Advances, Challenges and Opportunities in 3D CMOS Sequential Integration

P. Batude[1], M. Vinet[1], B. Previtali[1], C. Tabone[1], C. Xu[1], J. Mazurier[1], O. Weber[1], F. Andrieu[1], L. Tosti[1], L.Brevard[1], B. Sklenard[1,2], P. Coudrain[2], S. Bobba[3], H. Ben Jamaa[1], P-E. Gaillardon[1], A. Pouydebasque[1], O. Thomas[1], C. Le Royer[1], J.-M. Hartmann[1], L. Sanchez[1], L. Baud[1], V. Carron[1], L. Clavelier[1], G. De Micheli[3], S. Deleonibus[1], O. Faynot[1] and T. Poiroux[1].

[1]CEA- leti, Minatec Campus, F-38054 Grenoble, France/ [2]ST Microelectronics, F-38926 Crolles, France / [3]EPFL, Lausanne, Switzerland.

Abstract- *3D sequential integration enables the full use of the third dimension thanks to its high alignment performance. In this paper, we address the major challenges of 3D sequential integration: in particular, the control of molecular bonding allows us to obtain pristine quality top active layer. With the help of Solid Phase Epitaxy, we can match the performance of top FET, processed at low temperature (600°C), with the bottom FET devices. Finally, the development of a stable salicide enables to retain bottom performance after top FET processing. Overcoming these major technological issues offers a wide range of applications.*

Introduction- The 3D sequential integration scheme offers the possibility to fully use the third dimension potential, i.e, to connect two stacked layers at the transistor scale whereas 3D parallel integration is limited to connecting blocks of a few thousand transistors (Fig.1).

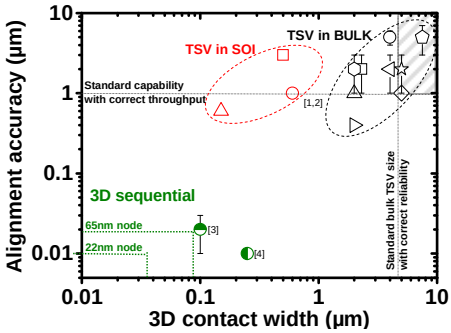


Fig.1: Alignment capability versus 3D contact width in parallel and sequential integration schemes. Reported also in the graph: Bulk TSV size and alignment capability limits with correct reliability & throughput respectively/ contact size and alignment capability for planar integration in 65 &22nm nodes.

However, its implementation faces the challenge of being able to process a high performance top transistor at Low Temperature (LT) in order to preserve the bottom FET from any degradation, as the stacked FETs are fabricated *sequentially* (Fig.2).

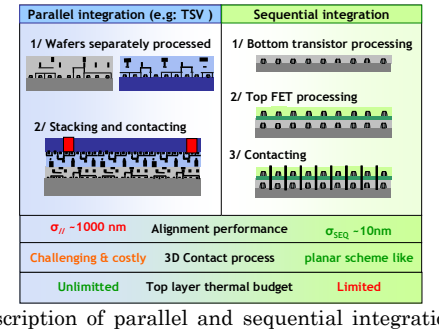


Fig.2: Description of parallel and sequential integration general process flows. TSV technology is one example of parallel integration. In this case, the stacked wafers are processed separately. In the sequential scheme, the transistors are processed sequentially above each other.

In this paper, the challenges of 3D sequential integration, (i.e. 1- stable performance bottom FET, 2- high quality top substrate fabrication, 3- top FET LT processing) will be presented as well as the proposed solutions to achieve such integration. Examples of potential applications are also reviewed.

Device fabrication- The process flow enabling to tackle the above mentioned challenges is presented in Fig.3. P- and N-FDSOI transistors with high-K/metal gate stack are fabricated on the bottom layer and standard high temperature spike anneal (1050°C) is used for dopant activation. Before bonding, thin Inter Layer Dielectric (ILD) is deposited and planarized on top of the patterned bottom transistors. LT (200°C) molecular bonding of SOI substrate enables full transfer of a monocrystalline Si layer. Top MOSFETs are then processed at low temperature (≤600°C). In particular, high temperature dopant activation is replaced by Solid Phase Epitaxy (SPE) at 600°C.

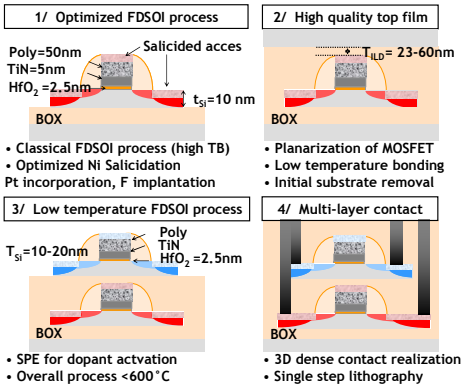


Fig. 3: Description of the process flow enabling to achieve stable performance bottom FET, high quality top substrate and high performance top FET with 600°C process

Bottom MOSFET performance- The first challenge in monolithic integration is to preserve bottom FETs performance during top FETs processing. To avoid additional dopant diffusion or interfacial oxide growth on bottom transistor, LT (<650°C) top FETs process is mandatory. Stabilized salicide is also required. Ni based salicide stability is obtained up to 650°C thanks to F &W implantation together with Pt incorporation (Fig.4-left). After complete 3D integration, R_s of the stabilized salicide shows no degradation (Fig.4-right).

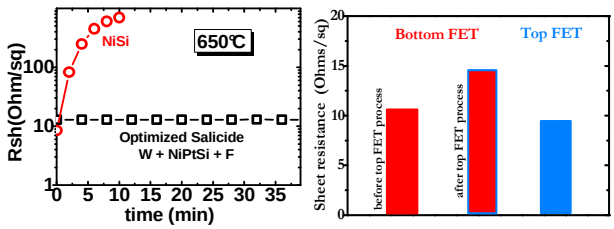


Fig.4-left: Development of an adapted salicide for 3D sequential integration: Adding Pt together with F &W implantation to the Ni based salicide enables to stabilize it up to 650°C. **Fig.4-right:** Sheet resistance of bottom salicided access (Ni+F) (before and after top FET processing @600°C) and of top salicided access.

Top active fabrication- Fig.5 presents the different techniques to obtain a crystalline semiconductor layer above processed transistors. Molecular bonding clearly stands apart from other techniques: first, it suppresses the need for Seed Windows (SW) required in recrystallization techniques [5-10] and thus allows higher integration density. In addition, bonding benefits from pristine crystalline quality and accurate thickness control.

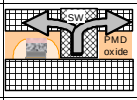
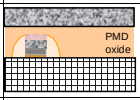
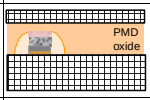
	Seed window (SW)	Poly-Si	Wafer bonding
Description			
Density	limited due to SW	Same than bottom level	Same than bottom level
Crystalline quality	Defect in SW region with controlled location	Random defects location	Perfect quality ~SOI supply quality
Thickness control	10s nm range	nm range	Å range
layer orientation	same orientation	random orientation for top substrate	different orientation possible
References	[5] [6] [7] [8] [9] [10]	[3] [11]	[12] [13] [14] [15]

Fig.5: Description, benchmark and main references for top active realization techniques.

Active layer transfers with semiconductor and interlayer dielectric thicknesses down to 10 and 25nm respectively have been demonstrated (Fig.6 (a) &12). Perfect bonding at the wafer scale is evidenced with acoustic and infrared characterization in Fig.6b&c.

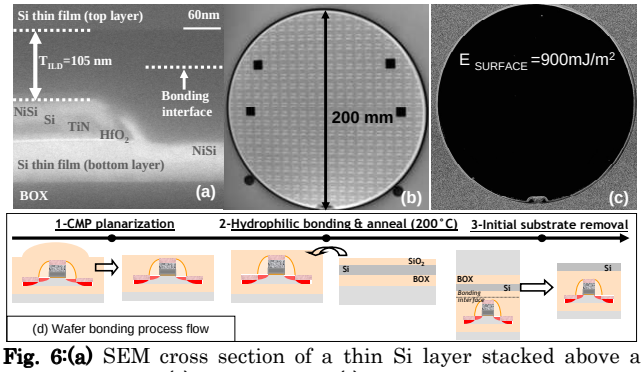


Fig. 6:(a) SEM cross section of a thin Si layer stacked above a transistor layer **(b)** Infrared and **(c)** acoustic characterization of bonded top active layer on bottom layer showing full transfer and no bonding defects for 200mm wafer **(d)** Wafer bonding flow.

Low temperature process (600°C) for top FET is achievable thanks to the use of SPE and high-K gate oxide. Indeed, as shown in Fig.7, SPE anneal at 600°C leads to similar I_{ON}/I_{OFF} trade-off than standard 1050°C spike anneal for both n&pFETs. These FDSOI transistors have been processed at CEA-leti.

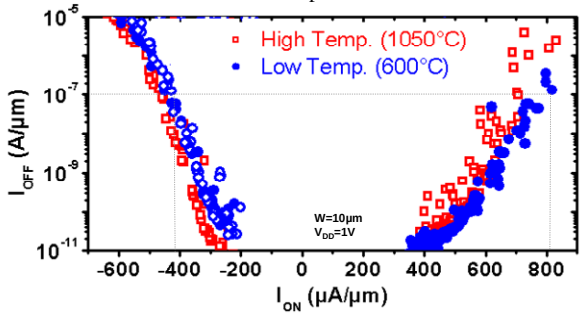


Fig.7: Comparison of I_{ON}-I_{OFF} trade-off for planar n- and p-FETS with low and high temperature dopant activation anneal.

The LT process does not increases GIDL leakage as shown in Fig.8. This is explained by the role of buried oxide on the end of range defects dissolution [16] which is enhanced in thin FDSOI devices (T_{Si}=6nm).

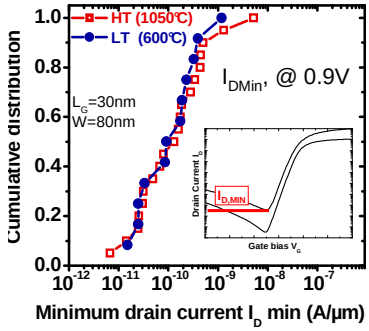


Fig.8: Similar junction quality for HT and LT planar devices is demonstrated by the plotting of I_{DMIN} cumulative distribution (V_D=0.9V).

Additionally, LT activation presents variability values in line with state of the art results for FDSOI devices (A_{VT}=1.35mV.µm) [17] as shown in Fig.9.

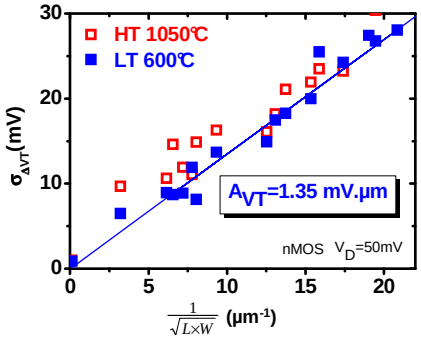


Fig.9: Pelgrom plot of planar LT & HT devices. LT A_{VT} value is in line with state of the art variability values on FDSOI [17]

Finally, LT process presents an improved gate leakage versus EOT trade-off than HT process (Fig.10-left). The EOT reduction is explained by a 4 Å thinner interfacial oxide (Fig.10-right).

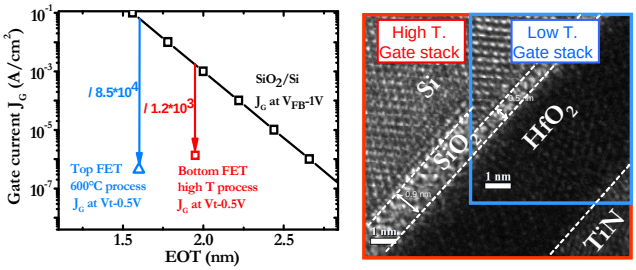


Fig.10-Left: Figure of merit of top (LT) and bottom (HT) oxides underlining the improved gate stack quality with LT process.
Fig.10-Right: TEM cross section of low temperature and standard high temperature process showing a reduction of interfacial SiO_2 oxide.

3D structures demonstration- Fig.11 benchmarks the technological options used in state-of-the art 3D sequential demonstrators. It highlights the interest of molecular bonding together with the 600°C process scheme, enabling the integration of bottom salicide.

	[14,15,18]	[3]	[6]	[12]
BottomFET				
Bottom salicidation	Stable Ni based salicide	Throughvia salicidation(CoSi)	throughvia salicidation (CoSi)	X
Top active realization				
Top active Technique	Molecular Bonding	Grain growth	Epitaxy (SEG)	Molecular bonding
Crystalline quality	=Original SOI	Poly crystalline	Defects in SW	=Original SOI
thickness control	=Original SOI	X	CMP planarizat°	=Original SOI
Top Fet & Top active Thermal Budget (TB)				
active TB	200°C	650°C	650°C	450°C
Gate Oxide	HfO_2 ALD (515°C)	Plasma SiO_2 (400°C)	Plasma SiO_2 (400°C)	Thermal SiO_2
Activation	SPE 600°C	Spike(>1000°C)	Spike(>1000°C)	RTP>1000°C
max TB	600°C	Spike(>1000°C)	Spike(>1000°C)	RTP>1000°C

Fig.11: 3D sequential technological options benchmark. The use of molecular bonding together with the 600°C LT process enables bottom standard salicidation.

Fig.12 displays a TEM cross section of a 3D sequential structure with two stacked FDSOI transistors with L_G=50nm. It corresponds to the smallest gate length demonstrated so far with a 3D sequential integration.

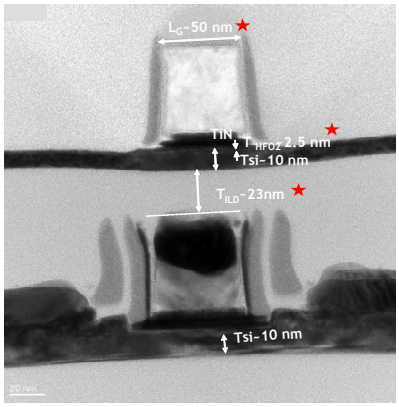


Fig. 12: TEM cross-sections of stacked transistors with record L_G=50nm and ultra thin interlayer dielectric T_{ILD}=23nm, T_{Si}=10nm.

Performance benchmark of the top pFETs with the state of the art is presented in Fig.13. For the same I_{OFF} of 100nA/µm, the 600°C top p-FET reaches comparable I_{ON} values (taking into account the smaller V_{DD}) than the top pFET of [3, 6] (processed at 650°C + spike anneal activation >1000°C). Note that using HT spike anneal for top dopant activation is detrimental for bottom FET performance (Ni salicide agglomeration and additional dopant diffusion).

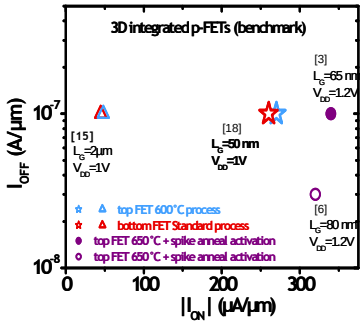


Fig. 13: Benchmark of top FET with 3D sequential integration literature with Lg<100nm.

Fig.14 presents the 3D inverter transfer voltage characteristics with such scaled gate length. Functional 3D 6T SRAMs cells have also been demonstrated, as shown in Fig.15.

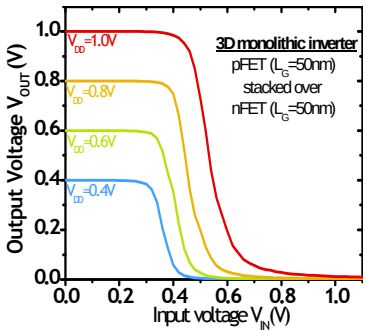


Fig. 14: Inverter transfer voltage characteristic with pFET stacked over nFET ($L_{G,P}$=$L_{G,N}$=50nm)

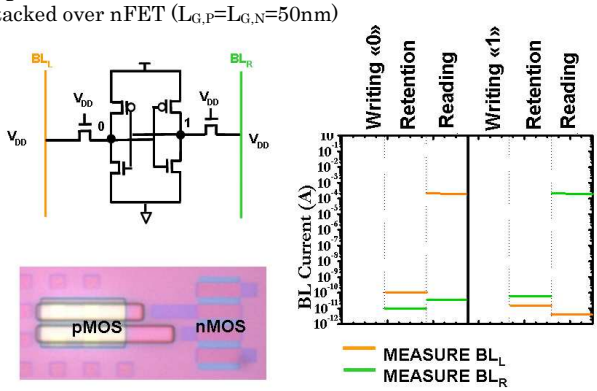


Fig. 15: Characterization of a 6T SRAM cell with pFETs stacked over nFETs. The BL current measurement evidence the SRAM cell functionality in the read, write and retention regimes

Application and perspectives- Sequential integration offers 3D contacts pitch close to planar contact pitch (Fig.1). This enables circuit partitioning at a fine granularity (i.e. at transistor/gate scale), which yields new potential applications. For example, such high density 3D contacts can be helpful for FPGAs, highly miniaturized imagers [19] and CMOS gates. Gain in performance is possible through the integration of the best suited technologies for different functions on distinct levels. Fig.16 summarizes the possibilities of co-integrations adapted to the different split functions.

	level	Example of partitioning	Best suited technology
Filed Programmable Gate Array (FPGA)	1	Pass gate	High Performance transistors
	2	6T SRAM	Low Standby Power transistors
Highly miniaturized pixels	1	Photodiode and transfer gate	1 µm thick SOI with back side illumination
	2	Readout transitors	Low noise transistors with relaxed (L,W) and gate oxide
CMOS gate	1	nFET	nFET gate stack, tensile-Si or InGaAs
	2	pFET	pFET gate stack, compressive-Si or Ge or (110) Si.

Fig.16: Examples of applications with gain when the partitioning is at the finest grain level. These applications are built from matrix of full custom cells, then their design are achieved without using 3D place and route tools. Gain in performance is possible through the possible integration of the best suited technologies for different functions on distinct levels.

Fig.17 shows an example of independent transistors optimisation in terms of gate stack, channel material (Ge/Si) and orientation (100/110).

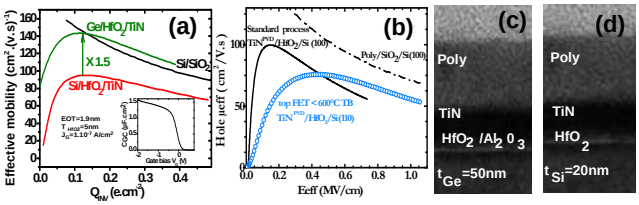


Fig.17: (a)p-Ge FET or (b) p-(110)Si-pFET stacked above (100)Si-n-FET. (c) adapted gate stack for Ge top FET (d) adapted gate stack for Si.

Such heterogeneous co-integration in a planar scheme would lead to complex and thus costly process. It is worth noting that the partitioning of the different functions described in Fig.16 can only be achieved in sequential integration thanks to its low 3D contacts pitch (in opposition to TSV technology with its 10µm pitch) as highlighted in Fig 18.

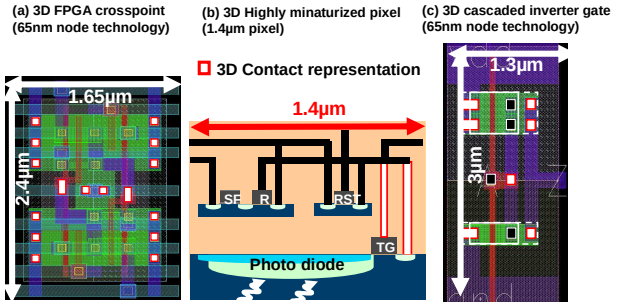


Fig.18: Layout or schematic of (a) 3D FPGA cross point in 65 nm node (b) Highly miniaturized pixel [19] (c) 3D cascaded inverter in 65 nm node. The partitioning described in Fig.16 is achievable thanks to the 3D contact at the transistor scale.

As these proposed applications are built from matrix of full custom cells, their design can be achieved without using 3D place and route tools. By extending this concept to complex digital ASIC, it is envisaged to stack logic cells in a 3D arrangement. The issue of 3D place and routing is settled thanks to a new 2D to 3D transformation technique [20], which is based on smaller standard cell stacking on top of bigger cells (Fig.19). This enables the use of standard 2D place and route algorithm.

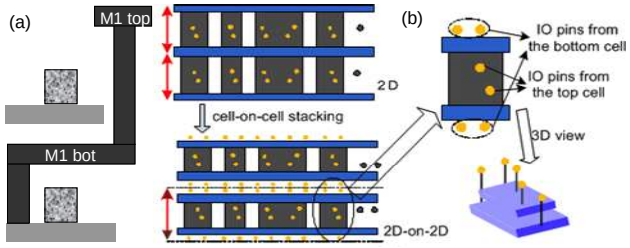


Fig. 19: Description of the principle of the 2D to 3D transformation enabling to use 2D standard place and route algorithm. This methodology is possible when integrating one metal level between the stacked transistors as described in (a).

Using this tool, a 15% reduction in the average interconnect length and a x1.8 improvement in overall power·delay·area product are predicted for the 45nm node (Fig. 20).

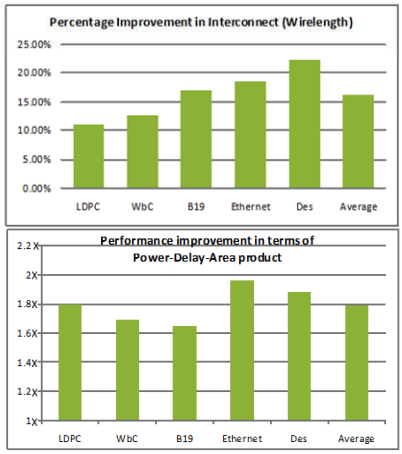


Fig. 20: Improvement in wirelength and power·delay·area product for different benchmark circuits [21] (45 nm node).

Finally, because of their regular and dense architectures, memories would largely benefit from 3D sequential integration. Indeed, it appears that, to continue to decrease bit cost, stacking will be more efficient than scaling [22].

Conclusion– Thanks to its ability to offer fine-grain circuit partitioning at the transistor scale, 3D sequential integration opens up a new field of applications and design. It enables both increasing the density and performances without resorting to aggressive scaling. Its key technological enablers are molecular bonding and low temperature top FET process which lead to design 3D transistors matching the targets of advanced nodes thanks to low access resistance, salicide, scaled EOT, optimized threshold voltage and mobility boosters.

Acknowledgments- This work has been carried out in the frame of the ST/IBM/CEA-Leti development alliance.

References:

[1]:P.Garrou *et al*., Handbook of 3D integration, vol 1 (Wiley Ed)
[2]: P.Garrou *et al.*, Handbook of 3D integration, vol 2 (Wiley Ed)
[3]: S-M.Jung *et al.*, VLSI 2005, pp220
[4]: P.Batude *et al.*, ECS journal 2008, VO16,pp47
[5]: Y-H. Son *et al*., VLSI 2007, p80
[6]: S-M. Jung *et al.*, VLSI 2007, pp82
[7]: J.Feng *et al*, Electron Device Letter 2006, Vol 27, Issue 11, pp911
[8]:H. Liu *et al*, IEDM 2001, pp729
[9]: T. Cheng *et al.*, IEDM 2009, pp179
[10]: T. Cheng *et al.*, IEDM 2010, pp496
[11]: T. Naito *et al*., VLSI 2010, pp219
[12]: L.Xue *et al.*, Trans. on Electron Dev, VOL. 50, NO. 3, pp601 (2003);
[13]: D-S. Yu *et al*., IEDM 2004, pp181
[14]: P. Batude *et al.*, VLSI 2009, pp166
[15]: P. Batude *et al.*, IEDM 2009; pp345
[16]: J-J. Hamilton *et al*., Master Sci Eng B, 2005;
[17]: O.Weber *et al.*, IEDM 2010, pp58
[18]: P. Batude *et al.*, VLSI 2011, pp158
[19]: P. Coudrain *et al.*, IEDM 2008, pp1
[20]: S. Bobba *et al.*; ASPDAC 2011, pp336
[21]: www.opencores.org
[22]: S-M. Jung *et al.*, IEDM 2006, pp1