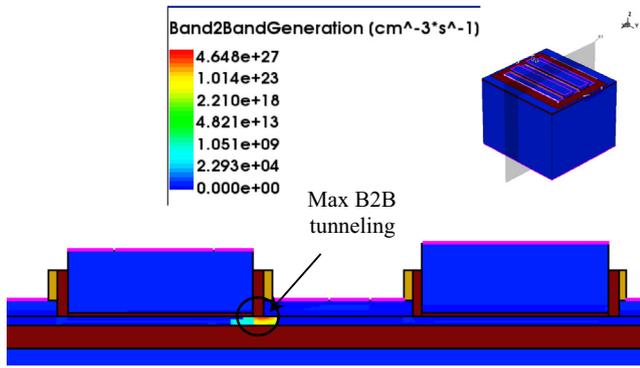


Fig. 7. Band to band tunneling generation at the end of ACS stress.

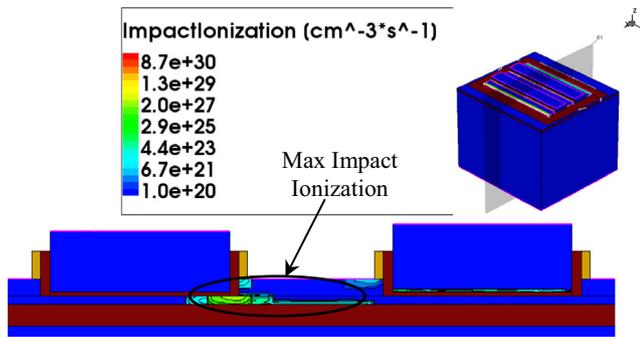


Fig. 8. Impact ionization during the ACS stress for device5 at time t1.

$$\vec{J}_n = -nq\mu_n(\nabla\Phi_n + P_n\nabla T) \text{ and } \vec{J}_p = -pq\mu_p(\nabla\Phi_p + P_p\nabla T) \quad (8)$$

and typical circuit equations.

Additionally a temperature boundary condition ($T = 300 \text{ K}$) was applied to the bulk silicon area below the BOX.

The temperature equation (6) as well the temperature gradients in Eqs. (7) and (8) are essential in the study of self-heating effects and electrothermal stability of the device. The electrothermal sta-

bility with the condition $\frac{\partial I}{\partial T}|_V = \frac{1}{V} \frac{dP}{dT}$ can cause the thermal breakdown of the device through the creation of mesoplasma states [11]. By performing temperature extraction, we note a significant difference between maximum and device average temperature (Fig. 9). This difference can be explained by the existence of a hot spot in the pn junction near the merged area (Fig. 10), the same area where the maximum impact ionization was observed.

4. Measurements and discussion

The devices were tested in DC and TLP [12] modes under different conditions.

4.1. DC measurements

The selection of gate stack affects strongly the breakdown voltage of the device (Fig. 11) with EG gate stack devices exhibiting higher breakdown voltage and lower leakage. On the other hand, the use of EG gate stack limits the minimum gate length to 150 nm something which strongly affects LBJT gain and in turn TLP behavior that will be discussed in the next paragraph. Breakdown voltage extraction was performed on multiple dies (Figs. 11–13). Devices 4, 5 and 6 exhibit superior performance and very good variability.

4.2. TLP measurements

TLP measurements had duration of 100 ns or 5 ns pulse width with native (300 ps) or 10 ns rise time. The stress was applied on anode, while cathode was grounded. During the first measurements, the gates were either grounded or floating. Our primary goal is to investigate the ESD behavior of the various structures and evaluate the efficiency of the floating-body (without base contact) lateral thyristor.

For the first type of devices (Fig. 14), [13] with the two gates connected, we observe snapback characteristics. Same trend is observed for in devices of second group that have separate MOS and diode gates (Fig. 15), [13], for identical biasing conditions. For these structures, there is no systematic differentiation between 100 ns and 5 ns TLP I–V curves. The trigger voltage reaches 5 V. No strong snapback behavior is observed primarily due to (i) the lack of direct control on the Emitter–Base p–n junctions of the Lateral Bipolar Junction Transistors (LBJT between source and drain of MOSFET), and (ii) the relatively high doping of LBJT base in the

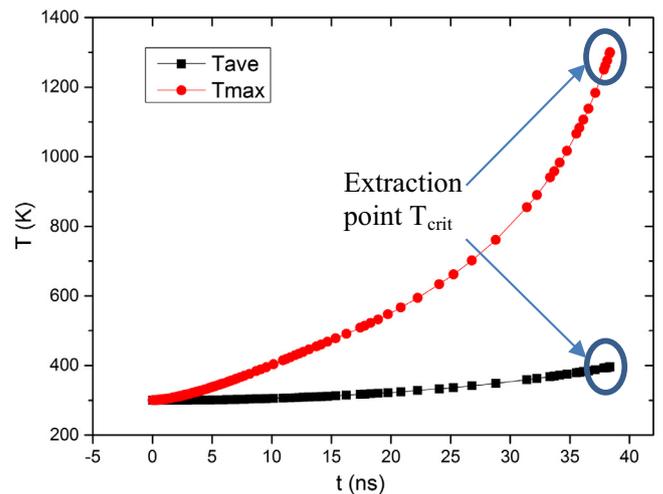


Fig. 9. Average and peak temperature during ESD stress due to self-heating.

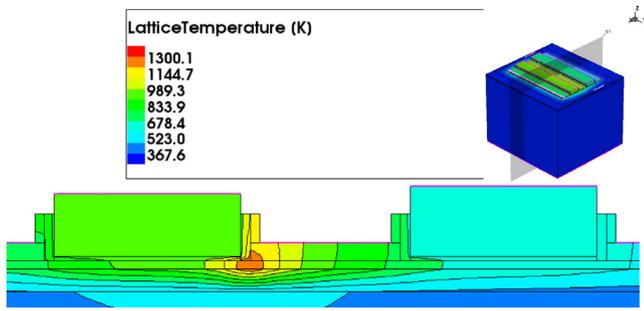


Fig. 10. 2D temperature extraction during ACS stress in thermodynamic simulation after 38 ns.

	V_{t1}	I_{t1}	I_{t2}	V_{hold}	V_{BR}
DEV1 GND	NA	NA	<20mA	NA	3.1V
DEV1 FLT	NA	NA	<180mA	NA	NA
DEV2 GND	NA	NA	<20mA	NA	4.2V
DEV2 FLT	NA	NA	<180mA	NA	NA
DEV3 GND	NA	NA	<20mA	NA	4.8V
DEV3 FLT	NA	NA	<120mA	NA	NA
DEV4 GND	4V	NA	>160mA	4V	4.3V
DEV4 FLT	4V	NA	>140mA	4V	NA
DEV5 GND	3.5V	NA	>160mA	3.5V	4.3V
DEV5 FLT	3.5V	NA	>140mA	3.5V	NA
DEV6 GND	4.5V	NA	>160mA	4.5V	4.9V
DEV6 FLT	4.3V	NA	>140mA	4.3V	NA

Fig. 11. Extracted results for different gate biasing and configurations. We observe the improved performance with implementations 4, 5 and 6. These devices exhibit higher I_{t2} , improving robustness during an ESD event.

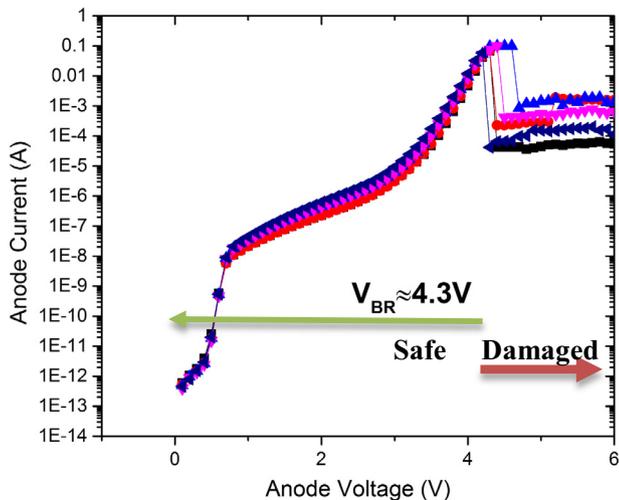


Fig. 12. Device 5 DC sweep with grounded G1 and G2 for breakdown voltage extraction at room temperature.

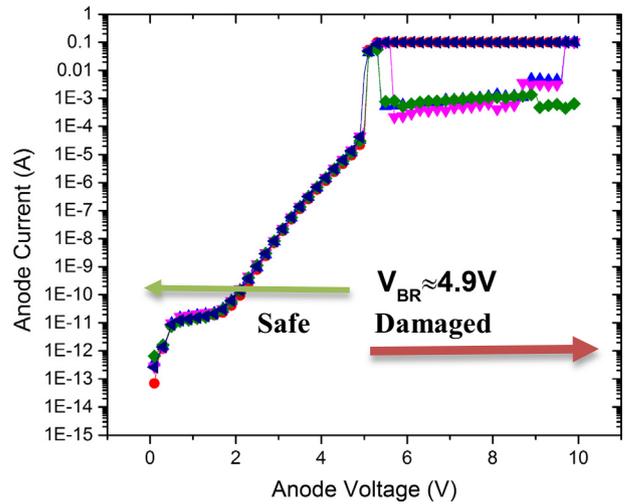


Fig. 13. Device 6 DC sweep with grounded G1 and G2 for breakdown voltage extraction at room temperature.

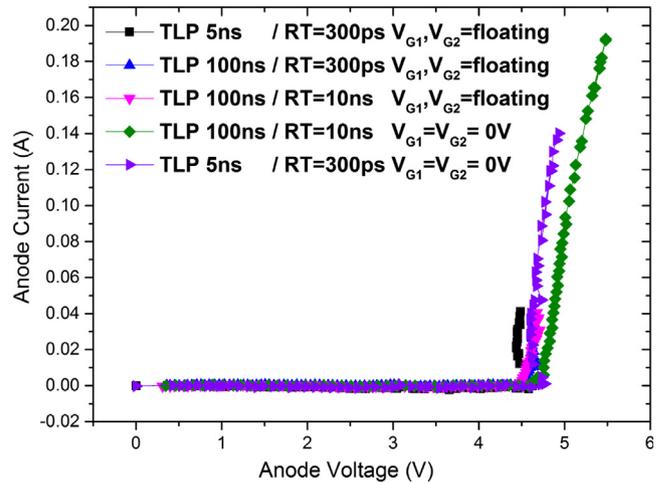


Fig. 14. Device 3 TLP measurements for different pulse width: 100 ns and 10 ns.

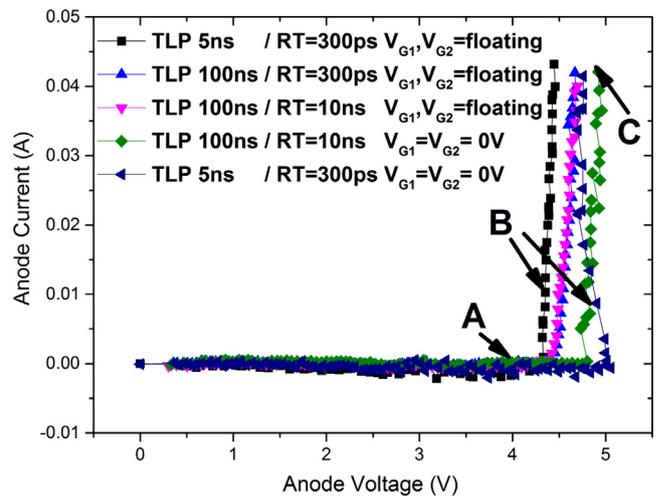


Fig. 15. Device 6 TLP measurements for different pulse width: 100 ns and 10 ns.

merged area. In terms of ESD protection capability, DEV5 and 6 provide the best performance with higher I_{t2} , lower leakage and higher breakdown voltage.

In Fig. 16 we see the chronogram extracted from the TLP measurements with grounded gates, at different current levels. Each TLP point is calculated as an average of voltage values between the two lines in Fig. 16. It is interesting to observe the over-voltage behavior during the TLP stress. The average voltage for the TLP curve is lower than the actual voltage applied to the device. In terms of reliability this means that for a limited time, when over-voltage peaks appear during the measurement, the structure is forced to a higher stress than the one anticipated.

4.3. Front and back gate effect

The device behavior during a TLP stress can be modified by biasing the MOS gate (G1). In particular, the gate bias modulates the nMOSFET LBJT gain. With positive bias for the nMOSFET gate, the energy bands in the base area are lowered (by $-qV_{G1}$). Electrons are attracted in the bipolar base, increasing the npn LBJT gain and the overall anode current (Fig. 17). This behavior is observed while we are in the nMOSFET subthreshold region ($V_{G1} < V_{T0}$). When biasing becomes higher than nMOSFET threshold voltage we activate additionally the MOS effect.

Reciprocally, by utilizing a negative gate bias the energy bands are raised. However, since there are no extra holes available to be attracted in the body (LBJT base), the modulation of I_A - V_A curves is weak while the gain is slightly decreased. These trends are verified by the TLP measurements presented in Fig. 18 (for $V_{G1} = 0$).

The n-type base of the pnp LBJT is located in the merged area so biasing the diode gate does not affect strongly the pnp gain. By applying a positive bias to the diode gate, the base width W_B of the pnp transistor expands under the gate and the gain is lowered. Conversely, for negative bias, it is the P⁺ area of the anode which extends underneath the gate, reducing the resistance of the silicon layer (under G2). Further analyzing Eq. (2), we note that the high doping level in the base is the prevailing term, resulting in a very low bipolar gain that can only marginally be affected by base width and resistance reduction. This behavior is confirmed by the TLP response of the device shown in Fig. 19: there is minor modulation of the TLP characteristics as a function of V_{G2} . The curves measured with $V_{G2} = -1$ V or $V_{G2} = +1$ V are superposed.

Another degree of flexibility is offered by the back gate (back-plane) biasing, albeit it is subject to super-coupling effects observed in UTBB FD-SOI process [14]. This is due to the existence of both type of carriers inside the device (holes and electrons) during thyristor mode biasing (when MOSFET effect is not active). A

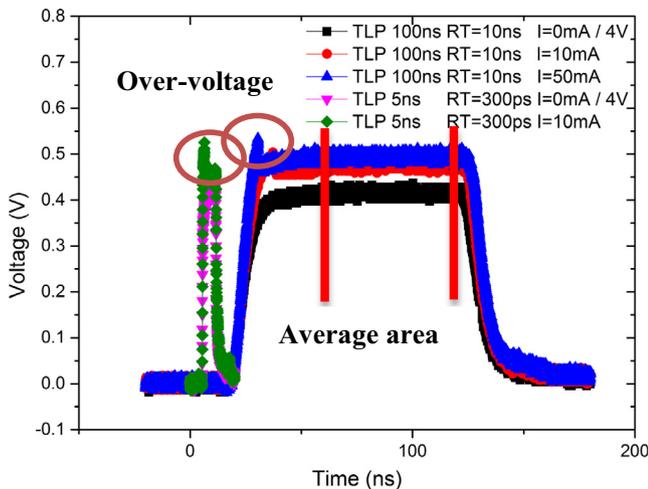


Fig. 16. Device 6 TLP chronograms at different current values as indicated in Fig. 15 for different pulse widths and rise times.

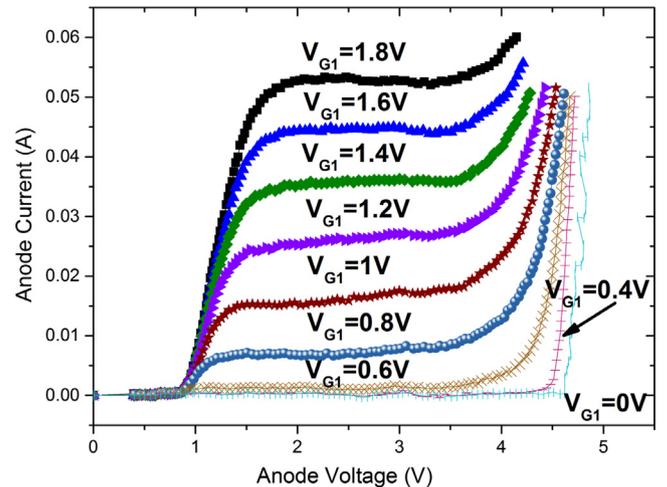


Fig. 17. Device 6. TLP 100 ns, RT 10 ns with different positive G1 biasing, $V_{G1} = 0$ V and $V_{BP} = 0$ V.

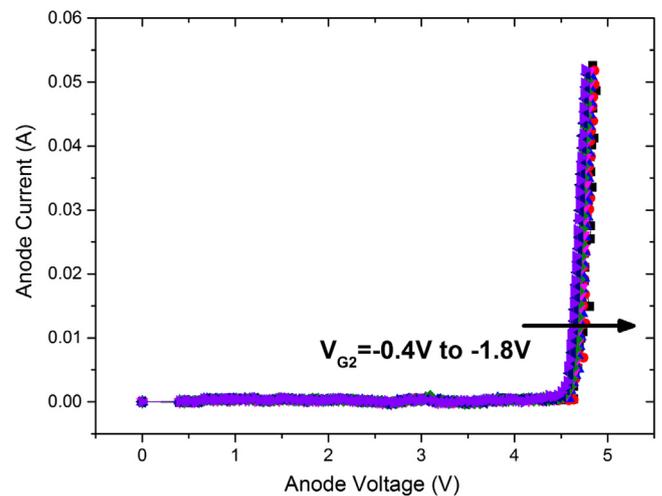


Fig. 18. Device 6. TLP 100 ns, RT 10 ns with different negative G2 biasing, $V_{G1} = 0$ V and $V_{BP} = 0$ V.

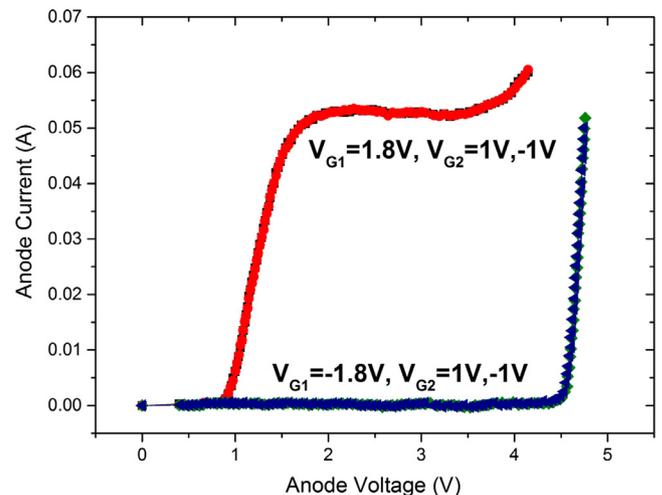


Fig. 19. Device 6. TLP 100 ns, RT 10 ns with different G2 biasing and $V_{BP} = 0$ V.

large bias is needed on the back plane in order to achieve the same results as using the front-gate biasing (Fig. 20 compared with Figs. 17 and 18). This is due to the difference between the front

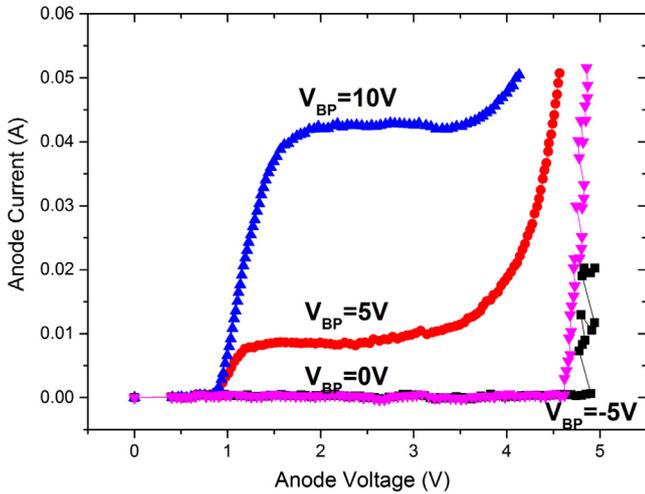


Fig. 20. Device 6. TLP 100 ns, RT 10 ns with different BP biasing for $V_{G1} = V_{G2} = 0 V$.

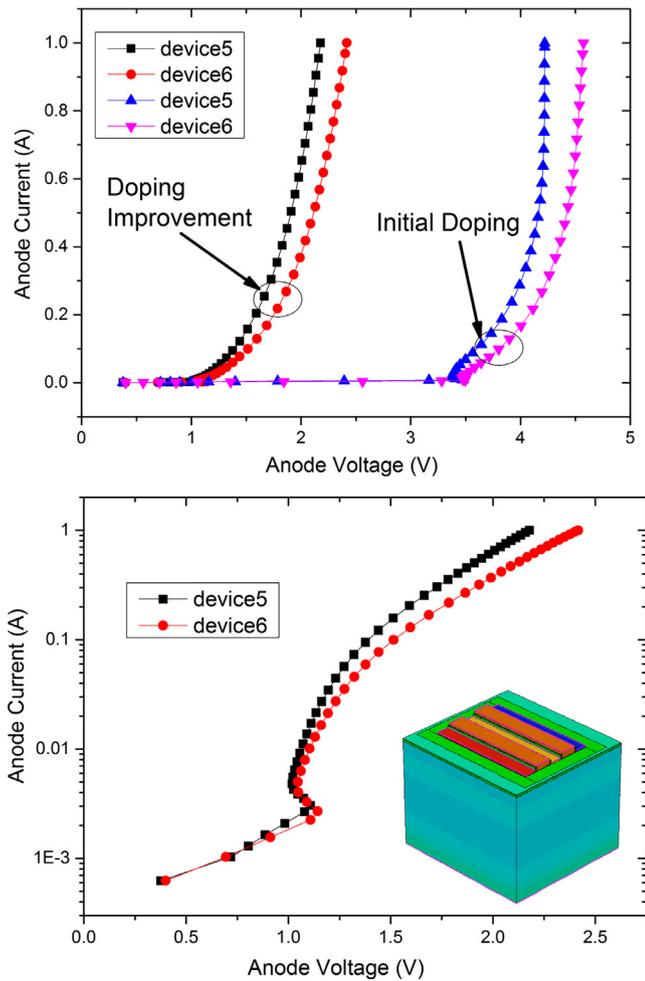


Fig. 21. ACS I-V TCAD simulation for devices 5 and 6. Top: Linear plot and comparison with previous solution, Bottom: Semi log plot.

gate oxide (EOT in the HKMG case) and the BOX oxide thickness with a coupling coefficient [15] $t_{OX}/t_{BOX} = 0.136$. The combined gain of the 2 BJTs is limited once more by the high doping in the merged area (base) of the pnp LBJT, resulting in a lower gain for the BJT formed with the gated diode. In the next section we will

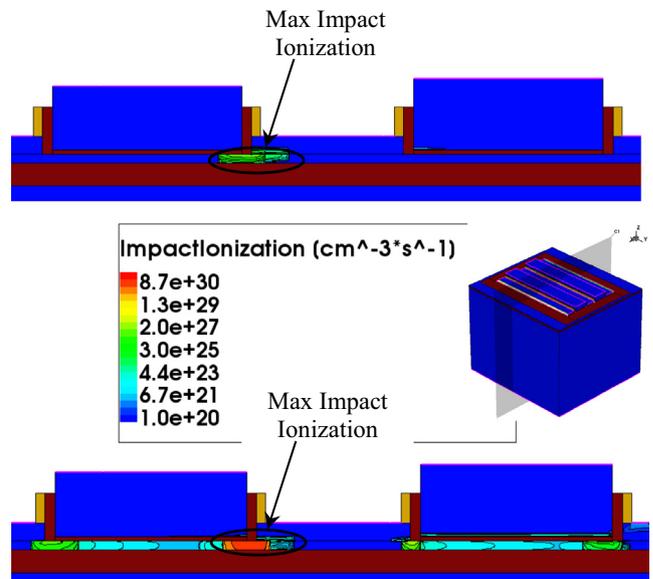


Fig. 22. Impact ionization during the ACS stress for device 5 with modified doping; at the trigger voltage point (t1) (TOP) at the end of the stress (BOTTOM).

show that doping calibration modulates the thyristor behavior leading to radically change the ESD stress response.

5. Thyristor behavior through doping modification

One of the critical parameters of the GDNMOS is the doping concentration in the merged area. This area acts like the base of the p-n-p LBJT of the thyristor with its doping and width reversely proportional to bipolar gain, as seen in Eq. (2). The doping profile selected for analysis in this section is compatible with the FD-SOI process and is equivalent to the LDD doping used in MOSFETS to reduce hot carrier effects by spreading the electric field [16]. TCAD simulations show a remarkable shift of around 2 V on the I-V curve with snapback behavior, when this LDD doping is utilized instead of the initial higher N^+ doping (Fig. 21). The reason for this shift is the increase in emitter efficiency and base transport factor of the gated diode LBJT for lower doping. Fig. 22 illustrates the impact ionization rate at the beginning and at the end of the stress. It is interesting to compare these results with those presented in Fig. 8 for the device with higher doping in merged area, where we observe impact ionization generation through a larger volume of silicon. With the new doping used the effect of impact ionization is induced in a smaller volume of silicon activating avalanche multiplication and triggering earlier with thyristor-like behavior.

6. Conclusion

In this work, we have introduced the ultrathin film GDNMOS device whose behavior was evaluated using both simulations and measurements on fabricated samples. No latch-up is observed, hence the device can be used for high-voltage protection. It benefits from simple integration on thin film and full process compliance. The ESD characteristics of GDNMOS can be further tuned by selecting a positive biasing for the nMOSFET gate as well as by changing the doping concentration of the merged area. The GDNMOS shows promising characteristics and remarkable versatility for adoption in the FD-SOI technology.

Acknowledgements

We would like to thank N. Planes for FD-SOI wafer fabrication and procurement, Dr. Olivier Bon and the rest of ST mask generation team for support. We would also like to thank our colleagues Dr. J. Bourgeat, B. Heitz and Dr. Nicolas Guitard for insight and expertise that proved beneficial for this project. Support from EU project WayToGoFast was highly appreciated.

References

- [1] Planes N et al. 28 nm FDSOI technology platform for high-speed low-voltage digital applications. In: Symposium on VLSI Technology (VLSIT), Jun 2012, Honolulu, US. p. 133–4. doi: <http://dx.doi.org/10.1109/VLSIT.2012.6242497>.
- [2] Benoist T et al. Experimental investigation of ESD design window for fully depleted SOI N-MOSFETs. *Microelectron Eng* 2011;88(7):1276–9. doi: <http://dx.doi.org/10.1016/j.mee.2011.03.112>.
- [3] Golanski D et al. First demonstration of a full 28 nm high-k/metal gate circuit transfer from Bulk to UTBB FDSOI technology through hybrid integration. In: Symposium on VLSI Technology (VLSIT). p. T124–5. ISSN: 2158-5601.
- [4] Galy Ph et al. BIMOS transistor in thin silicon film and new solutions for ESD protection in FDSOI UTBB CMOS technology. In: EurosoiUlis 2015 Italy. p. 29–32. <http://dx.doi.org/10.1109/UJIS.2015.7063765>.
- [5] Athanasiou S et al. Impact of back plane on the carrier mobility in 28 nm UTBB FDSOI devices, for ESD applications. In: EurosoiUlis 2015 Italy. p. 317–20. <http://dx.doi.org/10.1109/UJIS.2015.7063837>.
- [6] Athanasiou S et al. Preliminary 3D TCAD electro-thermal Simulations of BIMOS transistor in thin silicon film for ESD protection in FDSOI UTBB CMOS technology. In: ICICDT 2015. p. 1–4. <http://dx.doi.org/10.1109/ICICDT.2015.7165913>.
- [7] Solaro Yohann et al. Innovative ESD protections for UTBB FD-SOI technology. In: Electron Devices Meeting (IEDM), Washington, US. p. 7.3.1–4. doi: <http://dx.doi.org/10.1109/IEDM.2013.6724580>.
- [8] Synopsis Sentaurus TCAD, ver G-2012.06.
- [9] Jayant Baliga B. *Fundamentals of power semiconductor devices*. Springer Publishing Company, Incorporated; 2008.
- [10] Galy Ph et al. Numerical evaluation between Transmission Line Pulse (TLP) and Average Current Slope (ACS) of a submicron gg-nMOS transistor under Electrostatic Discharge (ESD). In: Workshop EOS/ESD/EMI, LAAS-CNRS Toulouse. p. 15–7.
- [11] Voldman Steven H. *ESD physics and devices*. John Wiley & Sons Ltd.; 2004.
- [12] Maloney T, Khurana N. Transmission line pulsing techniques for circuit modeling of ESD phenomena. In: EOS/ESD symposium proceedings. pp. 49.
- [13] Athanasiou S et al. GDNMOS: a new high voltage device for ESD protection in 28 nm UTBB FD-SOI technology. In: EurosoiUlis 2016 Austria. p. 151–54. <http://dx.doi.org/10.1109/UJIS.2016.7440075>.
- [14] Eminente S et al. Ultra-thin fully-depleted SOI MOSFETs: special charge properties and coupling effects. *Solid State Electron* 2007;51(2):239–44. doi: <http://dx.doi.org/10.1016/j.sse.2007.01.016>.
- [15] Cristoloveanu S, Li Sheng S. *Electrical characterization of SOI materials and devices*. Springer; 1995. doi: <http://dx.doi.org/10.1007/978-1-4615-2245-4>.
- [16] Ogura et al. Design and characteristics of the Lightly Doped Drain-Source (LDD) insulated gate field-effect transistor. *IEEE Trans Electron Dev* 1980;27(8):1359–67. doi: <http://dx.doi.org/10.1109/T-ED.1980.20040>.



Sharp-switching band-modulation back-gated devices in advanced FDSOI technology



Hassan El Dirani^{a,b,*}, Pascal Fonteneau^a, Yohann Solaro^b, Charles-Alex Legrand^a, David Marin-Cudraz^a, Philippe Ferrari^b, Sorin Cristoloveanu^b

^a STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, France

^b Univ. Grenoble Alpes, IMEP-LAHC, CNRS, F-38000 Grenoble, France

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by Viktor Sverdlov

ABSTRACT

A band-modulation device with a free top surface, named Z³-FET (Zero front-gate, Zero swing slope and Zero impact ionization) and fabricated in the most advanced Fully Depleted Silicon-On-Insulator technology, is demonstrated experimentally. Since the device has no front gate, the operation mechanism is controlled by two adjacent heavily doped buried ground planes acting as back-gates. Characteristics such as sharp quasi-vertical switching, low leakage, and tunable trigger voltage are measured and discussed. We explore several variants (thin and thick silicon or SiGe body) and show promising results in terms of high current, switching performance and ESD capability with relatively low back-gate and drain bias operation.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Fully Depleted Silicon On Insulator technology (FDSOI) features high performance by reducing the parasitic capacitances, varying the threshold voltage with back-gate biasing, improving the leakage, mobility and subthreshold swing (SS) [1,2]. In that context, devices with Ultra-Thin Body and Buried oxide (UTBB) attract increasing attention for RF and IoT applications. In addition, FDSOI technology benefits from simpler (planar) manufacturing process than FinFETs, offering lower power consumption [3,4]. First introduced at the CMOS 28 nm node [1], the Z³-FET [5] (Zero gate, Zero swing slope and Zero impact ionization) is a band-modulation device, like FED [6–8] and Z²-FET [9–11], but without top gates. The device is operated with back-gate voltage applied on two independent ground planes (GP). It exhibits ‘vertical’ switching, low leakage current (I_{Leak}), tunable triggering voltage (V_{t1}), and high ON current (I_{ON}). For better electrostatic control, lower operating voltage, improved power consumption and further CMOS down-scaling, it is suitable to decrease the BOX and silicon film thicknesses while making the source/drain terminals thicker (S/D) [12]. The impact of these parameters on Z³-FET is investigated in this paper.

The device has been proposed recently [5] and further documented in [13]. In this paper, we investigate in more detail the characteristics of Z³-FET fabricated with the most advanced, post-28 nm-node FDSOI technology [12]. The architecture, technology description and the operation principle of Z³-FET are described in Section 2. Section 3 is dedicated to the regular Z³-FET with undoped body. We discuss the impact of silicon film thickness and temperature on the device performance, the capability to sustain high voltage and the response in the high current regime. In Section 4, we introduce a doped variant of Z³-FET, explored with DC and high current TLP measurements.

2. Fabrication, architecture and principle of operation

2.1. Fabrication process

The proposed Z³-FET is fully compatible with FDSOI CMOS process flow. The process started from an SOI substrate with 20 nm BOX and 6 nm active Si layer thicknesses [12]. Variants consist of thicker film (12 nm) and either Si or SiGe body. The thin buried oxide separates the undoped film ($N_A = 10^{15} \text{ cm}^{-3}$) from two highly doped P and N type ground planes (with $N_A = N_D = 10^{18} \text{ cm}^{-3}$, respectively) acting as back-gates. In situ Si epitaxy (needed for reduced series resistance [14]) is performed only in source/drain regions. The processing of high-k dielectric and metal gate is omitted, making the device cost effective.

* Corresponding author at: STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles Cedex, France.

E-mail address: hassan.eldirani1@st.com (H. El Dirani).

2.2. Device structure and operation mechanism

The Z^3 -FET is a forward biased P-I-N diode with an undoped ultra-thin silicon film ($t_{Si} = 6$ nm, Fig. 1a). The source (N^+ doped) is grounded and the drain (P^+ doped) is positively biased ($V_A > 0$ V). Sharp switching is controlled by two separated ground planes that act as back-gates (GP-N and GP-P in Fig. 1). The heavily doped GPs are respectively positively (GP-N) and negatively (GP-P) biased. The device is normally OFF thanks to potential barriers that block the injection of electrons (from N^+ source) and holes (from P^+ drain) into the body. The twin GPs induce ‘electrostatic’ doping (N-type above GP-N and P-type above GP-P) and emulate a virtual NPN thyristor structure. The placement and the biasing of the two ground planes is selected to form adequate injection barriers while maintaining the buried N^+/P^+ diode underneath the BOX in reverse mode for low leakage current. The triggering from OFF state to ON state is achieved by increasing the anode voltage. When V_A reaches V_{t1} , a positive feedback mechanism occurs due to the flow of carriers from the anode to the cathode and vice versa, leading to a sudden collapse of barriers, as documented in [9]. Basically, the carriers injected through one barrier lowers the height of the opposite barrier, via charge-induced change of body potential, initiating a positive feedback. This band-modulation mechanism results in remarkable sharp transition from low to high current, and enables an I_{ON}/I_{OFF} ratio of 8 decades (Figs. 2–4). Comparing to Z^2 -FET, the Z^3 -FET does not have high-k metal gate stack which eliminates any issues related to high voltage reliability. The free surface can be functionalized for various applications like bio, light and radiation sensing. Another important feature of Z^3 -FET that benefits from the BOX as a gate oxide is the ability of sustaining high back-gate bias. These functionalities can be combined with the typical applications of band-modulation devices (memory, ESD, fast logic), leading for example to sensors with built-in memory.

Three variants have been fabricated. The first one has an ultra-thin silicon film ($t_{Si} = 6$ nm), Fig. 1a. The second (Fig. 1b) and the third (Fig. 1c) variants feature thicker Si film ($t_{Si} = 12$ nm). The L_n part of the third architecture (Fig. 1c) received optional high N-type doping.

3. Z^3 -FET characterization and analysis

The fabricated devices were systematically characterized in static and pulsed (TLP) modes. In the following, the undoped version of Z^3 -FET is explored with variable silicon film thickness (6 nm and 12 nm) and different types of channel doping (Si in Fig. 2a and b and SiGe in Fig. 2c). $V_{G_{bP}}$ and $V_{G_{bN}}$ represent the back GP-P and GP-N biases, respectively.

3.1. DC characteristics and hysteresis

Figs. 2–4 show the typical output device characteristics: sharp switch, low leakage current, tunable triggering voltage V_{t1} and hysteresis. The need to keep the PIN diode in OFF state requires robust barriers that block the injection of carriers into the channel. In ultrathin devices, the L_p barrier is strong enough even at $V_{G_{bP}} = 0$ V featuring a very low leakage current ($I_{leak} < 10^{-11}$ A) but degraded sharp switching. A thicker film ($t_{Si} = 12$ nm) is needed to reinforce the electron barrier while negatively biasing the GP-P ($V_{G_{bP}} = 2$ V). The sharp switch is recovered for $V_{G_{bN}} \geq 1$ V, as shown in Fig. 2b.

The feedback between the barriers is affected by the silicon film thickness. In ultrathin films, the recombination rate of carriers increases, as dominated by the interfaces, and the effective carrier lifetime decreases. Numerical simulations with reduced carrier lifetime confirmed the degradation of the switch sharpness [5]. In order to trigger the feedback mechanism between the barriers and retrieve the sharp switch, two solutions have been tested. The first approach consists in negatively biasing the GP-P; however, $V_{G_{bP}} = -2$ V is not sufficient to retrieve the sharp switch and the characteristics are similar to those in Fig. 2a. It seems that the lifetime is too short and does not allow the carriers injected through one barrier to reach the second barrier and initiate the feedback mechanism. The second solution is to increase the film thickness from 6 nm to 12 nm; a steep switch over >6 decades of current is indeed observed in Fig. 2b. With SiGe body, the barriers are weaker and the carriers can be injected more easily than in devices with Si channel. This leads to a slight reduction of triggering voltage V_{t1} , as seen in Fig. 2c, that is attractive for low-power circuits.

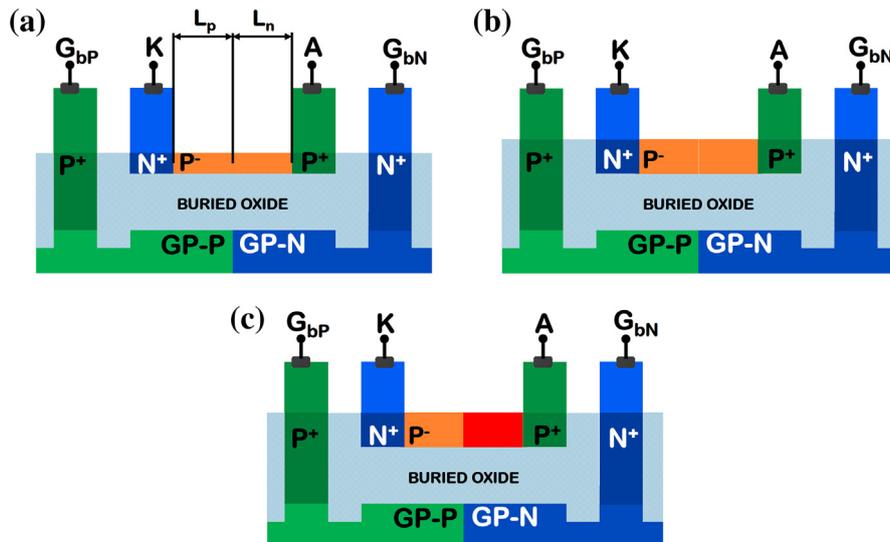


Fig. 1. Schematic of Z^3 -FET architecture in advanced FDSOI technology: (a) $t_{Si} = 6$ nm, (b) $t_{Si} = 12$ nm and (c) partially N-doped channel.

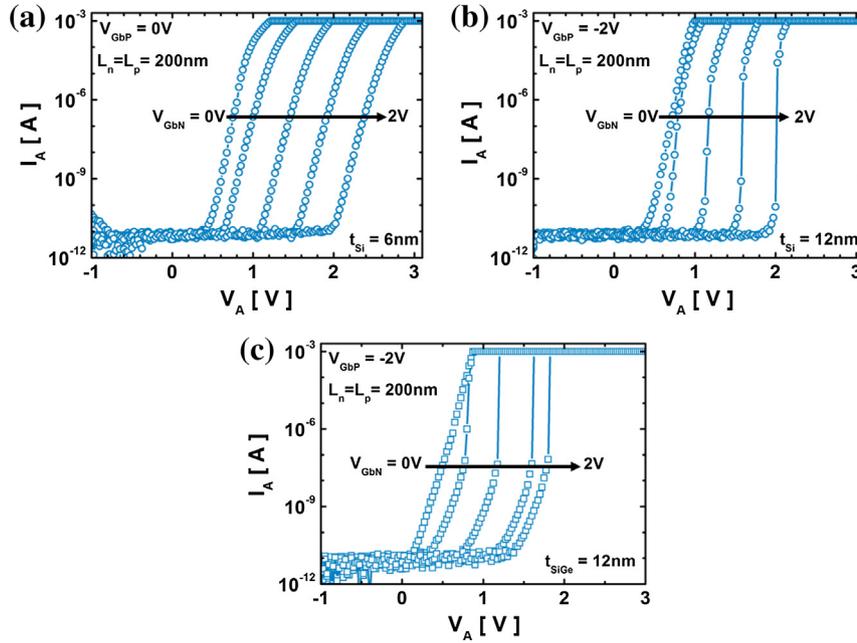


Fig. 2. Experimental DC I_A - V_A characteristics for different GP-N biases in Z^3 -FETs with (a) thin film $t_{Si} = 6$ nm and $V_{Gbp} = 0$ (similar curves for $V_{Gbp} = -2$ V); (b) thicker film $t_{Si} = 12$ nm and $V_{Gbp} = -2$ V; (c) thicker SiGe body $t_{SiGe} = 12$ nm and $V_{Gbp} = -2$ V. $L_n = L_p = 200$ nm. The saturation of ON current is an artefact due to the 1 mA current compliance used in the setup.

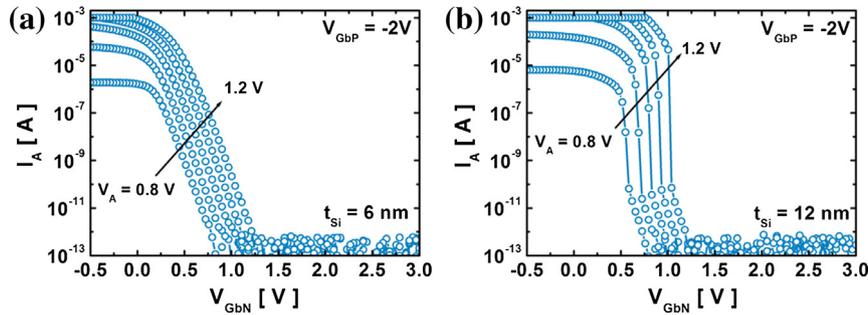


Fig. 3. I_A - V_{Gbn} transfer characteristics for various anode bias V_A in Z^3 -FET with (a) ultrathin body and (b) thicker body. $L_n = L_p = 200$ nm, 1 mA current compliance.

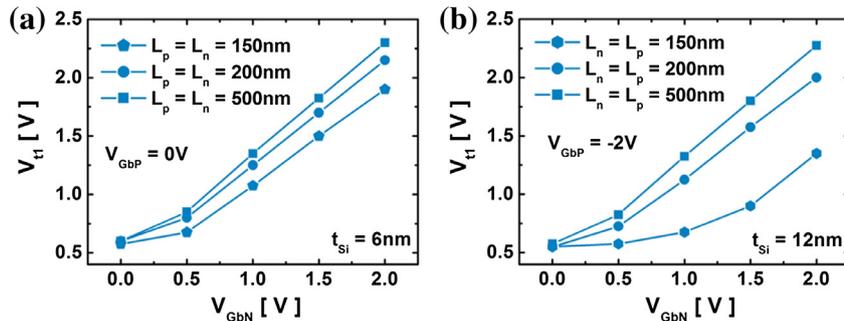


Fig. 4. Variation of triggering voltage V_{t1} with GP-N bias for (a) ultrathin body $t_{Si} = 6$ nm at $V_{Gbp} = 0$ V and (b) thicker body $t_{Si} = 12$ nm at $V_{Gbp} = -2$ V.

The transfer characteristics I_A - V_{Gbn} is shown in Fig. 3 which confirms the ability of the Z^3 -FET to switch from a low current in OFF state ($I_{OFF} < 10^{-13}$ A) to a high current in ON state ($I_{ON} > 10^{-3}$ A). While the ultrathin (6 nm) Z^3 -FET shows ~ 90 mV/decade sub-threshold swing (Fig. 3a), the thicker (12 nm) device exhibits a sharp switch (Fig. 3b). A higher anode voltage further improves the switching performance. Similar vertical slope characteristics are obtained by back-sweeping V_{Gbp} as shown in [5]. The Z^2 -FET, governed by its front-gate, features a sharper OFF to ON state tran-

sition (~ 1 mV/decade [10,11,15–17]) than the Z^3 -FET ($\Delta V_{Gbn}/\Delta I_A = 7$ mV/decade, Fig. 3b), which is operated by the back-gates through a thick oxide (BOX). The turn-on voltage is controlled by the potential difference between the anode and GP-N ($V_A - V_{Gbn}$). As V_A increases, a higher V_{Gbn} is needed to block the device, as shown in Fig. 3b.

As V_{Gbn} increases, the holes injection barrier is stronger, hence a higher V_{t1} is needed to turn on the device. The evolution of triggering voltage with V_{Gbn} is presented in Fig. 4a and b,

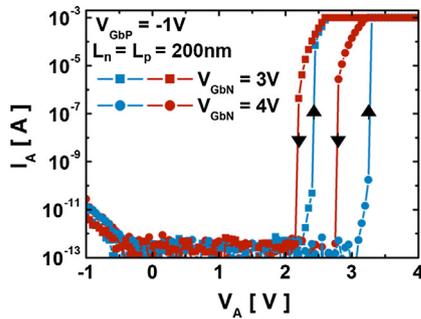


Fig. 5. Output I_A - V_A curves showing sharp switching and gate controlled hysteresis for various V_{GbN} . $V_{Gbp} = -1$ V, $L_n = L_p = 200$, $t_{Si} = 12$ nm, 1 mA current compliance.

showing that V_{t1} is very sensitive to the back-gate bias ($\Delta V_{t1}/\Delta V_{GbN} = 900$ mV/V).

In ultrathin devices with grounded GP-P, the impact of the channel length on the triggering voltage V_{t1} is low (Fig. 4a). On the other hand, devices with thicker t_{Si} need a negative $V_{Gbp} = -2$ V to achieve the sharp switch. The strong coupling between the GP bias and the channel makes the ON voltage V_{t1} adjustable up to 2 V. This coupling effect reduces as device length decreases (hexagon symbols, Fig. 4b) because the barriers are weaker.

In the following, we discuss the hysteresis. The output I_A - V_A curves in Fig. 5 show that the device is in OFF state at relatively low bias ($V_A < 2$ V) and turns ON sharply when V_A reaches the triggering voltage V_{t1} . As V_A is swept back, a small anti-clockwise hysteresis is observed until V_A is able to turn the device OFF. Since the robustness of L_n barrier is improved with V_{GbN} , the triggering voltage is increased and the hysteresis window is enlarged, as illustrated in Fig. 5.

It is important to note that the electrostatic control of the device barriers can decrease with increasing the reverse bias of the diode formed by the GPs (see Section 3.3). In this case, the barriers are relatively weaker and the OFF state is retrieved for lower sweep-back V_A . As a result, the hysteresis window is wider, which is ben-

eficial for promoting the Z^3 -FET as a competitor of 1T-DRAM memory.

3.2. Impact of operating temperature

Some applications require high device performance in a wide temperature range. For that purpose, the Z^3 -FET was characterized at room temperature and high temperature (125 °C). Fig. 6 shows that the Z^3 -FET is also functional at high temperature. The switch from OFF to ON state is slightly affected by the operating temperature, as seen in Fig. 6a. Fig. 6b shows that the reduction of V_{t1} is equal to 0.4 V ($\Delta V_{t1}/\Delta T = -3$ mV/°C). However, the quasi-linear dependence of V_{t1} on V_{GbN} is maintained from room temperature up to 125 °C. The V_{t1} variation also increases when reducing the device length (-5 mV/°C for $L_n = L_p = 200$ nm and -2 mV/°C for $L_n = L_p = 500$ nm), as presented in Fig. 6c. Finally, the I_{ON}/I_{OFF} ratio, 10^8 at room temperature, reduces to 10^7 at 125 °C, since the thermal energy of carriers increases at higher temperature [17,18] and they can easily cross the device from node to node. Considering these results, it can be concluded that the small temperature-dependence of Z^3 -FET is an asset for ESD protection schemes.

3.3. High voltage

The advantage of having the triggering voltage modulated by the GPs allows using the Z^3 -FET as a high voltage device. Increasing V_{GbN} to 7 V requires in principle a higher anode voltage V_A to turn ON the device. However, the actual mechanism is more subtle and depends on the channel length, as shown in Fig. 7. In long Z^3 -FET, $V_{t1}(V_{GbN})$ is a monotonic function (Fig. 7a), whereas in short device V_{t1} decreases for V_{GbN} higher than 5 V (Fig. 7b). This means that the hole injection barrier starts to shrink for $V_{GbN} > 5$ V, which seems counter-intuitive. We explain this surprising effect by the expansion of the depletion region of the reverse-biased diode formed by the two GPs below the BOX. The width W_{dep} of the junction depletion region is calculated for different doping concentrations in Fig. 7c. In the studied case ($N_A = N_D = 10^{18}$ cm $^{-3}$ for GP-P and

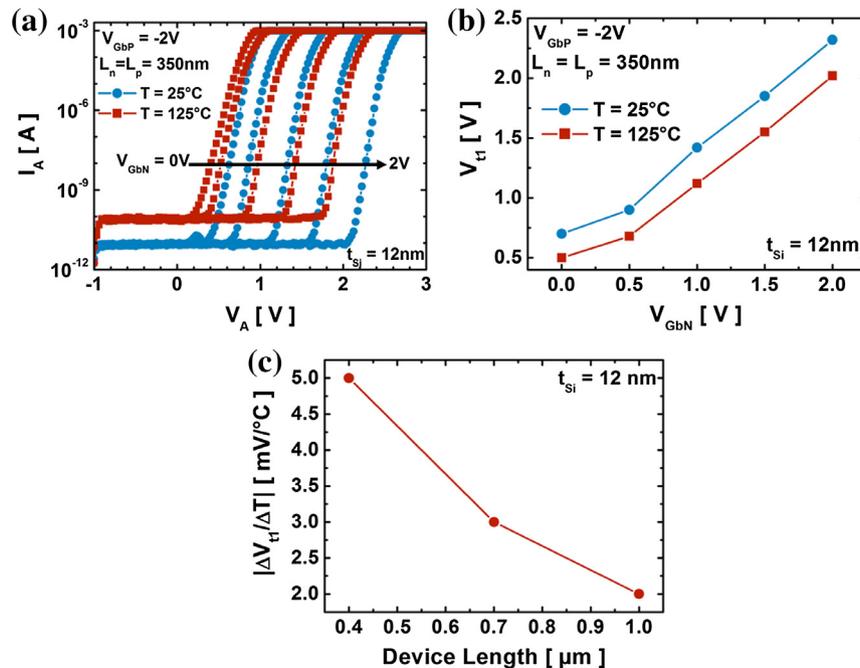


Fig. 6. (a) DC I_A - V_A curves for two temperatures: high temperature $T = 125$ °C (red square symbols) and room temperature (blue circle symbols) for various V_{GbN} and $V_{Gbp} = -2$ V. (b) V_{t1} versus GP-N bias at $T = 25$ °C and 125 °C. (c) Variation of $\Delta V_{t1}/\Delta T$ with device length. $t_{Si} = 12$ nm, 1 mA compliance. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

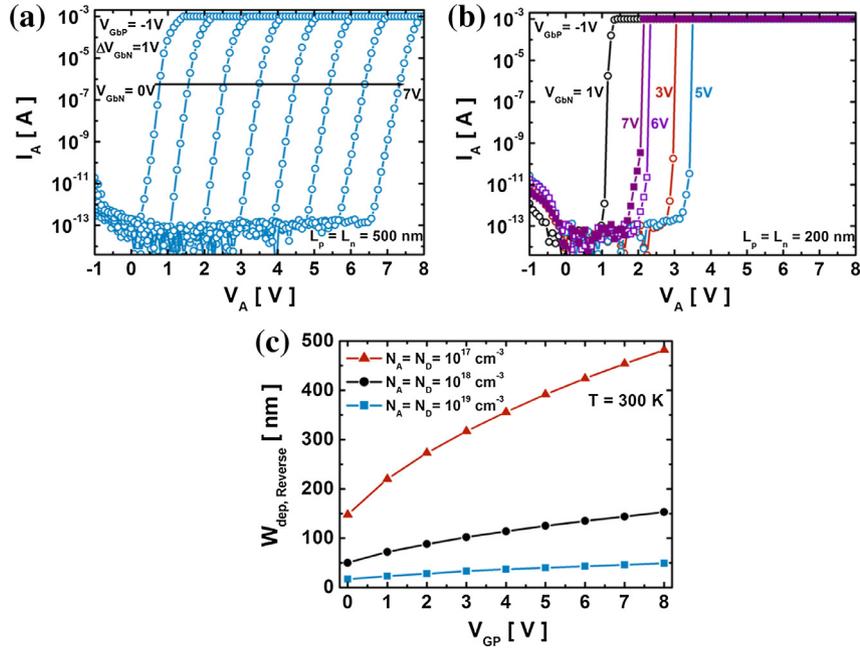


Fig. 7. Current versus anode voltage measured for GP-N bias reaching high voltage in (a) short Z^3 -FET, $L_n = L_p = 200$ nm and (b) long Z^3 -FET, $L_n = L_p = 500$ nm with $V_{G_{bP}} = -1$ V. (c) Evolution of lateral depletion width between the GPs versus reverse bias $V_{GP} = V_{G_{bN}} - V_{G_{bP}}$ for different doping concentrations. $t_{Si} = 12$ nm.

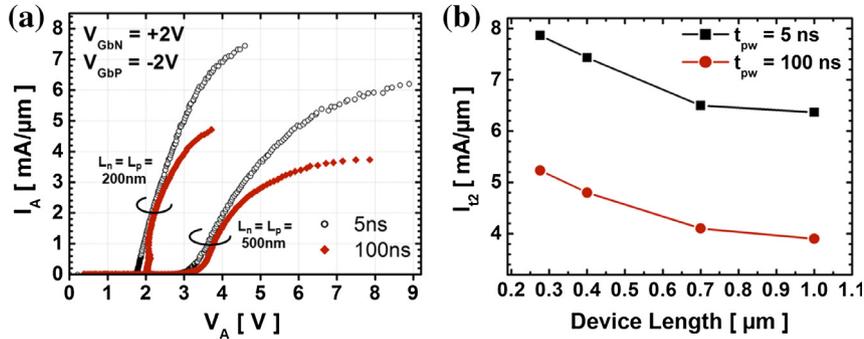


Fig. 8. TLP characterization results. (a) I_A - V_A curves in short and long Z^3 -FETs for 5 ns pulse width (open symbols) and 100 ns (closed symbols) with native rise time (~ 300 ps), $V_{G_{bN}} = 2$ V and $V_{G_{bP}} = -2$ V. (b) Variation of failure current I_{t2} with device length for 5 ns and 100 ns pulse width. $t_{Si} = 12$ nm.

GP-N respectively), the depletion zone reaches 150 nm at $V_{GP} = V_{G_{bN}} - V_{G_{bP}} = 8$ V. Thus the electrostatic control of the GPs on the channel decreases, implying narrower barriers and a V_{T1} reduction for $V_{GP} > 5$ V, as noticed in Fig. 7b. The channel control is retrieved by increasing the device length ($L_n = L_p \gg W_{dep}$). The long Z^3 -FET ($L_n = L_p = 500$ nm, Fig. 7a) can sustain high anode voltage up to 8 V without device breakdown. A second solution is to use heavier doping in GPs: for $N_A = N_D = 10^{19}$ cm^{-3} , the space charge region of the diode falls down below 40 nm even at high voltage (Fig. 7c). However, overdoping the ground planes (by ion implantation) may also affect the body doping. Note that the diode formed by the two GPs is always reversed biased and exhibits negligible leakage current.

3.4. TLP measurements

The ESD behavior was investigated with transmission line pulse (TLP) characterization [19,20]. We used different pulse widths ($t_{PW} = 5$ and 100 ns) and native rise time (~ 300 ps). Typical measurements are shown in Fig. 8. The high current regime is dominated by self-heating where the temperature rise degrades the mobility and eventually leads to thermal runaway causing device

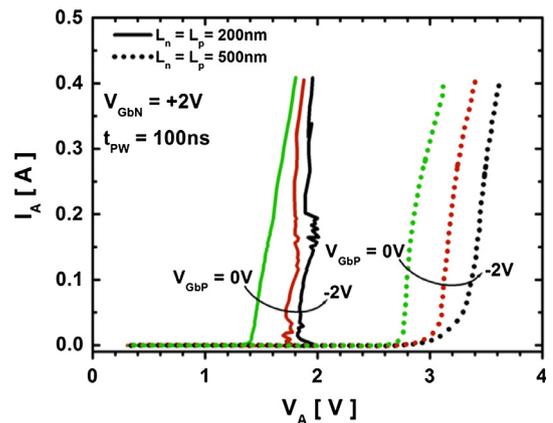


Fig. 9. Non-destructive TLP measurements for 100 ns pulse width in short (solid lines) and long (dotted lines) Z^3 -FETs for $V_{G_{bN}} = 2$ V and various GP-P bias. $t_{Si} = 12$ nm.

breakdown. As noticed in Fig. 8a, the heating is reduced for shorter pulse widths, where the failure current I_{t2} of short Z^3 -FET ($L_p =$

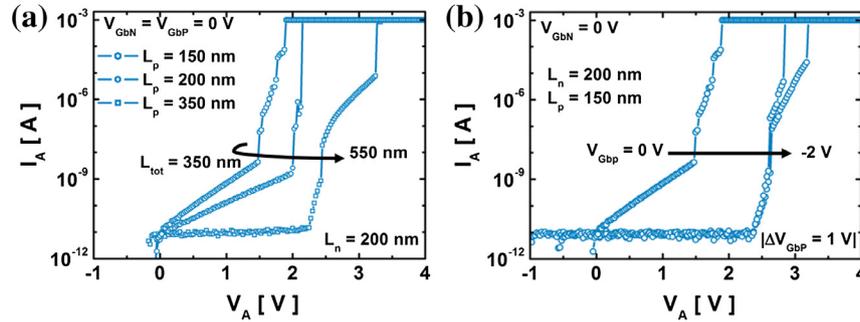


Fig. 10. Current versus drain voltage measured for (a) different geometries at $V_{GbN} = V_{GbP} = 0$ V and (b) various V_{GbP} bias ($L_n = 200$ nm, $L_p = 150$ nm at $V_{GbN} = 0$ V). Doped Z^3 -FETs with $t_{Si} = 12$ nm.

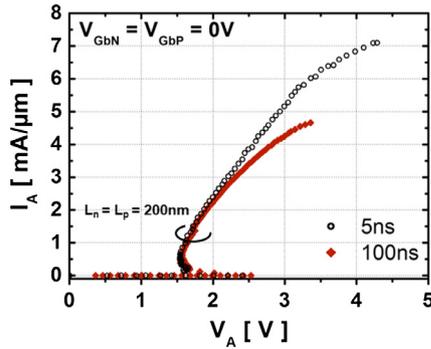


Fig. 11. TLP I_A - V_A characteristics for 5 ns (open symbols) and 100 ns (closed symbols) pulse width with native rise time (~ 300 ps). N-doped Z^3 -FET with $L_p = L_n = 200$ nm and $V_{GbN} = V_{GbP} = 0$ V. $t_{Si} = 12$ nm.

$L_n = 200$ nm) is improved from 4.9 mA/ μ m for $t_{pw} = 100$ ns up to 7.6 mA/ μ m for $t_{pw} = 5$ ns. The maximum current value and the triggering voltage V_{t1} depend on device length. Shorter Z^3 -FETs show improved performance: easier triggering with smaller V_{t1} and higher current capability (Fig. 8a). The evolution of failure current I_{t2} with device length is presented in Fig. 8b. It is confirmed that I_{t2} increases for shorter pulse widths and decreases with the device length. Thanks to the high performance, even at $t_{pw} = 100$ ns, Z^3 -FET stands as a viable candidate for protection against HBM-type (Human Body Model) ESD events.

Fig. 9 confirms that the triggering voltage of the device is tunable not only by GP-N but also by GP-P bias and device length. The higher the $|V_{GbP}|$ bias, the larger the triggering voltage. Taking advantage of these adjustable parameters, our devices are able to fulfill the ESD design window requirements showing ultra-low leakage current, high failure current and abrupt switching.

4. Doped Z^3 -FET

The undoped variant of Z^3 -FET features excellent performance but requires a negative bias on GP-P and a positive bias on GP-N, which is unsuitable in some applications. In order to address this issue, a new variant has been fabricated with a highly doped L_n part of the channel (Fig. 1c). The device behavior is discussed in the following section.

4.1. Static DC leakage characteristics

The I_A - V_A characteristics of the doped Z^3 -FET are presented in Fig. 10 for different geometries (Fig. 10a) and various V_{GbP} bias values (Fig. 10b). The N-doped L_n region of the channel forms a natural strong barrier against hole injection, avoiding the positive bias

of the GP-N. Hence, the device is blocked even without back-gate bias (Fig. 10a). By contrast, the L_p region is left undoped and shows a clear dependence on GP-P bias (Fig. 10b). In short devices, the GP-N barrier is narrow and cannot prevent completely the injection of holes toward the cathode. Since at $V_{GbP} = 0$ V the GP-P barrier is weak, the leakage current is too high. In order to avoid leakage, the device length can be increased (Fig. 10a). An alternative solution, effective in both short and long devices, is to negatively bias the GP-P (Fig. 10b). When the barriers are broader (long device) and higher ($V_{GbP} < 0$), the triggering voltage increases as shown in Fig. 10a.

4.2. TLP measurements

Fig. 11 shows the S-shaped negative-resistance characteristic for the partially N-doped device with no back-gate bias which is an important feature for ESD chip designers. Compared to undoped device with the same length ($L_p = L_n = 200$ nm, Fig. 8a), the doped Z^3 -FET exhibits a higher triggering voltage ($V_{t1} = 2.5$ V in Fig. 11) with similar failure current $I_{t2} = 7.2$ mA/ μ m. Nevertheless, the dynamic resistance R_{ON} is increased in doped devices due to the high recombination rate in the channel.

5. Conclusion

A gateless band-modulation device (Z^3 -FET) was proposed and demonstrated experimentally in the most advanced FDSOI technology. Several variants with different silicon film thicknesses and doping levels were studied. High performance in terms of triggering, leakage current and failure current have been reported. For proper operation, the film thickness is very important, thicker devices showing sharper characteristics. Experiments showed that the Z^3 -FET characteristics are weakly dependent on temperature. Finally, the high regime current, investigated with TLP measurements, confirmed that Z^3 -FETs can serve as robust ESD protection element, high-voltage device, 1T-DRAM memory and, thanks to the absence of the top gate, as ion-, photo-, and radiation sensors.

Acknowledgements

The layout, process and ESD/LU teams at STMicroelectronics are thanked for helpful support and fruitful discussions. We also thank the European project WayToGoFast for financial support.

References

- [1] Planes N, Weber O, Barral V, Haendler S, Noblet D, Croain D, et al. 28 nm FDSOI technology platform for high-speed low-voltage digital applications. In: Dig tech pap - symp VLSI technol, 33(4). p. 133–4.
- [2] Faynot O, Andrieu F, Weber O, Fenouillet-Béranger C, Perreau P, Mazurier J, et al. Planar fully depleted SOI technology: a powerful architecture for the

- 20 nm node and beyond. In: Tech dig – int electron devices meet (IEDM). p. 50–3.
- [3] Wu SY, Lin CY, Chiang MC, Liaw JJ, Cheng JY, Yang SH, et al. A 16 nm FinFET CMOS technology for mobile SoC and computing applications. In: Int electron devices meet. p. 9.1.1–4.
- [4] Arnaud F, Planes N, Weber O, Barral V, Haendler S, Flatresse P, Nyer F. Switching energy efficiency optimization for advanced CPU thanks to UTBB technology. In: Tech dig – int electron devices meet (IEDM). p. 48–51.
- [5] Solaro Y, Fonteneau P, Legrand CA, Fenouillet-beranger C, Ferrari P, Cristoloveanu S. A sharp-switching device with free surface and buried gates based on band modulation and feedback mechanisms. *Solid State Electron* 2016;116:8–11.
- [6] Raissi F. A brief analysis of the field effect diode and breakdown transistor. *IEEE Trans Electron Devices* 1996;43(2):362–5.
- [7] Yang Y, Salman AA, Ioannou DE, Beebe SG. Design and optimization of the SOI field effect diode (FED) for ESD protection. *Solid State Electron* 2008;52(10):1482–5.
- [8] Yang Y, Gangopadhyay A, Li Q, Ioannou DE. Scaling of the SOI Field Effect Diode (FED) for memory application. In: Int semicond device res symp, ISDRS'09. p. 9–10.
- [9] Wan J, Cristoloveanu S, Le Royer C, Zaslavsky A. A feedback silicon-on-insulator steep switching device with gate-controlled carrier injection. *Solid State Electron* 2012;76:109–11.
- [10] Solaro Y, Wan J, Fonteneau P, Fenouillet-Beranger C, Le Royer C, Zaslavsky A, Ferrari P, Cristoloveanu S. Z²-FET: a promising FDSOI device for ESD protection. *Solid State Electron* 2014;97:23–9.
- [11] El Dirani H, Solaro Y, Fonteneau P, Ferrari P, Cristoloveanu S. Sharp-switching Z²-FET device in 14 nm FDSOI technology. In: European solid-state device conf ESSDERC. p. 250–3.
- [12] Weber O, Josse E, Andrieu F, Cros A, Richard E, Perreau P, Baylac E, Degors N, Gallon C, Perrin E, Chhun S, Petitprez E, Delmedico S, Simon J, Druais G. 14 nm FDSOI technology for high speed and energy efficient applications. In: Dig tech pap – symp VLSI technol (VLSIT). p. 14–5.
- [13] El Dirani H, Solaro Y, Fonteneau P, Legrand CA, Marin-Cudraz D, Golanski D, Ferrari P, Cristoloveanu S. A sharp-switching gateless device (Z3-FET) in advanced FDSOI technology. In: EUROSOCI-ULIS conference. p. 131–4.
- [14] Fenouillet-Beranger C, Denorme S, Perreau P, Buj C, Faynot O, Andrieu F, et al. FDSOI devices with thin BOX and ground plane integration for 32 nm node and below. *Solid State Electron* 2009;53(7):730–4.
- [15] Wan J, Le Royer C, Zaslavsky A, Cristoloveanu S. A systematic study of the sharp-switching Z²-FET device: from mechanism to modeling and compact memory applications. *Solid State Electron* 2013;90:2–11.
- [16] Solaro Y, Wan J, Fonteneau P, Fenouillet-beranger C, Le Royer C, Zaslavsky A, Ferrari P, Cristoloveanu S. Z²-FET as a novel FDSOI ESD protection device. *Solid State Electron* 2014;97:23–9.
- [17] El Dirani H, Solaro Y, Fonteneau P, Ferrari P, Cristoloveanu S. Properties and mechanisms of Z²-FET at variable temperature. *Solid State Electron* 2015:8–13.
- [18] Rodriguez N, Andrieu F, Navarro C, Faynot O, Gamiz F, Cristoloveanu S. Properties of 22 nm node extremely-thin-SOI MOSFETs. In: Int SOI Conf. IEEE; 2011. p. 1–2.
- [19] Barth JE, Verhaege K, Henry LG, Richner J. TLP calibration, correlation, standards, and new techniques. *IEEE Trans Electron Packag Manuf* 2001;24(2):99–108.
- [20] Maloney T, Khurana N. Transmission line pulsing technique for circuit modeling of ESD phenomena. In: Proc 7th EOS/ESD symposium. p. 49–54.



Electrical characterization and modeling of 1T-1R RRAM arrays with amorphous and poly-crystalline HfO₂



Alessandro Grossi^{*}, Cristian Zambelli, Piero Olivo, Alberto Crespo-Yepes¹, Javier Martin-Martinez¹, Rosana Rodríguez¹, Monserrat Nafria¹, Eduardo Perez², Christian Wenger²

Università degli Studi di Ferrara, Dip. di Ingegneria, Via Saragat 1, Ferrara 44122, Italy

ARTICLE INFO

Article history:

Available online 18 October 2016

The review of this paper was arranged by Viktor Sverdlov

Keywords:

RRAM array
Forming
Read window
Energy saving

ABSTRACT

In this work, a comparison between 1T-1R RRAM arrays, manufactured either with amorphous or poly-crystalline Metal–Insulator–Metal cells, is reported in terms of performance, reliability, Set/Reset operations energy requirements, intra-cell and inter-cell variability during 10k endurance cycles and 100k read disturb cycles. The modeling of the 1T-1R RRAM array cells has been performed with two different approaches: (i) a physical model like the Quantum Point Contact (QPC) model was used to find the relationship between the reliability properties observed during the endurance and the read disturb tests with the conductive filament properties; (ii) a compact model to be exploited in circuit simulations tools which models the I – V characteristics of each memory cells technology.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Resistive Random Access Memories (RRAM) technology gathered significant interest for several applications [1–3]. RRAM behavior is based on the possibility of electrically modifying the conductance of a Metal–Insulator–Metal (MIM) stack: the Set operation moves the cell in a low resistive state (LRS), whereas Reset brings the cell in a high resistive state (HRS) [4,5]. To activate such a switching behavior, some technologies require a preliminary Forming operation [6–8].

The choice of a proper Metal–Insulator–Metal (MIM) technology for RRAM cells, exhibiting good uniformity and low switching voltages, is still a key issue for array structures fabrication and reliable electrical operation [9]. Such a process step is mandatory to bring this technology to a maturity level. In this work, a comparison between 1T-1R RRAM 4kbits arrays manufactured either with amorphous [5] or poly-crystalline [10] HfO₂ is performed. In amorphous HfO₂ the conduction mainly occurs through a conductive filament with a variable concentration of defects, whereas in poly-crystalline HfO₂ the conduction occurs only through grain boundaries with a very low defect concentration. The differences in terms of conduction properties and defect concentrations translate into

different switching properties [9], with several implications on inter-cell variability (variations between cells) and intra-cell variability (cycle-to-cycle variations of any given cell).

In this work, that is an extended yet complete picture of the results presented in [11], a comparison in terms of performance, reliability, Set/Reset operations energy requirements, intra-cell and inter-cell variability during 10k endurance cycles is reported. In addition to the previously presented results, 100k read disturb cycles were performed to deepen the understanding of the reliability of each technology. Moreover, to understand the relationship between the reliability properties observed during the endurance and read disturb tests and the conductive filament properties, Quantum Point Contact (QPC) modeling [12] was used, since it allows to correctly represent the measured I – V characteristics independently from the conduction mechanism. Even if the QPC allows to model the conductive filaments properties taking into account the cell-to-cell variability, it offers a technology description that sometimes is complex to be implemented in circuit simulation tools. To this extent, an equivalent circuit model able to offer a simpler description of the devices was applied and validated on both MIM technologies. The memory cells used in this work can be modeled using a diode-resistance equivalent circuit model. The model parameters extracted from the fittings of experimental I – V curves can provide additional information about electrical properties of the memory cells to be exploited in the design of RRAM arrays.

^{*} Corresponding author.

E-mail address: alessandro.grossi@unife.it (A. Grossi).

¹ Universitat Autònoma de Barcelona (UAB), Departament d'Enginyeria Electrònica, 08193 Bellaterra, Spain.

² IHP, Im Technologiepark 25, 15236 Frankfurt (Oder), Germany.

2. Experimental setup

The 1T-1R memory cells in the 4kbits arrays are constituted by a select NMOS transistor manufactured with a 0.25 μm BiCMOS technology whose drain is in series to a MIM stack. The wordline (WL) voltage applied to the gate of the NMOS transistor allows setting the cell current compliance. The cross-sectional Scanning Transmission Electron Microscopy (STEM) image of the cell and the 1T-1R cell schematic are reported in Fig. 1. The variable MIM resistor is composed by 150 nm TiN top and bottom electrode layers deposited by magnetron sputtering, a 7 nm Ti layer, and a 8 nm HfO_2 layer deposited with two different Atomic Vapour Deposition (AVD) processes resulting either in amorphous (A) or polycrystalline (P) HfO_2 films, respectively. The resistor area is equal to 0.4 μm^2 . Amorphous films have been integrated also with a resistor area equal to 1 μm^2 . This latter process option shows improved reliability and performance [4]. The Forming/Set/Reset operations on the arrays were performed by using an Incremental Pulse and Verify algorithm. The bitline (BL), sourceline (SL) and WL voltages applied during Forming, Set, Reset and Read operations are reported in Table 1. Reset operations were performed by applying the highest WL voltage available (2.8 V on array A and 2.5 V on array P) to maximize the cells switching yield while avoiding the breakdown of the MIM [13]. Pulses were applied during Forming by increasing V_{BL} with $\Delta V_{BL} = 0.01$ V, whereas during Set and Reset $\Delta V_{BL} = 0.1$ V and $\Delta V_{SL} = 0.1$ V have been used, respectively. Each pulse featured a duration of 10 μs , with a rise/fall time of 1 μs to avoid overshoot issues. Set operation was stopped on a cell when the read-verify current reached 20 μA , whereas Reset was stopped when 10 μA was reached. Forming, Set and Reset BL/SL voltages necessary to reach the requested read-verify current targets are extracted from the characterization data and labelled as V_{FORM} , V_{SET} and V_{RES} , respectively.

3. Experimental results

Arrays using A- HfO_2 (A-array) with resistor area of 0.4 μm^2 , 1 μm^2 and P- HfO_2 (P-array) resulted in a Forming Yield (calculated

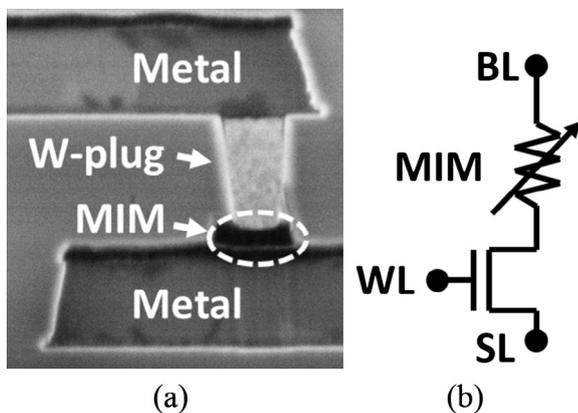


Fig. 1. Cross-sectional STEM image (a) and schematic (b) of the 1T-1R cell integrated in the arrays.

Table 1
Forming, set, reset and read voltage parameters.

Operation	V_{SL} [V]	V_{BL} [V]	V_{WL} [V]
Forming	0	2–3.2	1.5
Set	0	0.2–3.2	1.5
Reset	0.2–3.2	0	2.5 (A)/ 2.8 (P)
Read	0	0.2	1.5

as the cell percentage showing a read verify current after forming $I_{read} \geq 20$ μA) of 58%, 90% and 95%, respectively. Fig. 2 shows the average current ratios between Low Resistive State (LRS) and High Resistive State (HRS) read currents (I_{LRS}/I_{HRS}), calculated on the entire cells population during SET/RESET cycling at $V_{read} = 0.2$ V on A-array and P-array, and their relative dispersion coefficient. The minimum current ratio that allows to correctly discriminate between HRS and LRS, defined as $I_{LRS}/I_{HRS} > 2$, is indicated for comparison [5]. The average ratios of A-arrays with resistor area of 0.4 μm^2 and 1 μm^2 go under the minimum ratio limit after 200 and 1k cycles, respectively. To evaluate the cell-to-cell variability the dispersion coefficient of I_{LRS} and I_{HRS} distributions, defined as (σ^2/μ) , has been used. P-array showed higher Ratio (≈ 2.8) even after 10k cycles, but also a higher dispersion coefficient after Forming (i.e., cycle 1). The grain boundaries conduction mechanism in the poly-crystalline HfO_2 structure could be the reason of the higher cell-to-cell variability in P-arrays [14]. A-array with resistor area of 1 μm^2 shows a slightly higher average ratio than A-array with resistor area of 0.4 μm^2 .

Fig. 3 shows a comparison between I_{LRS} and I_{HRS} cumulative distributions measured at cycle 1 and after the endurance test: A-arrays show more compact distributions at cycle 1, however after the endurance test P-array shows a higher percentage of correctly switching cells reaching the Set/Reset verify targets. I_{HRS} cumulative distribution in P-array show a longer tail at cycle 1 compared to A-arrays. After 10k cycles only an increase of the tail in P-array can be observed whereas on A-arrays a strong shift of the distributions towards higher currents occurs, resulting in a higher number of cells not reaching the Reset threshold. I_{HRS} cumulative distribution in A-array with resistor area of 1 μm^2 shows lower currents at cycle 1 than A-arrays with resistor area of 0.4 μm^2 , however after 10k cycles I_{HRS} cumulative distributions are very similar. In I_{LRS} cumulative distributions a tail creation of cells not able to reach the Set threshold can be observed on P-arrays after 10k cycles, whereas on A-arrays a strong shift of the distributions towards lower currents occurs, resulting in a higher number of cells not reaching the Set threshold especially when cells with resistor area of 0.4 μm^2 are considered. A-array with resistor area of 0.4 μm^2 shows a high number of cells not reaching the Set threshold even at cycle 1.

Fig. 4 shows the average Set and Reset switching voltages (V_{SET} , V_{RES}) and their relative dispersion coefficients: lower V_{SET} and V_{RES} are required on P-array which shows no variations during the endurance test, whereas V_{SET} , V_{RES} increase on A-arrays during cycling. V_{RES} on P-array shows the highest variability. A-arrays show similar behavior of the average V_{SET} and V_{RES} (a lower average V_{SET} is observed on A-array with larger resistor area only up to 500

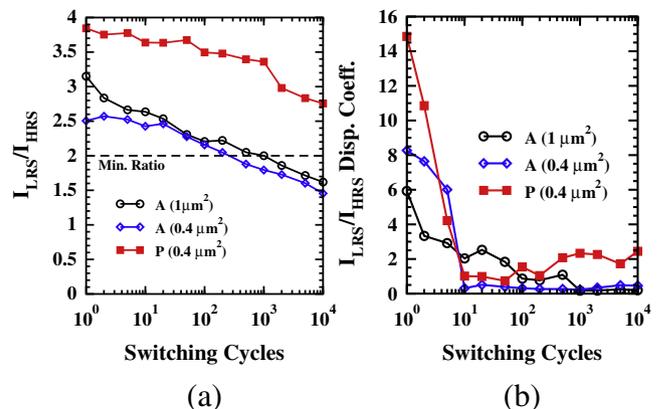


Fig. 2. I_{LRS}/I_{HRS} current ratio average values (a) and dispersion coefficients (b) calculated during cycling.

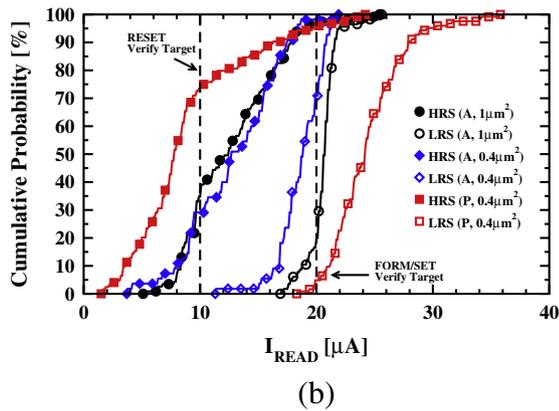
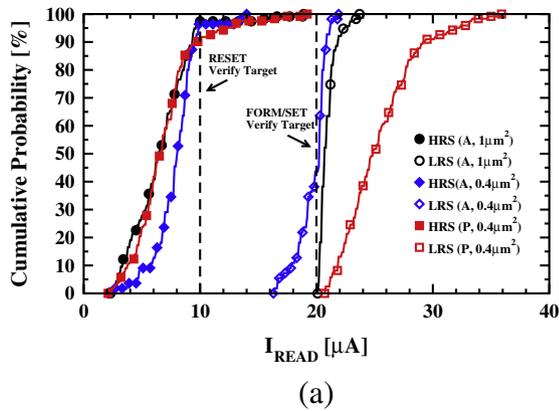


Fig. 3. I_{HRS} and I_{LRS} cumulative distributions at cycle 1 (a) and at cycle 10k (b).

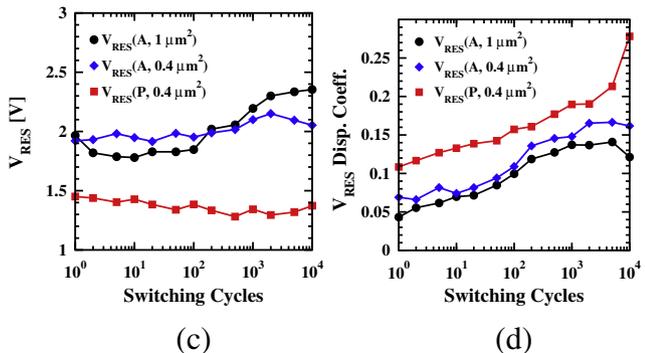
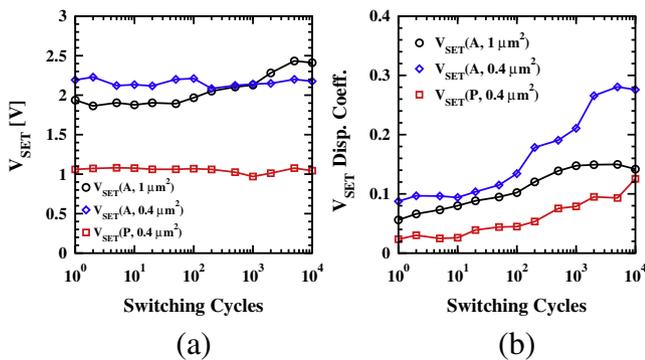


Fig. 4. V_{SET} and V_{RES} average values (a,b) and dispersion coefficients (c,d) calculated during cycling.

cycles), while a higher V_{SET} and V_{RES} dispersion can be observed in A-array with smaller resistor area.

Fig. 5 shows the cumulative distributions of Forming, Set and Reset switching voltages at cycle 1 and after the endurance test: Forming, Set and Reset algorithms starting point and last attempt are indicated, corresponding to the first and the last voltage pulse available in the incremental pulse and verify procedure. P-array requires lower V_{SET} and V_{RES} but higher V_{FORM} if compared to A-array with the same resistor area. A-array with larger resistor area requires higher V_{FORM} , moreover it can be observed that $\approx 40\%$ of the devices with smaller resistor area reached the forming threshold at $V_{FORM} = 2$ V, corresponding to the first attempt of the Forming Algorithm. Since P-array shows a more compact distribution on V_{SET} and a larger V_{RES} than A-arrays, faster Set operation could be reliably used on P-array, whereas on Reset an incremental pulse

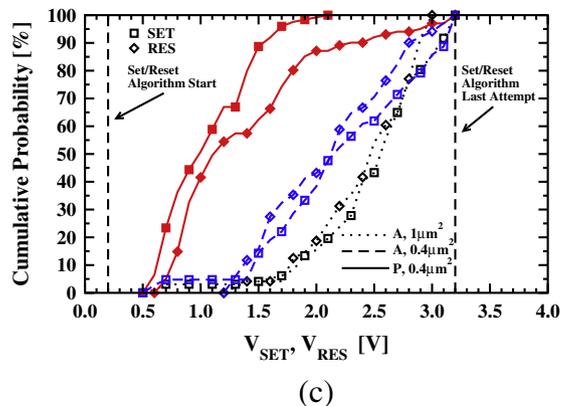
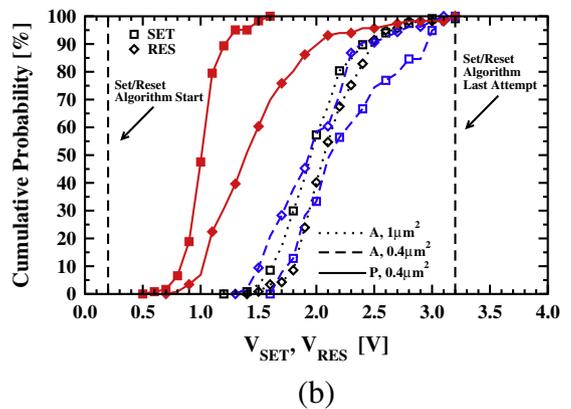
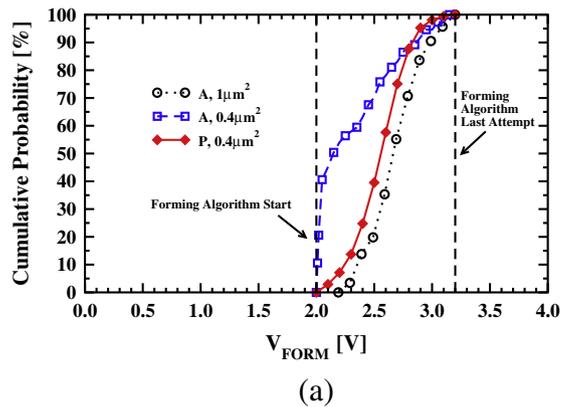


Fig. 5. V_{FORM} , V_{SET} and V_{RES} cumulative distributions at cycle 1 (a) and at cycle 10k (b).

with verify technique is required to ensure good reliability. A-arrays show large distributions on both V_{SET} and V_{RES} , hence the adaptation of incremental pulse with verify techniques is mandatory on such arrays.

Fig. 6 shows the average energy required to perform Set and Reset operations on a single cell: P-array shows lower power consumption with a lower increase during cycling. A-arrays with different resistor area show similar power consumption during Reset operation, whereas a lower consumption during Set is observed on A-array with larger resistor area only up to 500 cycles. The overall energy required to create/disrupt the conductive filament during Set/Reset operations has been calculated as:

$$E = \sum_{i=1}^n V_{pulse,i} * I_{pulse,i} * T_{pulse} + V_{read} * I_{read,i} * T_{read} \quad (1)$$

where n is the number of Reset pulses applied during incremental pulse operation, $V_{pulse,i}$ is the pulse voltage applied at step i , $I_{pulse,i}$ is the current flowing through RRAM cell during pulse i application, $T_{pulse} = 10 \mu s$ is the pulse length, $V_{read} = 0.2 V$ is the read voltage applied during verify operation, $I_{read,i}$ is the current read during read verify step i , and $T_{read} = 10 \mu s$ is the verify pulse length.

In the considered RRAM cells the read signals has the same polarization of the Set operation (both pulses are applied on the BL), hence the read disturb could only be a problem on cells in HRS state since a very long sequence of read pulses could slowly re-create the conductive filament, resulting into an undesired switch from HRS to LRS [13]. Read disturb has been evaluated only on cells in HRS state for each considered technology: Fig. 7 shows the average HRS read current and its relative standard deviation

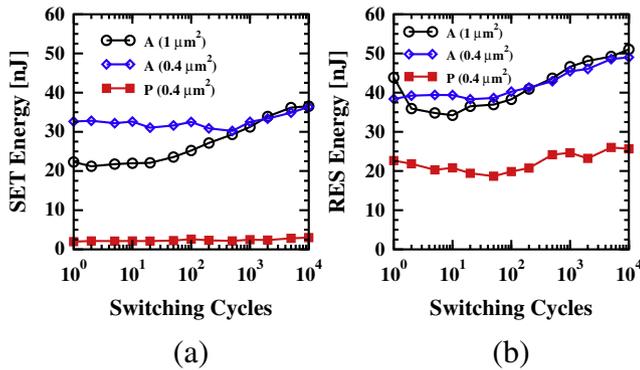


Fig. 6. Energy required to perform Set (a) and Reset (b) operations as a function of the Set/Reset cycle number.

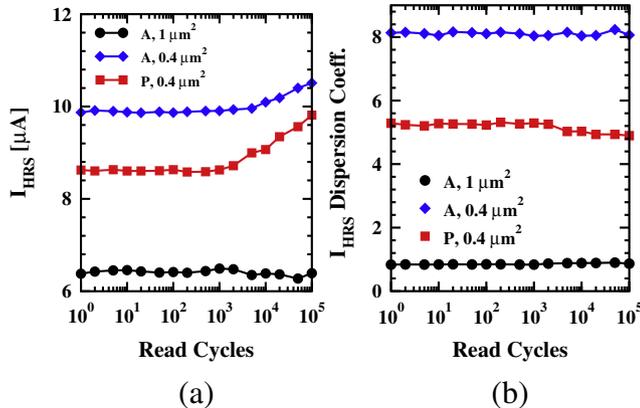


Fig. 7. Average read current variation (a) and dispersion coefficient evolution (b) of HRS calculated during 100k read disturb pulses, with $V_{pulse} = 0.2 V$.

measured during 100k read operations. P-array shows the highest read current variation, confirming that on such technology due to the high leakage currents it is easier to create conductive paths.

4. 1T-1R cells modeling

Extracting and modeling suitable parameters for the $I-V$ characteristics is important to gather statistical information for any kind of non-volatile memory [15]. In the RRAM arrays of this work, $I-V$ characteristics have been measured after-forming and modeled with two different approaches: in order to understand the differences on the conductive filament properties and variability QPC modeling has been used as in [16], while an equivalent circuit model [17] was used to obtain a description implementable in circuit simulation tools.

4.1. QPC modeling

Reset $I-V$ characteristics measured after-forming were used to analyze the conductive filament properties through QPC model. HRS current is calculated according to the expression:

$$I = \frac{2e}{h} G/G_0 \left(eV + \frac{1}{\alpha} \text{Ln} \left[\frac{1 + e^{\alpha(\Phi - \beta eV)}}{1 + e^{\alpha(\Phi + (1-\beta)eV)}} \right] \right) \quad (2)$$

where Φ is the barrier height (bottom of the first quantized level), $\alpha = t_B \pi^2 \hbar^{-1} \sqrt{2m^*}/\Phi$ is a parameter related to the inverse of the potential barrier curvature (assuming a parabolic longitudinal potential), $m^* = 0.44m_0$ is the effective electron mass and t_B is the barrier thickness at the equilibrium Fermi energy. β takes into account how the potential drops at the two ends of the filament: $\beta = 1$ has been used since the constriction is highly asymmetric [16]. G/G_0 is a conductance parameter equivalent to the number of filaments at very low voltages: in a very approximate way, a single highly conductive filament can be viewed as a parallel combination of elementary nanowires [18].

$I-V$ Reset operation has different impacts from cell-to-cell, resulting either into a break or a modulation of the conductive filament (CF) [8,16]. In the former case the presence of a potential barrier is assumed, hence fitting is performed considering $G/G_0 = 1$ and the average barrier length d and radius of the constriction r are calculated according to [12]. In the latter case, assuming the absence of a potential barrier, the normalized conductance of the filament G/G_0 is calculated. The percentage of cells resulting either into a CF break or modulation are reported in Table 2: the high leakage current in P-array makes very difficult to completely interrupt the conductive path hence the lowest percentage of CF break is obtained, whereas the highest percentage is obtained on A-array with the larger resistor area. The cumulative distributions of α and Φ fitting parameters calculated on the CF break cells are reported in Fig. 8.

Average value and standard deviation of the fitting parameters are reported in Table 3. The cumulative distributions of calculated barrier length d and radius r of the CF constriction are reported in Fig. 9, while the average value and standard deviations are reported in Table 4. A-array with the small resistor area shows the largest radius with the lowest barrier length: the presence of

Table 2
Reset condition comparison.

Technology	C.F. Break [%]	C.F. Modulation [%]
A, 1 μm^2	45	55
A, 0.4 μm^2	34	66
P 0.4 μm^2	20	80

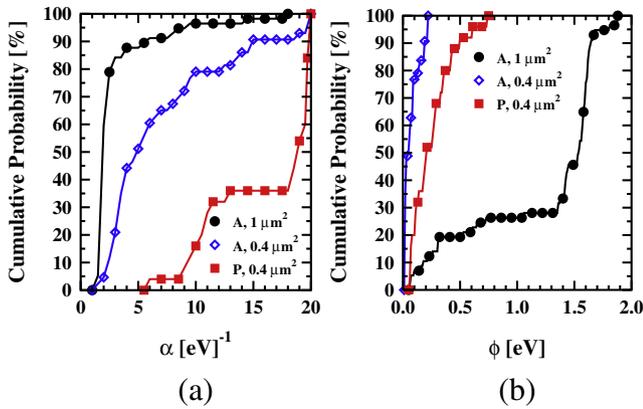


Fig. 8. Cumulative distribution of α and ϕ fitting parameters used on CF break cells.

Table 3

Average value and standard deviation of fitting parameters calculated on CF break cells.

Technology	α [eV ⁻¹]		ϕ [eV]	
	Avg.	Std.	Avg.	Std.
A, 1 μm^2	2.67	3.02	1.21	0.58
A, 0.4 μm^2	6.83	5.48	0.07	0.06
P, 0.4 μm^2	16.08	5.06	0.17	0.25

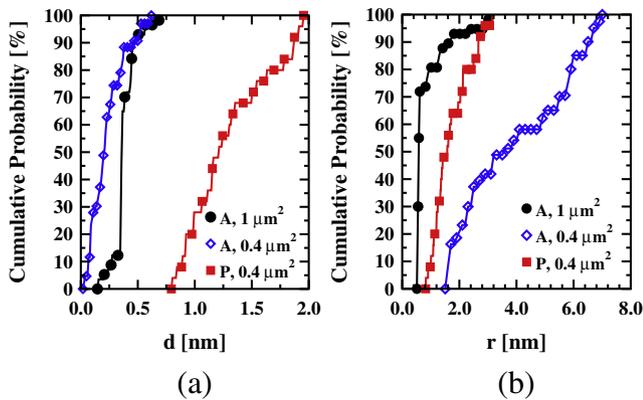


Fig. 9. Cumulative distribution of calculated barrier length d (a) and radius of the filament constriction r (b) on CF break cells.

Table 4

Average value and standard deviation of barrier length and filament radius calculated on CF break cells.

Technology	d [nm]		r [nm]	
	Avg.	Std.	Avg.	Std.
A, 1 μm^2	0.37	0.11	0.85	0.62
A, 0.4 μm^2	0.25	0.24	4.25	2.25
P, 0.4 μm^2	1.29	0.36	1.71	0.65

a very large constriction with a very low barrier explains the issues in controlling the cells' uniformity during Set and Reset operations. A-array with the larger resistor area shows higher barrier and smaller radius, resulting into a higher controllability during Set and Reset. Moreover, the highest parameters uniformity is observed, which translates into the highest HRS and LRS currents uniformity. A possible reason to explain the difference in the potential barrier between the two amorphous films is related to the defects concentrations in the HfO_2 . Indeed, amorphous films

integrated with lower area are affected by a higher defect concentration that eases the Reset process and therefore results in a lower potential barrier in the HRS. P-array shows the largest barrier with the highest variability: the highest barrier is the reason of the higher average ratio between HRS and LRS, while the high variability generates the high current variability observed in HRS.

In case of CF modulation fitting has been performed assuming large negative Φ values, α fixed to 1 (even if α and Φ play no role in such condition) and $G/G_0 \geq 1$ due to the presence of the residual filament. Fig. 10 shows the cumulative distribution of G/G_0 conductance values fitting parameters used on hard to disrupt cells: it can be observed that A-array with the larger resistor area shows the lowest variability, which is the reason of the lowest HRS current variability observed during Reset with the Incremental Pulse and Verify algorithm. Average value and standard deviation of the fitting parameter G/G_0 are reported in Table 5.

4.2. Equivalent circuit modeling

Electrical models are a power tool to analyze memory cells and circuits based on Resistive Switching (RS) devices allowing evaluating characteristics like power consumption or performance in large RS devices arrays [17,19]. To model the experimental I - V curves during both RS states (i.e., LRS and HRS) we use a Diode-Resistor based circuit (Fig. 11) where the resistance (R), the diode saturation current (I_s) and diode ideality factor (n) are the parameters of the model [17]. V_{APP} represents V_{BL} or V_{SL} , which are the applied voltages to produce the Set and Reset processes respectively.

To fit all the experimental I - V curves an automatized process has been developed to extract the model parameters values (Tables 6 and 7) for each curve. Fig. 12 shows some examples of experimental LRS I - V curves (circles) before the Reset process and the simulated curves using the circuit model of Fig. 11 with suitable parameters (continuous lines). As can be observed, the model fits perfectly with the experimental results for both amorphous and poly-crystalline samples. For each kind of samples, the analyzed

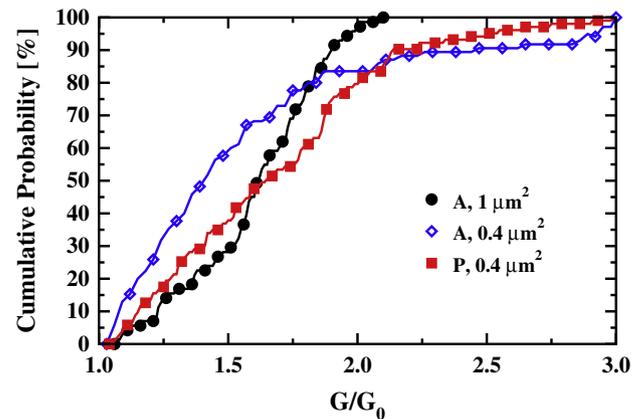


Fig. 10. Cumulative distribution of G/G_0 fitting parameters used on hard to disrupt cells.

Table 5

G/G_0 average value and standard deviation.

Technology	G/G_0	
	Avg.	Std.
A, 1 μm^2	1.61	0.26
A, 0.4 μm^2	1.56	0.53
P, 0.4 μm^2	1.67	0.42

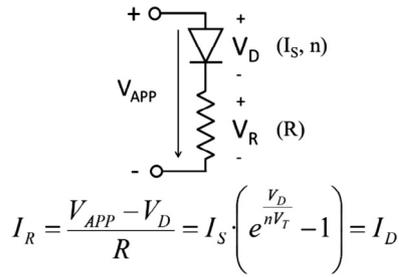


Fig. 11. Equivalent model for the 1T-1R device based on a Diode-Resistor circuit. Resistance (R), saturation current (I_S) and ideality factor (n) of the diode are the parameters used to fit the conduction of the 1T-1R devices at both resistive states, LRS and HRS, and for both types of samples, amorphous and poly-crystalline.

Table 6

Model parameters of the amorphous samples I - V fittings for both states, LRS and HRS.

	Set		Reset	
	Avg.	Std.	Avg.	Std.
R [K Ω]	25.84	29.6	6.25	4.18
I_S [A]	9.75e-5	4.56e-5	6.39e-8	1.37e-7
n	8.95	7.06	5.76	5.19

Table 7

Model parameters of the poly-crystalline samples I - V fittings for both states, LRS and HRS.

	Set		Reset	
	Avg.	Std.	Avg.	Std.
R [K Ω]	1.34	1.61	8.88	12.9
I_S [A]	5.67e-5	8.4e-5	1.52e-5	4.29e-7
n	3.95	5.03	1.18	2.35

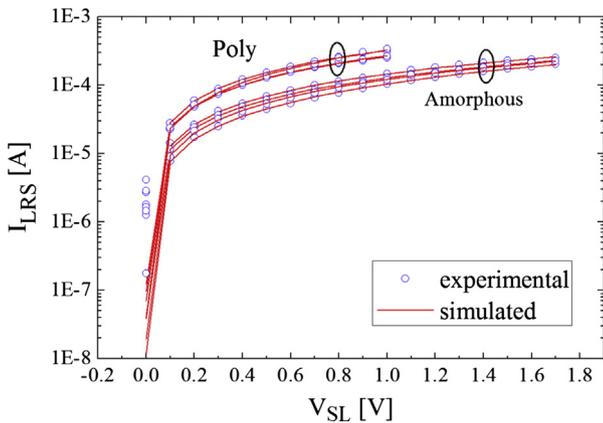


Fig. 12. Experimental LRS I - V curves (circles) and the simulated curves (continuous lines) obtained using the Diode-Resistor model. With a suitable parameter Set, the model reproduces properly the experimental curves for both poly-crystalline and amorphous samples.

voltage range was limited by the Reset voltage that is lower for the poly-crystalline samples.

The same automatic process was also used to fit HRS I - V curves for both samples types. Fig. 13 shows experimental HRS curves (circles) before the Set process and the corresponding simulated curves (continuous lines). Amorphous samples show very noisy I_{HRS} currents at low voltages (<1 V) that could be caused by the nature of the memory cell and the array structure where the drive transistor effect on the electrical characteristics of the memory

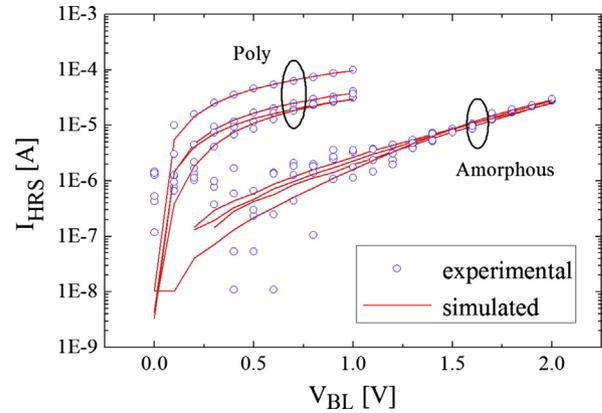


Fig. 13. Experimental HRS I - V curves (circles) and simulated HRS curves (continuous lines) using the Diode-Resistor model for both amorphous and poly-crystalline samples. Noisy currents at low voltages cannot be fitted by the model, especially for the amorphous samples where current values for V_{BL} below 1 V are not considered.

must be analyzed in detail. This noisy current must be neglected to avoid errors during the fitting process. For this reason, I_{HRS} values for V_{BL} below 1 V are not considered to force better fittings for voltages larger than 1 V, where the I - V curves are not affected by the noise. This consideration affects the obtained model parameters values (Tables 6 and 7). Thus, the best fittings for the amorphous HRS I - V curves require a mean R parameter value (6.25 K Ω) which is lower than the one obtained for the LRS (25.84 K Ω). This low value of the R parameter at HRS combined to the very low value of I_S at HRS (6.39e-8A) provides the best fitting between the experimental and the modeled curves.

5. Conclusions

1T-1R RRAM arrays manufactured with P-HfO₂ shows several advantages compared to A-HfO₂ even considering their improved process: higher current Ratio, lower switching voltages, lower power consumption, minor endurance degradation and higher overall yield. Moreover, P-array show very low V_{SET} variability, hence faster Set operation could be reliably performed. P-array disadvantages are represented by the larger HRS distribution after Forming, the higher Reset voltage dispersion, the lower read disturb immunity and the higher V_{FORM} if compared to A-array with the same resistor area, however it must be pointed out that such operation is performed only once. The grain boundaries conduction mechanism in the poly-crystalline HfO₂ structure could be the reason of the higher cell-to-cell variability observed in P-arrays. QPC modeling allowed showing that the higher uniformity observed on A-array with the large resistor area can be ascribed to a lower conductive filament shape variability in terms of radius of the constriction and barrier height, whereas the P-array shows the highest variability in terms of conductive filament shape: the reason could be ascribed again to the different conduction mechanism and the higher leakage currents observed on such technology. A diode-resistor equivalent circuit model correctly fits the experimental RS I - V characteristics of poly-crystalline and amorphous samples for both LRS and HRS. However, noisy current levels at low voltages, especially for amorphous samples, could lead to a non-well fitted curve. Thus, it is needed to remove them for a suitable current fitting at larger voltages.

Acknowledgments

This work was supported by the European Union's H2020 research and innovation programme under grant agreement No.

640073 and by ENIAC Joint Undertaking 2013-2, PANACHE No. 621217. UAB authors acknowledge funding from the Spanish MINECO and ERDF (TEC2013-45638-C3-1-R) and the Generalitat de Catalunya (2014SGR-384).

References

- [1] Tanachutiwat S, Liu M, Wang W. FPGA based on integration of CMOS and RRAM. *IEEE Trans Very Large Scale Integr (VLSI) Syst* 2011;19(11):2023–32.
- [2] Garbin D, Vianello E, Bichler O, Raffay Q, Gamrat C, Ghibaudo G, et al. HfO₂-based OxRAM devices as synapses for convolutional neural networks. *IEEE Trans Electron Devices* 2015;62(8):2494–501.
- [3] Xu T, Leppanen V. Analysing emerging memory technologies for big data and signal processing applications. In: 2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC). p. 104–9.
- [4] Zambelli C, Grossi A, Walczyk D, Bertaud T, Tillack B, Schroeder T, et al. Statistical analysis of resistive switching characteristics in ReRAM test arrays. In: *IEEE Int. Conf. on Microelectronics Test Structures (ICMTS)*. p. 27–31.
- [5] Zambelli C, Grossi A, Olivo P, Walczyk D, Dabrowski J, Tillack B, et al. Electrical characterization of read window in ReRAM arrays under different SET/RESET cycling conditions. In: *IEEE Int. Memory Workshop (IMW)*. p. 1–4.
- [6] Walczyk D, Bertaud T, Sowinska M, Lukosius M, Schubert MA, Fox A, et al. Resistive switching behavior in TiN/HfO₂/Ti/TiN devices. In: *Int. Semiconductor Conf. Dresden-Grenoble (ISCDG)*. p. 143–6.
- [7] Lorenzi P, Rao R, Irrera F. Forming kinetics in HfO₂-based RRAM cells. *IEEE Trans Electron Devices* 2013;60(1):438–43.
- [8] Grossi A, Walczyk D, Zambelli C, Miranda E, Olivo P, Stikanov V, et al. Impact of intercell and intracell variability on forming and switching parameters in RRAM arrays. *IEEE Trans Electron Devices* 2015;62(8):2502–9.
- [9] Morgan K, Huang R, Pearce S, de Groot C. The effect of atomic layer deposition temperature on switching properties of HfOx resistive RAM devices. In: *IEEE Int. Symp. on Circuits and Systems (ISCAS)*. p. 432–5.
- [10] Kim H-D, Crupi F, Lukosius M, Trusch A, Walczyk C, Wenger C. Resistive switching characteristics of integrated polycrystalline hafnium oxide based one transistor and one resistor devices fabricated by atomic vapor deposition methods. *J Vac Sci Technol B* 2015;33(5):052204.1–4.5.
- [11] Grossi A, Perez E, Zambelli C, Olivo P, Wengeru C. Performance and reliability comparison of 1t-1r RRAM arrays with amorphous and polycrystalline HfO₂. In: *Joint int. EUROSOI workshop and int. conf. on Ultimate Integration on Silicon (EUROSOI-ULIS)*. p. 80–3.
- [12] Miranda EA, Walczyk C, Wenger C, Schroeder T. Model for the resistive switching effect in HfO₂ MIM structures based on the transmission properties of narrow constrictions. *IEEE Electron Device Lett* 2010;31(6):609–11.
- [13] Grossi A, Zambelli C, Olivo P, Miranda E, Stikanov V, Schroeder T, et al. Relationship among current fluctuations during forming, cell-to-cell variability and reliability in RRAM arrays. In: *IEEE Int. Memory Workshop (IMW)*. p. 1–4.
- [14] Iglesias V, Porti M, Nafria M, Aymerich X, Dudek P, Bersuker G. Dielectric breakdown in polycrystalline hafnium oxide gate dielectrics investigated by conductive atomic force microscopy. *J Vac Sci Technol B* 2011;29(1):1–4.
- [15] Chimenton A, Zambelli C, Olivo P, Pirovano A. Set of electrical characteristic parameters suitable for reliability analysis of multimegabit phase change memory arrays. In: *Joint non-volatile semiconductor memory workshop and int. conf. on memory technology and design*. p. 49–51.
- [16] Grossi A, Zambelli C, Olivo P, Miranda E, Stikanov V, Walczyk C, et al. Electrical characterization and modeling of pulse-based forming techniques in RRAM arrays. *Solid State Electron* 2016;115(Part A):17–25.
- [17] Crespo-Yepes A, Martin-Martinez J, Rama I, Maestro M, Rodriguez R, Nafria M, et al. Intra-device statistical parameters in variability-aware modelling of resistive switching devices. In: *Joint int. EUROSOI workshop and int. conf. on Ultimate Integration on Silicon (EUROSOI-ULIS)*. p. 84–7.
- [18] Lian X, Cartoix X, Miranda E, Perniola L, Rurali R, Long S, et al. Multi-scale quantum point contact model for filamentary conduction in resistive random access memories devices. *J Appl Phys* 2014;115(24):1–8.
- [19] Martin-Martinez J, Kaczer B, Degraeve R, Roussel PJ, Rodriguez R, Nafria M, et al. Circuit design-oriented stochastic piecewise modeling of the postbreakdown gate current in MOSFETs: application to ring oscillators. *IEEE Trans Device Mater Reliab* 2012;12(1):78–85.



Inverse-magnetostriction-induced switching current reduction of STT-MTJs and its application for low-voltage MRAM



Yota Takamura^{a,*}, Yusuke Shuto^b, Shu'ichiro Yamamoto^b, Hiroshi Funakubo^c,
Minoru Kuribayashi Kurosawa^a, Shigeki Nakagawa^a, Satoshi Sugahara^{a,b}

^aSchool of Engineering, Tokyo Institute of Technology, Tokyo, Japan

^bLaboratory for Future Interdisciplinary Research of Science and Technology, Tokyo Institute of Technology, Yokohama, Japan

^cSchool of Materials and Chemical Technology, Tokyo Institute of Technology, Yokohama, Japan

ARTICLE INFO

Article history:

Available online 15 October 2016

The review of this paper was arranged by
Viktor Sverdllov

ABSTRACT

A new spin-transfer torque (STT) magnetic tunnel junction (MTJ) using an inverse magnetostriction (IMS) material for the free layer is proposed for low-voltage MRAMs. The MTJ is surrounded by a piezoelectric gate structure so that a pressure for introducing the IMS effect can efficiently be applied to the free layer without any high-yield-strength support structure. During STT-induced magnetization switching, the energy barrier height for the switching can be lowered by the IMS effect, and thus a critical current density (J_C) for the magnetization switching can dramatically be reduced. Energy performance of a low-voltage STT-MRAM cell using the proposed MTJ and a FinFET is also demonstrated.

© 2016 Published by Elsevier Ltd.

1. Introduction

Low-voltage (or near-threshold voltage) operations of complementary metal-oxide-semiconductor (CMOS) logic systems have attracted considerable attention owing to the ability of dramatic reduction of dynamic and static power dissipation [1]. In particular, low-voltage (~ 0.3 – 0.4 V) operations can minimize the energy dissipation (or maximize the energy efficiency) of logic systems [1], and thus this operation mode is promising for always-on applications such as various wearable devices [2]. Nonvolatile data retention adaptable to low-voltage operations is highly requested for these applications. However, in general, it is difficult for flash and other emerging nonvolatile memories to satisfy this requirement, since the low-voltage write operation degrades the data retention performance.

Spin-transfer torque (STT) magnetic tunnel junctions (MTJs) could be a candidate for a low-voltage nonvolatile memory element owing to the current-driven operation behaviour of the MTJs, i.e., as far as the critical current density (J_C) for their current-induced magnetization switching (CIMS) can be obtained at a desired low-voltage, there is no limitation for the (write-) operation voltage. Various efforts including perpendicular magnetic anisotropy electrodes have been paid to reduce J_C [3]. Although J_C can be reduced by lowering the energy barrier (E_B) of MTJs, this leads to the degradation of the thermal stability and thus the retention

ability. Techniques for energy barrier reduction applied *only during CIMS* would be promising for managing low J_C and high thermal stability. Saito et al. [4] proposed a switching field reduction technique based on the inverse magnetostriction (IMS) effect for field-induced magnetization switching of MTJs using a super magnetostriction material for the free layer. This technique would also be applied to J_C reduction for CIMS of STT MTJs. Note that pressures (<1 GPa) required for energy barrier deformation of the IMS layer of such MTJs would be obtained using a piezoelectric (PE) material with a low voltage bias (discussed later).

In this paper, we propose a new STT MTJ using an IMS material for the free layer (hereafter, referred to as an IMS-MTJ) and computationally investigate IMS-induced switching current reduction of the IMS-MTJ. Energy performance of a low-voltage MRAM cell using the proposed IMS-MTJ and a high performance FinFET is also demonstrated.

2. Proposed IMS-MTJ

Fig. 1(a) shows schematic device structures of the proposed STT MTJ using an IMS material for the free layer. The device is comprised of perpendicularly magnetized ferromagnetic electrodes (pinned and free layers), a tunnel barrier (TB), and a PE gate surrounding the MTJ part. The torus-shape PE gate structure is originally proposed for a piezoelectronic transistor [5]. The free layer consists of the IMS layer and high spin-polarization interfacial layer facing the TB. These two layers are magnetically coupled, i.e., the magnetization direction of the interfacial layer is governed

* Corresponding author.

E-mail address: takamura@ee.e.titech.ac.jp (Y. Takamura).

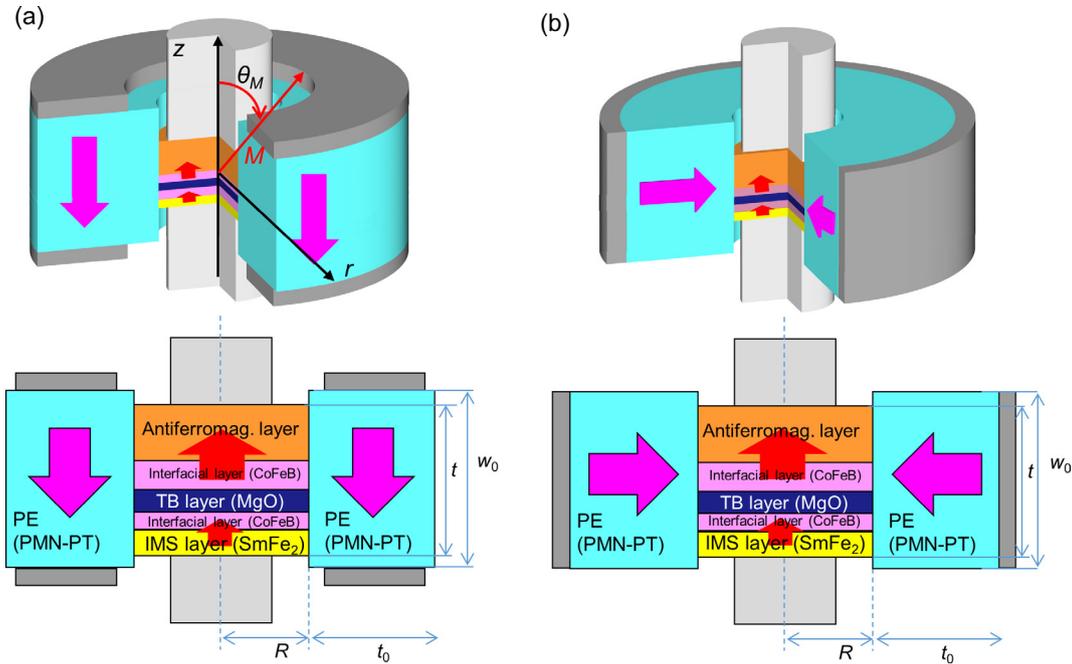


Fig. 1. (a) Schematic device structure and cross section of a proposed IMS-MTJ. Another possible structure is shown in (b).

by that of the IMS layer. The PE gate acts as a stressor to the free layer. The polarization of the PE material is parallel to the z -axis (see Fig. 1(a)), and the electrodes of the PE gate are placed so as to apply an electric field parallel to the polarization. The PE gate exerts a compressive stress to the IMS layer. Therefore, the IMS layer needs to possess a negative magnetostrictive constant to reduce the energy barrier for the magnetization switching. Another possible configuration of the PE gate is shown in Fig. 1(b). The polarization of the PE material is toward the central axis of the MTJ pillar as shown in the figure, and thus the electrode is formed to surround the outer periphery of the PE material.

3. Modelling and calculation procedure

IMS-induced STT magnetization switching behaviour was analysed based on the LLG (Landau-Lifshits-Gilbert) equation with Slonczewski's STT term [6,7]. In this study, we assumed that the IMS free layer had no in-plane magnetic anisotropy. When STT is exerted on the IMS layer of the IMS-MTJ, the time-evolution of the tilt angle (θ_M) component of the magnetization can be derived from the LLG equation:

$$\frac{M}{\gamma} \frac{d\theta_M}{dt} = -\alpha \frac{\partial}{\partial \theta} \{E_i(\theta_M) + E_{STT}(\theta_M)\}, \quad (1)$$

where γ is the gyromagnetic constant, M the saturation magnetization of the IMS layer, α the damping factor, θ the axis for the tilt angle of the magnetization. $E_i(\theta_M)$ is the internal energy of the magnetization given by

$$E_i(\theta_M) = \left(K_U - 2\pi M^2 + \frac{2}{3}\lambda P\right) V \sin^2 \theta_M \quad (2)$$

where K_U the uniaxial magnetic anisotropy energy density, λ the magnetostrictive constant (<0), P the pressure applied to the IMS layer, and V the volume of the IMS free layer. The first, second, and third terms in $E_i(\theta_M)$ represent the uniaxial magnetic anisotropy energy, the demagnetization energy, and the magnetostrictive energy that expresses the IMS effect, respectively. Here, the IMS layer is assumed to have a cylindrical shape, and P is applied toward

the central axis of the IMS-MTJ from the outer periphery of the IMS layer. $E_{STT}(\theta_M)$ in Eq. (1) represents the potential energy for STT and it is given by [8],

$$E_{STT}(\theta_M) = \frac{\hbar \eta}{2e\alpha} J \cos \theta_M, \quad (3)$$

where η is the spin transfer efficiency and I the current passing through the IMS layer. In this paper, the CIMS behaviour accompanied with/without the thermal excitation effect [9,10] were analysed using the energy curve $E(\theta_M) = E_i(\theta_M) + E_{STT}(\theta_M)$ with its effective energy barrier height E_B for the magnetization switching. E_B is given by the difference between the maximum and initial value (at $\theta_M = 0$) of $E(\theta_M)$ as follows:

$$\begin{aligned} E_B &= E(\theta_{\max}) - E(0) \\ &= \left(K_U - 2\pi M^2 + \frac{2}{3}\lambda P\right) V \left[1 - \frac{\hbar \eta}{2e\alpha l_m} \frac{J}{2(K_U - 2\pi M^2 + \frac{2}{3}\lambda P)}\right]^2 \end{aligned} \quad (4)$$

where J is the current density passing through the IMS-MTJ and l_m the thickness of the IMS layer. θ_{\max} is given by

$$\theta_{\max} = \cos^{-1} \frac{\frac{\hbar \eta}{2e\alpha} J}{2(K_U - 2\pi M^2 + \frac{2}{3}\lambda P) l_m} \quad (5)$$

Eq. (4) clearly shows that E_B can be reduced by P and J , i.e., by both the IMS and STT effects. The critical current density $J_{CO}(P)$ for the magnetization switching at $T = 0$ K can be defined as a current density when $E_B = 0$:

$$J_{CO}(P) = \frac{4e\alpha l_m}{\hbar \eta} \left(K_U - 2\pi M^2 + \frac{2}{3}\lambda P\right), \quad (6)$$

Note that when $P = 0$, $J_{CO}(P)$ corresponds to Slonczewski's critical current density at $T = 0$ K for ordinary STT-MTJs [7]. For the description of the critical current density J_{CT} at finite temperatures, a bit error rate (BER) of the IMS-MTJ needs to be introduced. Using BER of 10^{-x} , J_{CT} can be written using the Koch model [10,11]:

$$J_{CT}(P) = J_{CO}(P) \left[1 - \frac{k_B T}{E_B(P) V} \ln \left(\frac{1}{x \ln 10} \frac{t_p}{\tau_0}\right)\right], \quad (7)$$

Table 1
Simulation parameters for IMS-MTJ.

Tunneling magnetoresistance: TMR	100%
Resistance-area product: RA	$2 \Omega \cdot \mu\text{m}^2$
Voltage at half-maximum TMR: V_{half}	0.5 V
J_{CT} without the IMS effect	$2.5 \times 10^6 \text{ A/cm}^2$
J_{CT} with the IMS effect	$0.1 \times 10^6 \text{ A/cm}^2$
Resistance: $R_{\text{P}}(0)$ (Parallel mag.)	6.36 k Ω
$R_{\text{AP}}(0)$ (Antiparallel mag.)	12.7 k Ω

where k_{B} is the Boltzmann constant, T the temperature, t_{p} the pulse width of the current and pressure applied to the IMS layer, and τ_0 the attempt time.

In this study, SmFe_2 [12] is used for the IMS material, since it possesses a large negative magnetostrictive constant. A device structure consisting of a $\text{CoFeB}/\text{MgO}/\text{CoFeB}/\text{SmFe}_2$ MTJ part and a PMN-PT [13] gate surrounding it is employed. The material constants and device parameters used in this study are shown in Tables 1 and 2, which were determined by reference to reported data [12–16]. Note that Δ is defined by the energy barrier height for the switching at $P = 0$ MPa, i.e., $E_{\text{B}}(P = 0) = \Delta$.

Table 2
Material constants and device parameters for IMS-MTJ.

MTJ part										PE gate part					
K_{u} (MJ/m ³)	M (T)	λ (ppm)	α	R (nm)	l_{m} (nm)	t (nm)	Δ	H	τ_0 (ns)	Y_{MTJ} (GPa)	Y_{PE} (GPa)	ν	ϵ_{33}^{S}	d_{31} (nm/V)	w_0 (nm)
0.56	0.64	−1258	0.005	10	2	10	60	1	1	40	60	0.3	741	−0.852	10

Y_{MTJ} , Y_{PE} : Young modules of the MTJ and PE parts, ν : Poisson's ratio, ϵ_{33}^{S} : Relative permittivity, d_{31} : Piezoelectric strain constant.

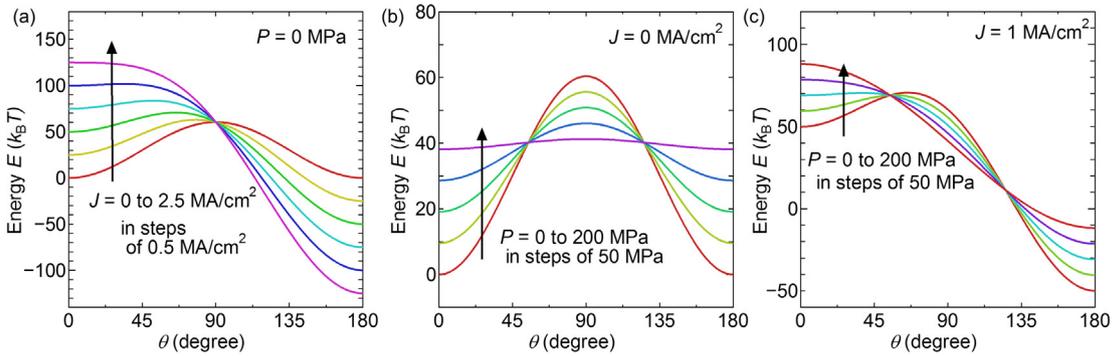


Fig. 2. Energy curves $E(\theta_{\text{M}})$ for the IMS free layer of the IMS-MTJ, in which (a) P is varied with $J = 0$, (b) J is varied with $P = 0$, and (c) P is varied with $J = 1 \text{ MA/cm}^2$. J and P represent current density passing through the device and pressure applied to the IMS layer, respectively.

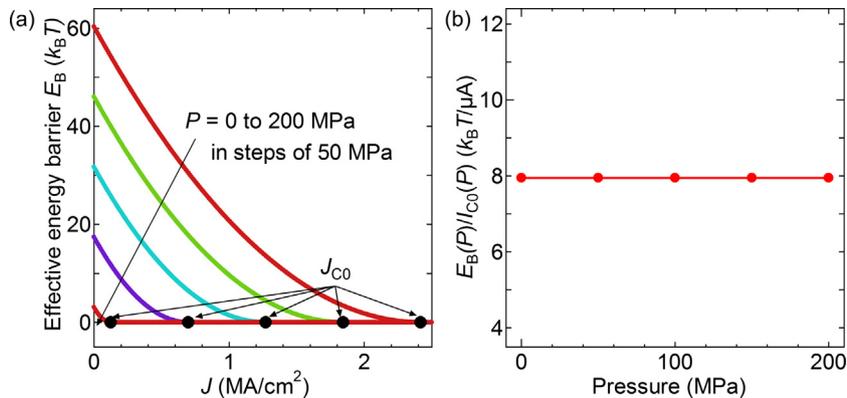


Fig. 3. (a) Effective energy barrier E_{B} as a function of J , in which P is varied from 0 to 200 MPa in steps of 50 MPa. (b) Spin-transfer torque magnetization switching efficiency of the IMS-MTJ as a function of P .

Operations of proposed STT-MRAM cells (shown later) were analysed by HSPICE with a 20-nm-technology FinFET PTM [17] and our developed MTJ macromodel [18]. This macromodel can closely fit experimentally observed electrical characteristics of ordinary MTJs within an error of 1.5% [18].

4. IMS-MTJ characteristics

Fig. 2 shows $E(\theta_{\text{M}})$ for the IMS free layer as a function of tilt angle θ_{M} of the magnetization (see Fig. 1(a)). The energy curve can be deformed by the STT effect, resulting that the effective energy barrier height E_{B} for the magnetization switching is reduced by J (Fig. 2(a)). This behaviour is consistent with the free layer of conventional STT-MTJs. This effective energy barrier can be also lowered by applying P to the free layer owing to the IMS effect (Fig. 2(b)). The STT magnetization switching (CIMS) occurs when the energy barrier disappears, and thus the IMS-induced barrier lowering is effective at reducing the critical current of the CIMS, as shown in Fig. 2(c). The effective barrier decreases with increasing J depending on P , and J_{C0} can be reduced with increasing P , as shown in Fig. 3(a). For ordinary STT-MTJs, J_{C0} can be reduced by

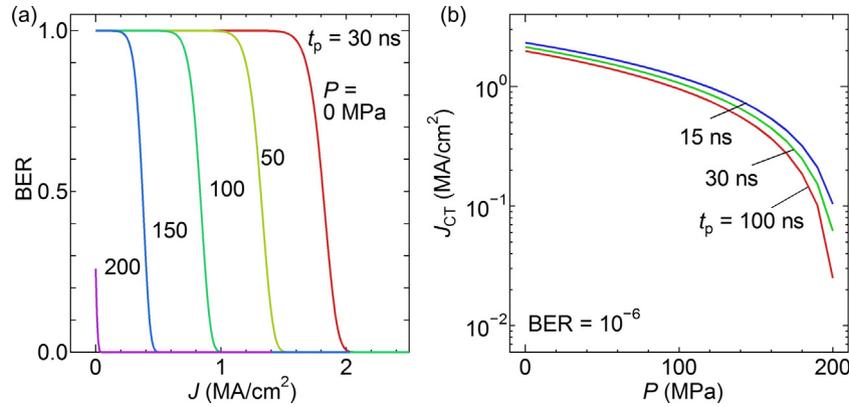


Fig. 4. (a) BER as a function of J , in which P is varied from 0 to 200 MPa in steps of 50 MPa. (b) J_{CT} as a function of P for various t_p .

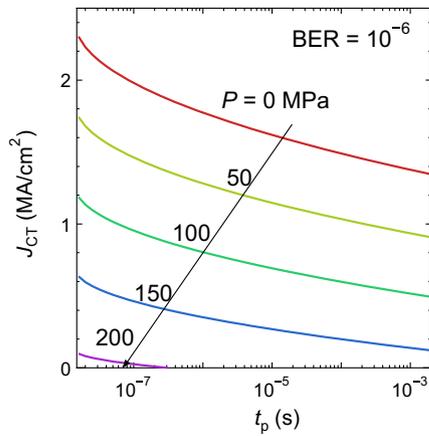


Fig. 5. J_{CT} as a function of t_p for $BER = 10^{-6}$, in which P is varied from 0 to 200 MPa in steps of 50 MPa.

lowering Δ ($=E_B|_{J=0}$), which degrades thermal stability. However, J_{C0} also diminished by P without reducing Δ for the proposed IMS-MTJ, i.e., the IMS-MTJs can possess high thermal stability or retention ability. Fig. 3(b) shows CIMS efficiency of IMS-MTJs as a function of P . The CIMS efficiency is given by a ratio of the effective energy barrier $E_B(P)$ for the switching to the critical current $I_{C0}(P)$ [19]. The CIMS efficiency remains constant at a reasonable value regardless of P .

Fig. 4(a) shows the BER as a function of J at room temperature, in which P is varied and the pulse width t_p for J and P are fixed at 30 ns. The BER rapidly increases with decreasing J for each P condition, when J is lower than $J_{C0}(P)$. Therefore, the critical current density J_{CT} determined by a given BER is important in practice and used for the following discussion. Fig. 4(b) shows J_{CT} for $BER = 10^{-6}$ as a function of P , in which pulse width t_p is varied. J_{CT} can be effectively reduced by increasing P depending on t_p .

Fig. 5 shows J_{CT} as a function of t_p for $BER = 10^{-6}$. The non-linear behaviour (concave upward decreasing function) of the J_{CT} - t_p curve, which is commonly observed for ordinary STT-MTJs [9], appears for all the pressure conditions and J_{CT} increases with decreasing t_p . The magnetization switching with $J_{CT} = 0.1 \text{ MA/cm}^2$ and $BER = 10^{-6}$ can be achieved for a moderate t_p (several tens of nanoseconds).

5. Low-voltage MRAM application

Fig. 6(a) and (b) shows two types of MRAM cells using the IMS-MTJ shown in Fig. 1(a). The cell shown in Fig. 6(a) is the same

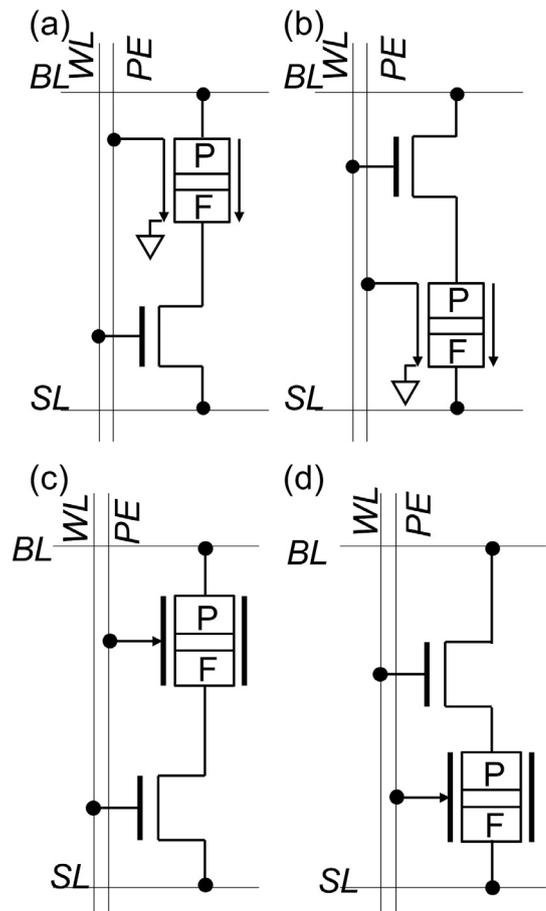


Fig. 6. Low-voltage MRAM cells using an IMS-MTJ with (a) drain-side MTJ connection (DSM) and (b) source-side MTJ connection (SSM) configurations for the IMS-MTJs shown in Fig. 1(a). (c) and (d) also show DSM and SSM cells using the IMS-MTJ shown in Fig. 1(b), respectively.

Table 3
Read/write operation architecture.

	Read	Write (AP → P)	Write (P → AP)
BL	H	L	H
SL	L	H	L
WL	H	H	H
PE	L	H	H

configuration as conventional MRAM cells, i.e., the cell configuration with the drain-side MTJ connection (DSM cell). The cell shown

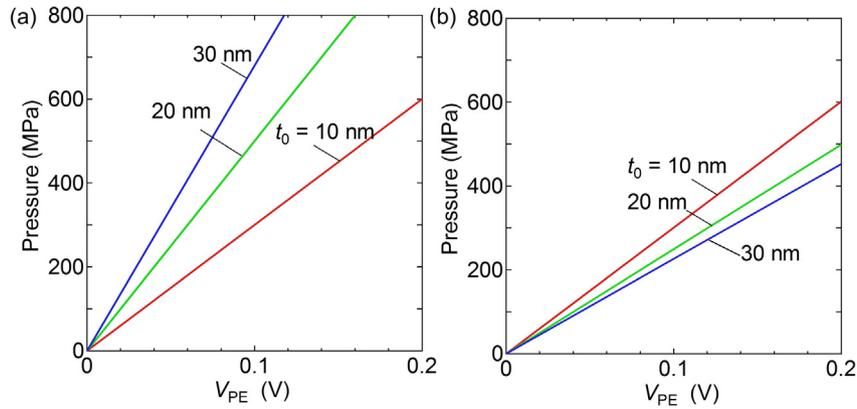


Fig. 7. P as a function of V_{PE} for the IMS-MTJs shown in Fig. 1(a) and (b). V_{PE} represents a bias voltage of the PE gate.

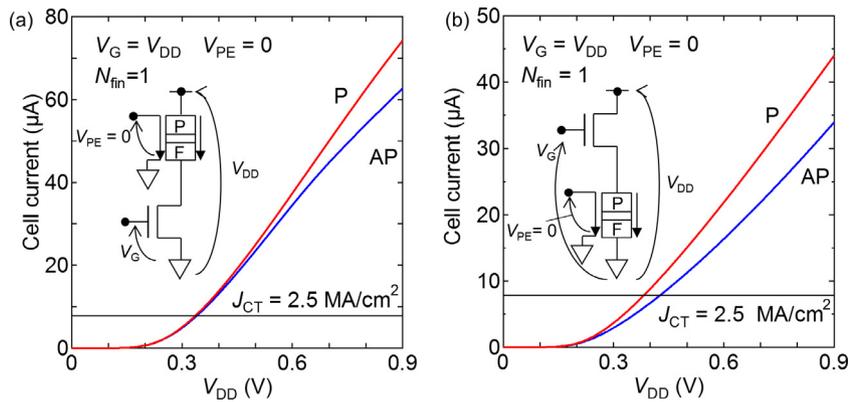


Fig. 8. Cell currents as a function of V_{DD} for (a) the DSM cell and (b) the SSM cell during the read operation mode.

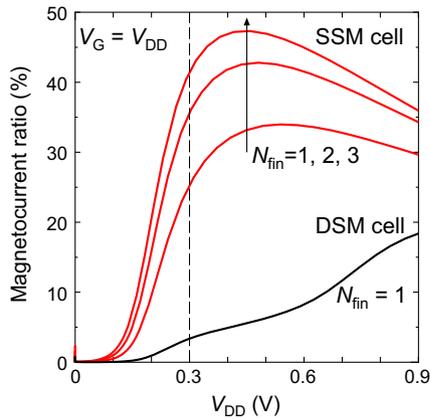


Fig. 9. Magnetocurrent ratio as a function of V_{DD} for the DSM and SSM cells during the read operation mode.

in Fig. 6(b) has a configuration with the source-side MTJ connection (SSM cell), in which the connected MTJ feeds back its voltage drop to the gate of the MOSFET. Since the degree of this negative feedback depends on the resistance states of the MTJ, the cell currents of the SSM cell can be effectively controlled by the magnetization configuration of the MTJ (shown later). In this study, a FinFET is used as a selector transistor for both the cells. Table 3 shows an operation architecture of these cells. Note that the IMS effect is not induced during the read operation, and it is employed only during the write operation. Fig. 7(a) and (b) shows P as a function of V_{PE} (that is a bias voltage of the PE gate; see Fig. 1 and

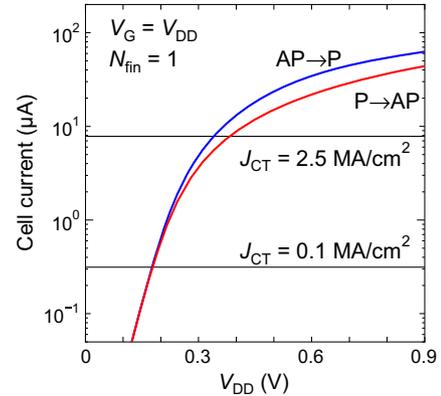


Fig. 10. Cell currents as a function of V_{DD} for the SSM cell during the write operation mode.

Table 2) for the IMS-MTJs shown in Fig. 1(a) and (b), respectively. P required for the IMS-induced switching current reduction described above (several hundreds of MPa) can easily be yielded by $V_{PE} = 0.2$ V or less, when a thin piezoelectric material is employed for the PE gate (see Table 2). Fig. 6(c) and (d) shows DSM and SSM cells using the IMS-MTJ shown in Fig. 1(b).

Fig. 8(a) and (b) shows cell currents as a function of V_{DD} for the DSM and SSM cells, respectively, during the read operation, in which V_{PE} and V_G are set to zero and V_{DD} , respectively. Although the cell currents of the SSM cell are lower than those of the DSM cell, its magnetocurrent ratio (that is a rate of change of the cell

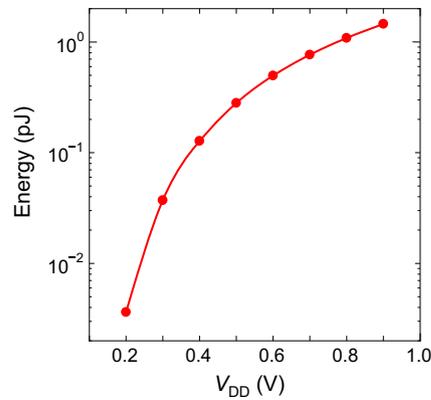


Fig. 11. Write energy as a function of V_{DD} for the SSM cell.

currents in the parallel and antiparallel configurations: $\gamma_{MC} = (I_P - I_{AP})/I_{AP}$ is sufficiently high to distinguish the parallel and antiparallel states even at $V_{DD} = 0.3$ V, as shown in Fig. 9. In addition, the magnetocurrent ratio can be enhanced by the number of the fin channel of the FinFET, as shown in Fig. 9. On the other hand, the magnetocurrent ratio of the DSM cell is severely degraded for lower V_{DD} operations. Note that when V_{DD} is less than 0.4 V, the magnetization switching does not occur during the read operation (since the cell currents do not exceed J_{CT} for $V_{PE} = 0$ V; see Table 1).

Fig. 10 shows cell currents during the write operation for the SSM cell (in which V_{PE} is applied). The cell current can exceed J_{CT} for the IMS-induced switching, when $V_{DD} \geq 0.2$ V. Fig. 11 shows the write energy for the SSM cells as a function of V_{DD} . Here, the energy to drive STT currents and the static energy for the PE gate to apply a pressure to the IMS free layers are included in the calculation. By reducing V_{DD} from 0.9 V to 0.2 V, the write energy can be considerably reduced to 1/400. Note that the energy consumption of the PE gate is negligibly small in comparison with that of CIMS.

6. Conclusions

A new IMS-MTJ was proposed and computationally analysed for low-voltage STT-MRAM applications. The free layer of the IMS-MTJ consists of a thin IMS material film and high spin-polarization interfacial layer facing the tunnel barrier, and it is surrounded by a PE gate so that a pressure for introducing the IMS effect can efficiently be applied to the free layer without any high-yield-strength support structure. During STT-induced magnetization switching,

the energy barrier height for the switching can be lowered by the IMS effect, and thus J_C can be dramatically reduced. The write energy of a low-voltage STT-MRAM cell using an IMS-MTJ with a FinFET can be hugely reduced compared to conventional MRAM cells.

Acknowledgement

This study was partly supported by JSPS KAKENHI Grant and JST CREST program. This work is also supported by VLSI Design and Education Center (VDEC), The University of Tokyo with the collaboration with Synopsys Corporation.

References

- [1] Jain S, et al. A 280 mV-to-1.2 V wide-operating-range IA-32 processor in 32 nm CMOS. In: 2012 IEEE ISSCC, paper 3.6.
- [2] Lee K, Kan JJ, Kang SH. Unified embedded non-volatile memory for emerging mobile markets. In: 2014 IEEE ISLPEd. p. 131–6.
- [3] Mangin S et al. Current-induced magnetization reversal in nanopyllars with perpendicular anisotropy. *Nat Mater* 2006;5(March):210–5.
- [4] Saito N, Yamada M, Nakagawa S. Improvement of stress-induced magnetization reversal process of DyFeCo thin films. *J Appl Phys* 2008;103 (January). 07A706/1–3.
- [5] Sugahara S, Shuto Y, Yamamoto S, Funakubo M, Kurosawa MK., in preparation.
- [6] Sun JZ. Spin-current interaction with a monodomain magnetic body: a model study. *Phys Rev B* 2000;62(July):570–8.
- [7] Slonczewski JC. Currents and torques in metallic magnetic multilayers. *J Magn Magn Mater* 2002;247(March):324–38.
- [8] Takamura Y, Nakagawa S, Sugahara S., in preparation.
- [9] Hosomi M, Yamagishi H, Yamamoto T, Bessho K, Higo Y, Yamane K, et al. A novel nonvolatile memory with spin torque transfer magnetization switching: spin-RAM. *IEDM Technical Digest*; 2009. p. 459–62.
- [10] Koch RH, Katine JA, Sun JZ. Time-resolved reversal of spin-transfer switching in a nanomagnet. *Phys Rev Lett* 2004;92(February):088302/1–2/4.
- [11] Butler WH et al. Switching distributions for perpendicular spin-torque devices within the macrospin approximation. *IEEE Trans Magn* 2012;48 (December):4684–700.
- [12] Samata H, Fujiwara N, Nagata Y, Uchida T, Lan MD. Crystal growth and magnetic properties of SmFe_2 . *Jpn J Appl Phys* 1998;37(October):5544–8.
- [13] Newns D, Elmegeen B, Liu XH, Martyna G. A low-voltage high-speed electronic switch based on piezoelectric transduction. *J Appl Phys* 2012;111 (April):084509/1–084509/18.
- [14] Ikeda S et al. A perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction. *Nat Mater* 2010;9(July):721–4.
- [15] Park JH, et al. Enhancement of data retention and write current scaling for sub-20 nm STT-MRAM by utilizing dual interfaces for perpendicular magnetic anisotropy. In: 2012 VLSI, paper 7.1.
- [16] Gajek M et al. Spin torque switching of 20 nm magnetic tunnel junctions with perpendicular anisotropy. *Appl Phys Lett* 2012;100(March):132408/1–8/3.
- [17] Predictive Technology Model (PTM) <<http://ptm.asu.edu/>>.
- [18] Yamamoto S, Sugahara S. Nonvolatile static random access memory using magnetic tunnel junctions with current-induced magnetization switching architecture. *Jpn J Appl Phys* 2009;48(November):043001/1–1/7.
- [19] Hu G, et al. STT-MRAM with double magnetic tunnel junctions. In: 2015 IEDM, paper 26.3.

CONTENTS—continued from inside back cover]

C. NAVARRO, S. BARRAUD, S. MARTINIE, J. LACORD, M.-A. JAUD and M. VINET: Reconfigurable field effect transistor for advanced CMOS: Advantages and limitations 155

M. M. VIGNETTI, F. CALMON, P. LESIEUR and A. SAVOY-NAVARRO: Simulation study of a novel 3D SPAD pixel in an advanced FD-SOI technology 163

S. ATHANASIOU, C.-A. LEGRAND, S. CRISTOLOVEANU and P. GALY: Reconfigurable ultra-thin film GDNMOS device for ESD protection in 28 nm FD-SOI technology. 172

H. EL Dirani, P. FONTENEAU, Y. SOLARO, C.-A. LEGRAND, D. MARIN-CUDRAZ, P. FERRARI and S. CRISTOLOVEANU: Sharp-switching band-modulation back-gated devices in advanced FDSOI technology 180

A. GROSSI, C. ZAMBELLI, P. OLIVO, A. CRESPO-YEPES, J. MARTIN-MARTINEZ, R. RODRÍGUEZ, M. NAFRIA, E. PEREZ and C. WENGER: Electrical characterization and modeling of 1T-1R RRAM arrays with amorphous and poly-crystalline HfO₂ 187

Y. TAKAMURA, Y. SHUTO, S. YAMAMOTO, H. FUNAKUBO, M. K. KUROSAWA, S. NAKAGAWA and S. SUGAHARA: Inverse-magnetostriction-induced switching current reduction of STT-MTJs and its application for low-voltage MRAM. 194

CONTENTS—continued from outside back cover

C. SCHULTE-BRAUCKS, S. GLASS, E. HOFMANN, D. STANGE, N. VON DEN DRIESCH, J. M. HARTMANN, Z. IKONIC, Q. T. ZHAO, D. BUCA and S. MANTL: Process modules for GeSn nanoelectronics with high Sn-contents.	54
B. C. PAZ, M. CASSÉ, S. BARRAUD, G. REIMBOLD, M. VINET, O. FAYNOT and M. A. PAVANELLO: Study of silicon n- and p-FET SOI nanowires concerning analog performance down to 100 K.	60
A. S. N. PEREIRA, G. DE STREEL, N. PLANES, M. HAOND, R. GIACOMINI, D. FLANDRE and V. KILCHYTSKA: An in-depth analysis of temperature effect on DIBL in UTBB FD SOI MOSFETs based on experimental data, numerical simulations and analytical models	67
R. BERTHELON, F. ANDRIEU, S. ORTOLLAND, R. NICOLAS, T. POIROUX, E. BAYLAC, D. DUTARTRE, E. JOSSE, A. CLAVERIE and M. HAOND: Characterization and modelling of layout effects in SiGe channel pMOSFETs from 14 nm UTBB FDSOI technology	72
H. J. PARK, L. PIRRO, L. CZORNOMAZ, I. IONICA, M. BAWEDIN, V. DJARA, V. DESHPANDE and S. CRISTOLOVEANU: Back-gated InGaAs-on-insulator lateral N ⁺ NN ⁺ MOSFET: Fabrication and typical conduction mechanisms	80
V. DESHPANDE, V. DJARA, E. O'CONNOR, P. HASHEMI, K. BALAKRISHNAN, D. CAIMI, M. SOUSA, L. CZORNOMAZ and J. FOMPEYRINE: DC and RF characterization of InGaAs replacement metal gate (RMG) nFETs on SiGe-OI FinFETs fabricated by 3D monolithic integration.	87
D. TOMASZEWSKI, G. GŁUSZKO, L. ŁUKASIAK, K. KUCHARSKI and J. MALESIŃSKA: Elimination of the channel current effect on the characterization of MOSFET threshold voltage using junction capacitance measurements	92
D. BOUDIER, B. CRETU, E. SIMOEN, R. CARIN, A. VELOSO, N. COLLAERT and A. THEAN: Low frequency noise assessment in n- and p-channel sub-10 nm triple-gate FinFETs: Part I: Theory and methodology	102
D. BOUDIER, B. CRETU, E. SIMOEN, R. CARIN, A. VELOSO, N. COLLAERT and A. THEAN: Low frequency noise assessment in n- and p-channel sub-10 nm triple-gate FinFETs: Part II: Measurements and results	109
C. MARQUEZ, N. RODRIGUEZ, F. GAMIZ and A. OHATA: Systematic method for electrical characterization of random telegraph noise in MOSFETs	115
B. KAZEMI ESFEH, S. MAKOVEJEV, D. BASSO, E. DESBONNETS, V. KILCHYTSKA, D. FLANDRE and J.-P. RASKIN: RF SOI CMOS technology on 1 st and 2 nd generation trap-rich high resistivity SOI wafers	121
C. JUNGEMANN, T. LINN, K. BITTNER and H.-G. BRACHTENDORF: Numerical investigation of plasma effects in silicon MOSFETs for THz-wave detection	129
V. ŠIMONKA, G. NAWRATIL, A. HÖSSINGER, J. WEINBUB and S. SELBERHERR: Anisotropic interpolation method of silicon carbide oxidation growth rates for three-dimensional simulation.	135
P. MANSTETTEN, L. FILIPOVIC, A. HÖSSINGER, J. WEINBUB and S. SELBERHERR: Framework to model neutral particle flux in convex high aspect ratio structures using one-dimensional radiosity.	141
T. BALDAUF, A. HEINZIG, J. TROMMER, T. MIKOLAJICK and W. M. WEBER: Tuning the tunneling probability by mechanical stress in Schottky barrier based reconfigurable nanowire transistors	148



CONTENTS

Abstracted / Indexed in: Res. Alert, Cam. Sci. Abstr., Chem. Abstr. Serv., Curr. Cont./Phys. chem. & Earth Sci., Curr Cont./Eng. Tech. & Appl. Sci., Curr. Tech. Indx, Eng. Indx, INSPEC Data., PASCAL-CNRS Data., Curr. Cont. Sci. Indx, Curr. Cont. SCISEARCH Data., SSSA/CISA/ECA/ISMEC, Mater. Sci. Cit. Indx, Appl. Sci. & Tech. Indx, Wilson Appl. Sci. & Tech. Abstr. Also covered in the abstract and citation database Scopus®. Full text available on ScienceDirect®.

ISSN 0038-1101

SPECIAL ISSUE: EXTENDED PAPERS SELECTED FROM EUROSOI-ULIS 2016

Guest Editors

Viktor Sverdlov and Siegfried Selberherr

V. SVERDLOV and S. SELBERHERR: Special Issue of Solid-State Electronics, dedicated to EUROSOI-ULIS 2016	1
J. M. C. STORK IEEE Fellow and G. P. HOSEY: SOI technology for power management in automotive and industrial applications.	3
B. MOHAMAD, C. LEROUX, D. RIDEAU, M. HAOND, G. REIMBOLD and G. GHIBAUDDO: Reliable gate stack and substrate parameter extraction based on C-V measurements for 14 nm node FDSOI technology	10
M. A. ELMESSARY, D. NAGY, M. ALDEGUNDE, N. SEOANE, G. INDALECIO, J. LINDBERG, W. DETTMER, D. PERIĆ, A. J. GARCIA-LOUREIRO and K. KALNA: Scaling/LER study of Si GAA nanowire FET using 3D finite element Monte Carlo simulations	17
N. SANO: Variability and self-average of impurity-limited resistance in quasi-one dimensional nanowires.	25
T. A. KARATSORI, C. G. THEODOROU, S. HAENDLER, C. A. DIMITRIADIS and G. GHIBAUDDO: Drain current local variability from linear to saturation region in 28 nm bulk NMOSFETs	31
S. STRANGIO, P. PALESTRI, M. LANUZZA, D. ESSENI, F. CRUPI and L. SELMI: Benchmarks of a III-V TFET technology platform against the 10-nm CMOS FinFET technology node considering basic arithmetic circuits	37
P. G. D. AGOPIAN, J. A. MARTINO, A. VANDOOREN, R. ROOYACKERS, E. SIMOEN, A. THEAN and C. CLAEYS: Study of line-TFET analog performance comparing with other TFET and MOSFET architectures	43
C. MEDINA-BAILON, C. SAMPEDRO, F. GÁMIZ, A. GODOY and L. DONETTI: Confinement orientation effects in S/D tunneling.	48

[continued on inside back cover



0038-1101(201702)128:C;1-K