Sensors & Transducers, Vol. 235, Issue 7, July 2019, pp. 15-22

Sensors & Transducers

Published by IFSA Publishing, S. L., 2019
http://www.sensorsportal.com

A Novel 3D Integrated Circuit Technology Based on Stacked FinFET-CMOS Structure

[*]Jin HE, Yuan REN, Jun PAN, Jingjing LIU and Mansun CHAN
SoC Key Laboratory, Peking University Shenzhen Institute and PKU-HKUST Shenzhen-Hong Kong Institution, 518057, China
Tel.: 86-755-64321156, fax: 86-755-64321155
[*] E-mail: frankhe@pku.edu.cn

Received: 15 April 2019 /Accepted: 28 May 2019 /Published: 31 July 2019

Abstract: This paper describes a methodology to combine the most scalable FinFET structure and CMOS integration processing to form stacked FinFET-CMOS technology used in 3-D integration circuits. The scalability of the proposed 3-D technology in the nanocale dimension is achieved through the promise of the double-gate structure offered by the FinFETs. The stack FinFET-CMOS process is realized in FEOL processing and can achieve extremely high packing density, which results in the 40 to 60 % reduction of capacitive loading. Various standard cells have been re-designed by the FinFET-CMOS technology resulting in significant reduction in cell size. The advantages in conventional circuit performance with the new 3-D standard cell library have also been demonstrated.

Keywords: FinFET, 3-D integration circuits, Processing technology, Performance enhancement, Capacitive load.

1. Introduction

There are a number of challenges in the migration of processing technologies to the sub-50 nm regime. At the circuit level, interconnect RC delay has become the major barrier to high circuit performance due to the fact it plays the almost same role in the circuit function determination with the intrinsic transistor in the nanoscale circuit technology [1]. From device perspective, the issues of short-channel effects, gate dielectric reliability, doping fluctuations, and series source/drain resistance have been listed in the ITRS 2004 roadmap [2]. The practical solutions to the above problem should be compatible with the well-established CAD infrastructure for circuit and system design. All these challenges made semiconductor industry and the microelectronics experts do intensive effort to find various solutions for the circuit processing technology, device new structure, and the CAD tools [3-5].

As traditional solid state CMOS integrated circuit technologies are scaling down to its technology and performance limits, FinFET device structure, invented by BSIM group from University of California at Berkeley may become the most possible alternative to the conventional bulk MOSFET because of a number of advantages such as ideal 60-mV/dec, excellent short-channel-effects (SCEs) immunity, and unique mobility enhancement. Since the first demonstration on IEDM'1998, the process technology for fabricating this device, the device physics for understanding of the device behavior, and a compact model which serves as a link between process technology and circuit design, allows large-scale evaluation of the FinFET performance, even the circuit performance evaluation and test, made significant advancement in recent several years. All these paved a smooth way for

FinFET application in the near future ULSI integrated circuits [6-10]. However, the past FinFET works only focused on the intrinsic transistor performance and circuit function demonstration and application, there has not been reported the impact of the interconnection on the circuit performance. Moreover, the CAD methodology for the FinFET and related circuit design still followed the traditional bulk CMOS routine. For the nanoscale FinFET circuits, it is expected that the traditional interconnection technology will significantly influence the intrinsic transistor performance benefit. On the same time, the CAD methodology may need any change to achieve the FinFET advantages over the bulk MOSFETs.

In this work, we proposed and demonstrated that a 3D Stacked FinFET-CMOS technology can potentially provide a solution to the above challenges. 3D integrated circuits (3D IC's) with multiple active layers stacked in the vertical direction have been suggested by ITRS roadmap [2] as a method to reduce the impact of interconnect loading by providing shorter connection paths between transistors through vertical wires. However, the approach to implement 3D IC's has been based on pre-fabrication of circuits on different layers and vertically stacked through metal-to-metal bonding [11-12] or through silicon vias [13-14]. These approaches have to overcome a number of issues including (1) a special CAD is needed to perform the vertical routing; (2) low vertical routing density; and (3) poor heat dissipation for transistors on the upper layers leading to a significant increase in device temperature during operation.

In this paper, we describe a methodology which combines the FinFET technology and 3D integration using a local-clustering technique to utilize the advantages of both FinFET and 3D integration with a relatively simple fabrication method based on past several works [15-19, 21]. The design methodology and fabrication process is a quasi-2D and compatible with the existing CAD tools. In this approach, only the standard cells in the design library are redesigned with closely coupled 3D stacked transistors. By making the cell size smaller, the total lateral interconnect length can be reduced.

2. Fabrication Process and Device Performance

The structure of the stacked FinFET is illustrated in Fig. 1. It consists of a π-shape gate straddles on a stacked fin that serves as the shared gate for both the NMOSFET and PMOSFET. The shared gate is used as an implant mask to achieve self-alignment for both N- and P-MOSFET source/drain regions. Between the two silicon active layers, there is an insulation layer to isolate the top and bottom devices.

The fabrication process of the 3D stacked FinFET-CMOS technology together with the SEM cross-section of a fabricated inverter are shown in Fig. 2 to illustrate the main features. The key processing steps are described here. Starting with a double layer of SOI wafer, a layer of LTO was deposited to serve as a hard mask for the subsequent processing as shown in Fig. 2(a). The multi-layer films are etched together to form the stacked fins as shown in Fig. 2(b). The gate oxide is then grown and followed by the deposition of in-situ doped n+ gate polysilicon (Fig. 2(c)). The gate is defined by lithography resulting in the structure as shown in Fig. 2(d). Doping of the source/drain region of the top and bottom FinFETs is performed by ion implantation. The doping of the bottom is achieved by high energy implant through the top layer (Fig. 2(e)). In order to connect to the bottom layer, the unwanted regions of the top silicon layer are etched to expose the bottom layer at the desired locations. The contacts where the top transistors are connected to the bottom ones, are also opened in this step (Fig. 2(f)). It is followed by Phosphosilicate Glass (PSG) passivation and contact opening to various layers.

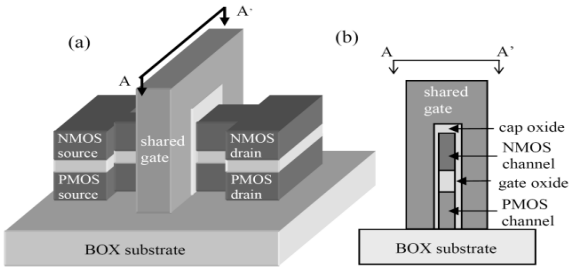


Fig. 1. (a) 3-D schematic view, and (b) cross-section of the stacked FinFET structure.

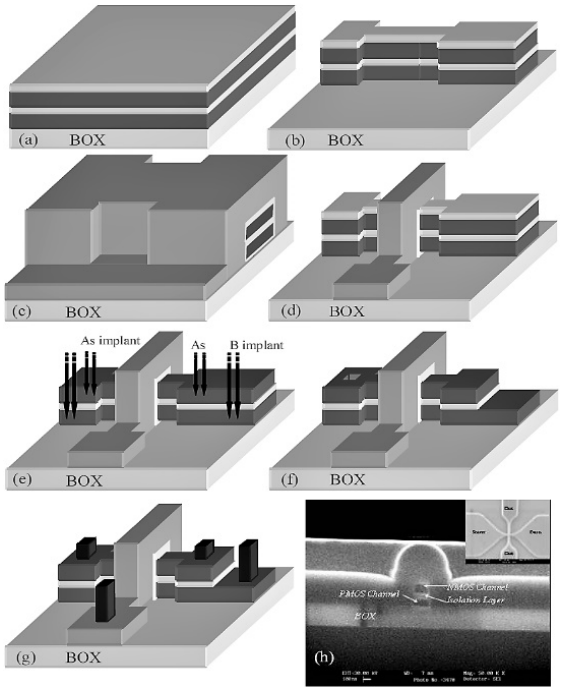
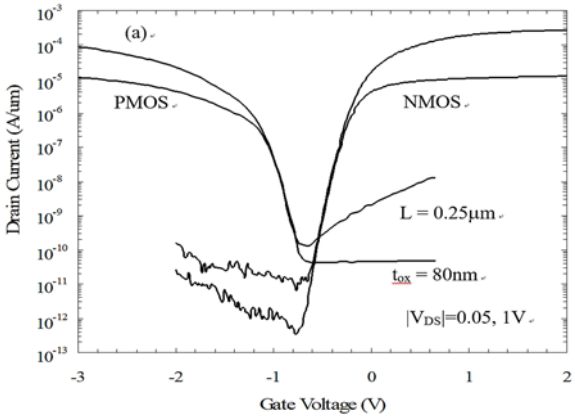


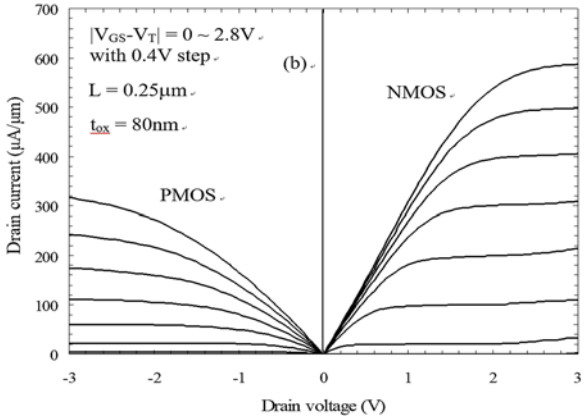
Fig. 2. (a)-(g) The key processing steps to form the stacked FinFET-CMOS inverter (h) The SEM image of the top view and cross-section of the fabricated FinFET-CMOS inverter.

The contact vias are filled to form the interconnect (Fig. 2(g)). The Scan Electron Microscopy (SEM) images of the cross-section and top view of the stacked FinCMOS inverter are shown in Fig. 2(h).

The I-V Characteristics of the upper and lower devices are shown in Fig. 3. The technology provides a number of advantages to implement 3D IC in highly scaled technology including: (1) high scalability inherent from the FinFET structure; (2) high density with more than 50 % area reduction compared to the conventional 2D architecture; (3) reduced interconnect wiring distance among active devices; and (4) process compatibility with traditional 2D CMOS technology, and (5) different device width between the top and bottom which can be implemented using different fin-heights that can be covered in the same floor plan. As shown in Fig. 3, the I-V characteristics of the fabricated device are not ideal due to a number of reasons. First of all, the current drives of the NMOSFET and PMOSFET in our fabricated devices are different due to the availability of the silicon film thickness available in the initial double SIMOX wafers used. On the other hand, the vias used to contact the source and drain regions of the bottom transistor are further away from the channel than for the top transistor. As a consequence, the series source and drain resistances are higher for the bottom than for the top transistor. These series resistance effects can also be found from the measured I-V characteristics as shown in Fig.3b. However, these problems can be adjusted using different fin-height and the optimization processing. In addition, the threshold voltage of the N and P devices are not symmetrical due to the use of n+ polysilicon for both NMOSFETs and PMOSFETs. This has to be corrected by using mid-gap metal gate in more advanced CMOS process.



(a)



(b)

Fig. 3. Measured (a) Ids-Vgs and (b) Ids-Vds characteristics from the stacked FinFET-CMOS transistors.

3. Processing Considerations in the Stacked FinFET-CMOS Technology

Despite the simplicity of the stacked FinFET-CMOS technology, a number of technology challenges exists and require careful optimization in the process design. First of all, process starts with a double layer SOI wafers which is not readily available in the market. Additional processing steps are needed to form the double SOI wafers either by wafer bonding or double SIMOX as described in Fig. 4. In our work, we start with a SIMOX wafer and a low energy oxygen implant was performed to separate the top silicon film into 2 layers with an additional buried oxide layer. The film thicknesses resulting from the process are given in Table 1. A TEM micrograph given in Fig. 4 shows the atomic structure and interface property of the double SIMOX wafers.

A particular challenge in the FinFET technology is the formation of a tall silicon fin with vertical sidewall and high aspect ratio.

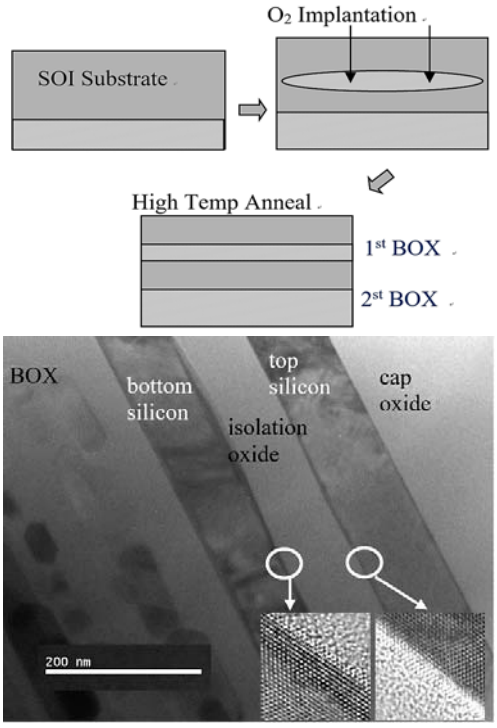


Fig. 4. Method to form the double silicon film and the TEM view of the material structure to form the double FinFET.

Table 1. Physical parameters of the double SIMOX wafers used in the formation of the stacked FinFET-CMOS technology.

Physical Parameter	Value
Top Silicon Thickness (nm)	109 + 2
Isolation Oxide Thickness (nm)	93 + 0.7
Bottom Silicon Thickness (nm)	81 + 0.8
BOX Thickness (nm)	321 + 2
Conductivity Type/Dopant	P/Boron
Orientation	(100)
Resistivity (Ohm.cm)	10-20

In the FinFET-CMOS process, this problem is magnified because the total fin height composed of the top silicon fin, bottom silicon fin and the isolation layer. By using inductive coupled plasma (ICP) etching, high aspect ratio fins with vertical sidewall angle of 89.5°-90° can be formed as shown in Fig. 5, but the sidewall quality has not been carefully studied.

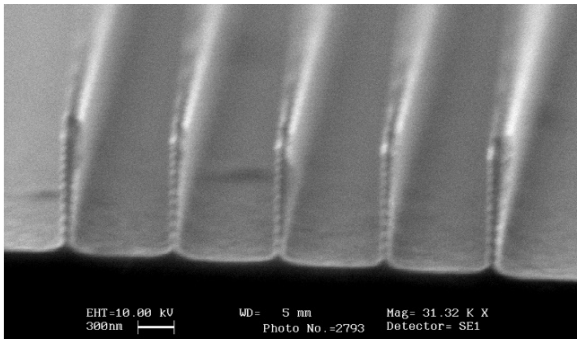


Fig. 5. SEM image of tall silicon film with high aspect ratio and vertical side wall (89.5°-90°) achieved by inductive coupled plasma (ICP) etching.

The FinFET-CMOS technology processes the top and bottom transistors together. It has the advantage of no accumulation of thermal cycle at the bottom layer when processing the top layer. However, doping the bottom with the existence of the top layer is very challenging. The most commonly used method is to apply a high energy implant to introduce dopant through the upper layers and target the implant peak at the location of the desired lower active layers as shown in Fig. 6. After the implant, dopant activation of all layers can be performed at the same time to reduce the thermal budget. However, the heavy implant through the upper layers can potentially create damage, and it is not easy to perform deep implants without a large lateral dopant spread. This limits the maximum implant depth and dictates the need to have only a small number of layers, minimal layer-to-layer separation and minimum layer thickness. As boron has deeper implant depth relative to phosphorus or arsenic, it is generally preferred to have PMOSFETs at the lower layers and NMOSFETs as in the demonstrated process.

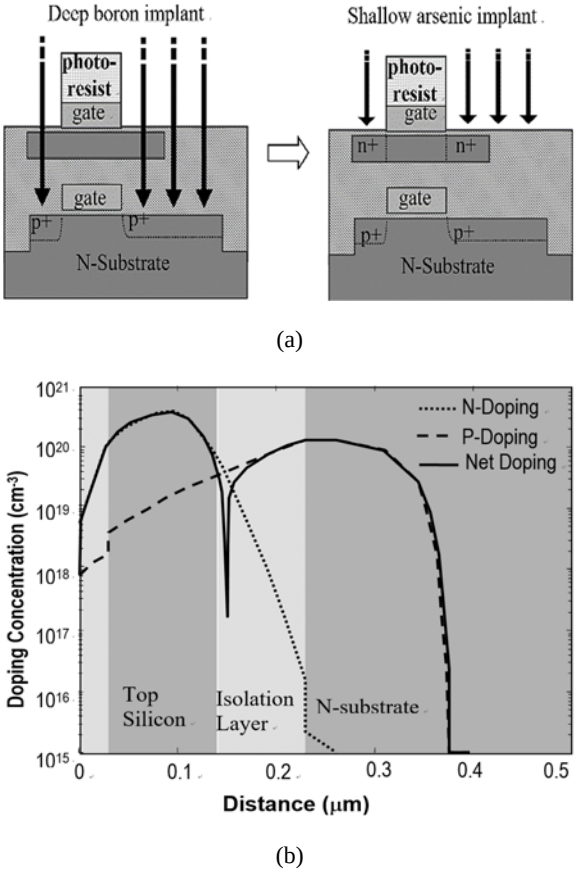
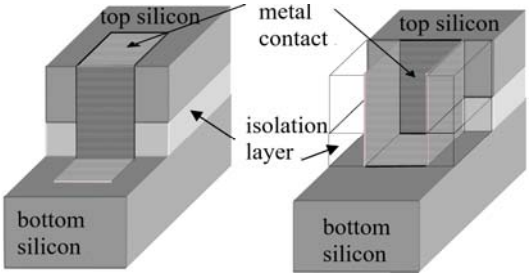


Fig. 6. (a). Introduction of dopants to the top and bottom active layers by implants with different energy; (b) The resulting doping profiles of boron and arsenic are demonstrated through numerical process simulation.

Contact between the top and bottom layers also requires special considerations. It can be achieved easily by opening contacts at different locations and use the metal wires above the device layers to connect them together. This approach consumes silicon area and reduces circuit density. Some better approaches to connect these layers are shown in Fig. 7. In Fig. 7 (a), two lithography steps are performed with the first one open up a deep contact through the top layer and stop at an etch-stop layer. The isolation layer may serve the purpose if it is sufficiently thin. In the second lithography step, a larger contact covering the deep contact is opened at the top silicon layer. Without removing the photo-resist, another etch is performed to remove the etch stop material covering the bottom silicon at the deep contact. The via filling then connects the top and bottom layers using the combined contact. Fig. 7 (b) illustrates another method to contact the top and bottom layers using a deep via formed by etching through the top layer directly down to the bottom layer. The via filling then connects the inside sidewall of the top silicon layer to the top surface of the bottom silicon layer at the same contact. The second method provides an effective way to contact both top and bottom layer without increasing contact dimension. But the contact resistance can be very high for very thin top layer silicon films.

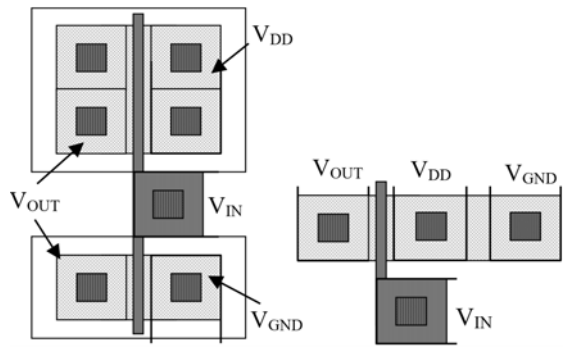


(a) (b)

Fig. 7. Illustration of a method to connect the top and bottom silicon layer through (a) combined contact to the top and bottom silicon layer; and (b) contact to connect the top and bottom silicon layer using the inside sidewall surface of the top silicon layer and the top surface of the bottom silicon layer.

4. Circuit Performance

The layout of an inverter formed using the stacked CMOS process is shown in Fig. 8. A large amount of area saving compared with conventional 2-D layout can be achieved by stacking the transistors. Besides CMOS inverter, other circuits such as 2-input NAND gate, 2-input NOR gate and 4-input NAND gate are also fabricated with layout as shown in Fig. 9 with area reduction around 50 %. To further evaluate the performance of the 3D stacked IC technology, we have extracted the device parameters from the fabricated device to obtain models for circuit simulation. Simulation results show that capacitive loading reduction from 40 % to 60 % can be achieved as shown in Fig. 10. The reduction comes from both more compact active devices placement and shorter interconnect distance between active devices within the cell. One detail performance estimation test and parameter extraction experiment from the measurement data indicated that the transistor compact active areas results in the 60 % reduction of the capacitance loading while the 3-D integration interconnection distance resulting in 40 % ratio of the capacitance loading reduction in the proposed stack FinFET technology.



(a) (b)

Fig. 8. Layout of (a) conventional 2-D inverter and (b) stacked CMOS inverter.

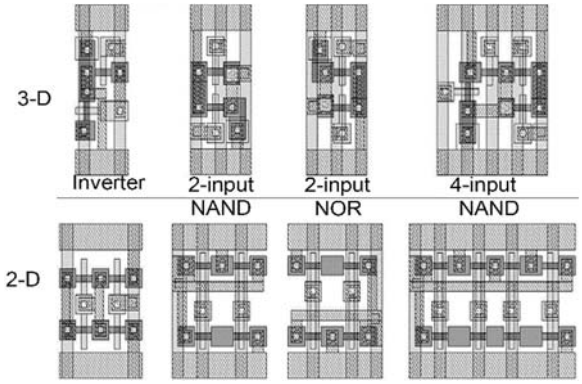
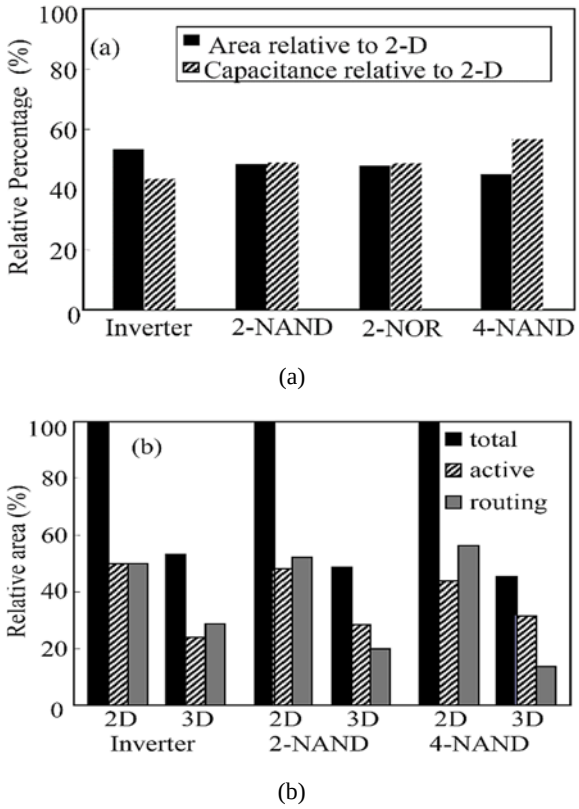


Fig. 9. Layouts of some of the 3-D FinFET-CMOS standard cells compared with the conventional 2-D layout.



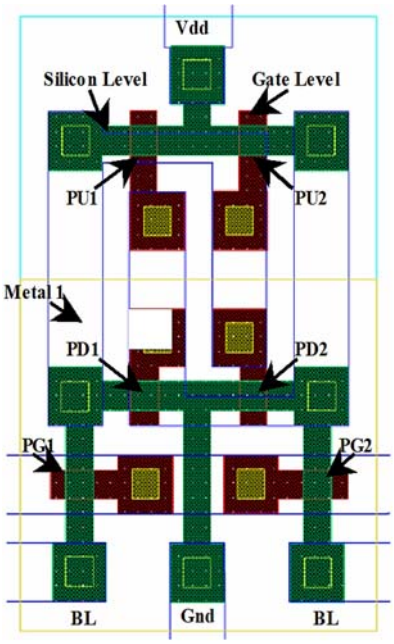
(a)

(b)

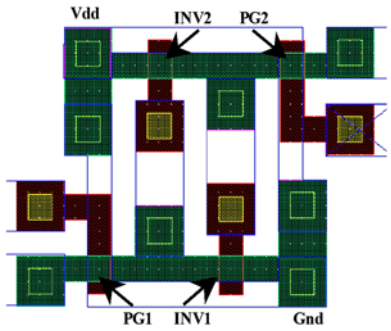
Fig. 10. Comparison of area and capacitance between 2-D planar gates and 3-D stacked FinFET-CMOS gates. (a) shows the percentage of area and capacitance of the gates constructed by the FinFET-CMOS structure relative to the 2-D case, and (b) shows the relative reduction of the area coming from the active devices and the intra-gate interconnects.

Among different circuits, memory array has the most regular structure with repeating units. Re-design the individual unit in a memory array with stacked CMOS technology is expected to produce the most significant area saving and performance improvement among all circuits. In particular, SRAM cells compose of mainly CMOS transistors and can quickly adopt the stacked CMOS technologies. Fig. 11 shows the layout of SRAM cell in 2-D [20] and using the stacked FinFET-CMOS technology, together with the output

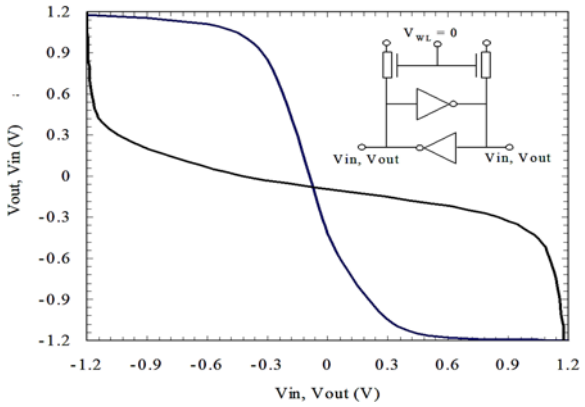
characteristics of the fabricated circuits. The advantage in terms of compactness is very obvious.



(a)



(b)

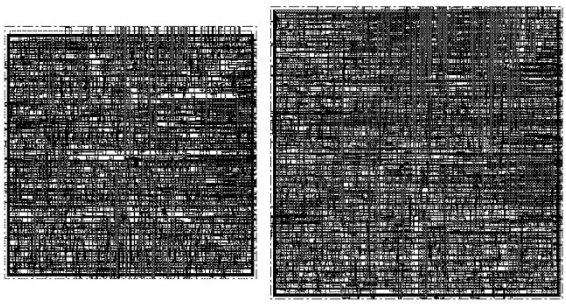


(c)

Fig. 11. Comparison of SRAM cell layout from with 2-D planar method (a) and the 3-D stacked FinFET-CMOS technology (b), indicating the high compactness achievable using the 3-D stacked FinFET-CMOS technology. (c) The inset shows the measured output characteristics of the SRAM cell constructed using the stacked FinFET-CMOS process.

The standard cell approach is very useful in developing 3-D integrated circuit because it is compatible with the current planar IC design methodology and CAD Tools. Once the standard cells have been reconfigured using the 3-D process, 3-D IC can be designed using high-level hardware description languages (e.g. VHDL and Verilog) and automatic placement and routing. This approach allows the benefit provided by 3-D IC to be quickly incorporated into the current technology.

To compare the performance of conventional planar and the 3D stacked-FinFET-CMOS structures, a number of circuits have been designed using high level Verilog language and doing auto placement and routing based on the new 3D standard cells, together with the conventional 2D planar standard cells. An example of 4-bit carry-look-ahead adder and 128-, 256-, 512-bit carry-select adder base on the 4-bit carry-look-ahead adder are studied. The layout of 128-bit adder is shown in Fig. 12 to indicate the area saving achieved by the 3D stacked-FinFET-CMOS technology. We found that the area of 3D circuits is about 70 %~75 % of the 2D circuits representing an area saving of over 25 %. However, if the circuit is small, the 3D circuit shows little improvement in the critical path delay as the RC delay of each stage is dominated by the intrinsic gate capacitance and output capacitance. When the circuit becomes more complex with increased transistor count, long interconnect with parasitic capacitance larger than that of the gate capacitance become dominant in the RC delay in the critical path. As shown in Fig. 13, the speed advantage of the proposed 3D circuit will become more and more prominent with increased number of transistor count.



(a) (b)

Fig. 12. (a) 3D 128- bit Adder obtained by automatic layout and placement; (b) 2D 128- bit Adder obtained by automatic layout and placement.

5. Conclusion

In summary, we have developed a stacked FinFET-CMOS technology and form circuit building blocks into multi-layer clusters. The technology has been verified and a number of circuit building blocks have been designed. Through circuit simulation based on physical parameters, the stacked FinFET-CMOS architectures shows better performance and lower

critical path delay when the number of transistors in the circuit increases. Despite the initial success of the demonstration, there are still a lot of technological challenges in the Stack FinFET-CMOS technology, including the formation of tall silicon fin, doping the bottom silicon layer with the presence of the top silicon layer, and using mid-gap gate material to form symmetric NMOSFET and PMOSFET transfer characteristics.

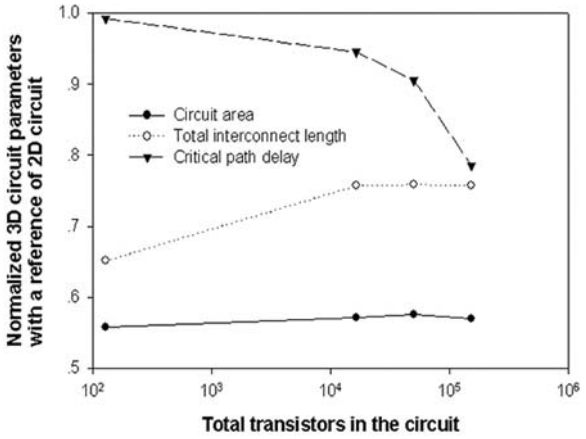


Fig. 13. Relative critical path delay, critical path length and circuit area of 2D and 3D circuits with different transistor count using the 2D circuit as a reference.

Acknowledgements

This work is funded by Guangdong Shenzhen Scientific and Technological Innovation Fund JCYJ20170817112527562, JCYJ20180227175343343, JCYJ20170412153845293, JCYJ20170817112848591, JCYJ20180507182843072.

References

[1]. J. D. Meindl, Beyond Moore's Law: The Interconnect Era, in *IEEE Computing in Science and Engineering*. Vol. 5, Issue 1, Jan/Feb 2003, pp. 20-24.

[2]. http://public.itrs.net

[3]. D. Lenoble, K. G. Anil, A. De Keersgieter, P. Eybens, N. Collaert, R. Rooyackers, S. Brus, P. Zimmerman, M. Goodwin, D. Vanhaeren, W. Vandervorst, S. Radovanov, L. Godet, C. Cardinaud, S. Biesemans, T. Skotnicki, M. Jurczak, Enhanced performance of pMOS MUGFET via integration of conformal plasma-doped source/drain extensions, in *Proceedings of the Symp. VLSI Technol*., 2006, pp. 168-169.

[4]. V. Subramanian, B. Parvais, J. Borremans, A. Mercha, D. Linten, P. Wambacq, J. Loo, M. Dehan, N.Collaert, S. Kubicek, R. J. P. Lander, J. C. Hooker, F. N. Cubaynes, S. Donnay, M. Jurczak, G. Groeseneken, W. Sansen, S. Decoutere, Device and circuit-level analog performance trade-offs: a comparative study of planar bulk FETs versus FinFETs, in *Proceedings of the IEDM Tech. Dig.*, 2005, pp. 898-901.

[5]. K. Bernstein, C. Chuang, R. Joshi, R. Puri, Design and CAD challenges in sub-90 nm CMOS Technologies, in *Proc. Int. Conf. Computer Aided Design*, Nov. 2003, pp. 129-136.

[6]. X. Huang, W. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. Cho, K. Asano, V. Subramanian, T. K. King, J. Bokor, C. Hu, Sub 50-nm FinFET:PMOS, in *Proceedings of the IEDM Tech. Dig*., 1999, pp. 67-70.

[7]. D. Hisamoto, W. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T. J. King, J. Bokor, C. Hu, FinFET-A self-aligned double-gate MOSFET scalable to 20nm, *IEEE Trans. Electron Devices*, Vol. 47, Issue 12, Dec. 2000, pp. 2320-2325.

[8]. M. V. Dunga, C. H. Lin, D. D. Lu, W. Xiong, C. R. Cleavelin, P. Patruno, J. Hwang, F. Yang, A. M. Niknejad, C. Hu, BSIM-MG: A versatile multi-gate FET model for mixed-signal design, in *Proceedings of the Symp. VLSI Technol.*, 2007, pp. 60-61.

[9]. E. J. Nowak, B. A. Rainey, D. M. Fried, J. Kedzierski, M. Ieong, W. Leipold, J. Wright, M. Breitwisch, A functional FinFET-DGCMOS SRAM cell, *IEDM Tech. Dig*., 2002, pp. 411-414.

[10]. T. Hsu, H. Lue, Y. King, J. Hsieh, E. Lai, K. Hsieh, R. Liu, C. Lu, A High-Performance Body-Tied FinFET Bandgap Engineered SONOS (BE-SONOS) for nand-Type Flash Memory, *IEEE Electron Device Lett.*, Vol. 28, Issue 5, May 2007, pp. 443-445.

[11]. K. Banerjee, S. J. Souri, P. Kapur, K. C. Saraswat, 3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-on-Chip Integration, in *Proc. IEEE*, Vol. 89, Issue 5, May 2001, pp. 602-633.

[12]. Lei Xue, C. C. Liu, H. S. Kim, S. Kim, S. Tiwari, Three-Dimensional Integration: Technology, Use, and Issues for Mixed-Signal Applications, *IEEE Trans. on Electron Device*, Vol. 50, Issue 3, Mar. 2003, pp. 601-608.

[13]. R. J. Gutmann, J.-Q. Lu, S. Pozder, Y. Kwon, D. Menke, A. Jindal, M. Celik, M. Rasco, J. J. McMahon, K. Yu, T. S. Cale, A Waferlevel 3-D IC Technology Platform, in *Proceedings of the Adv. Metallization Conf*., Oct. 2003, pp. 19–26.

[14]. K. Yamazaki, Y. Itoh, A. Wada, K. Morimoto, Y. Tomita, 4-layer 3-D IC Technologies for Parallel Signal Processing, *IEDM Tech. Dig*., 1990, pp. 599-602.

[15]. X. Lin, S. Zhang, X. Wu, M. Chan, Local Clustering 3-D Stacked CMOS Technology for Interconnect Loading Reduction, *IEEE Trans. on Electron Devices*, Vol. 53, Issue 6, June 2006, pp. 1405-1410.

[16]. X. Wu, P. C. H. Chan, S. Zhang, C. Feng, M. Chan, A Three-Dimensional Stacked Fin-CMOS Technology for High-Density ULSI Circuits, *IEEE Trans. on Electron Devices*, Vol. 52, Issue 9, September 2005, pp. 1998-2003.

[17]. X. Wu, P. C. H. Chan, S. Zhang, C. Feng, M. Chan, Stacked 3-D Fin-CMOS Technology, *IEEE Electron Device Letters*, Vol. 26, Issue 6, June 2005, pp. 416-418.

[18]. S. Zhang, R. Han, X. Lin, X. Wu, M. Chan, A Stacked CMOS Technology on SOI Substrate, *IEEE Electron Device Letters*, Vol. 25, Issue 9, September 2004, pp. 661-663.

[19]. M. Chan, Reduction of Interconnect Loading in Sub-100 nm Technology by 3D Stacked-FinCMOS, in

Proceedings of the IEEE Conference on Electron Devices and Solid-State Circuits, Taiwan, 20-22 December 2007, pp. 71-74.

[20]. Xiaomeng He, Jun Pan, Jinjing Liu, Cunlai Li, Guoqing Hu, Guangjin Ma, Jin He, and Mansun Chan,An effective mobility model for tunneling double-gate MOSFET (T-FinFET), Informatics, Electronics and Microsystems: *TechConnect Briefs*., 2018, pp. 169-172.

[21]. Jin He, Yuan Ren, Xiaomeng He, Xiaomeng Wang, Jun Pan, Jingjing Liu, Mansun Chan, Stacked FinFET-CMOS: A Promising FEOL Based Process Technology in Nanoscale 3-D Integration Circuits, in *Proceedings of the 2nd International Conference on Microelectronic Devices and Technologies (MicDAT' 2019)*, Amsterdam, The Netherlands, 22-24 May 2019, pp.19-21.