Back-end Limitations in Advanced Nodes and Alternatives

A. Ayres[1,2,3], O. Rozeau[1], B. Borot[2], L. Fesquet[3], G. Cibrario[1], M.Vinet[1]
[1]CEA, LETI, Minatec campus, 17 rue des Martyrs, 38054 Grenoble, France
[2]STMicroelectronics, 850 Rue Jean Monnet, F-38926 Crolles, France
[3]Université Grenoble-Alpes / TIMA & CNRS / TIMA, Grenoble Institute of Technology, 46 Avenue Felix Viallet, 38031 Grenoble, France
e-mail: alexandre.ayres@st.com

***Abstract*—The back-end scaling is increasing the interconnections parasitic elements. In this paper, we analyze the performance of advanced nodes back-end considering a first order evaluation layout. Parasitic elements are extracted and then SPICE simulations were done using compact models for 14, 10, 7 and 5nm. Finally, we present the simulation of possible candidates to increase BEOL performance, such air-gaps and 3D sequential integration.**

Keywords—BEOL extraction simulation; BEOL scaling; back-end performance, 3DVLSI.

I. Introduction

Moore's scaling is roughly reducing transistor dimension by two each eighteen months. To keep the trend alive, new transistors architectures were introduced in the 28nm/22nm nodes; such as FD-SOI and FinFET [1], [2] shown in Fig.1. Those new transistors were designed to overcome limitations of the traditional bulk devices, which suffers of leakage when the gate length is scaled too aggressively. The FD-SOI and FinFET (FF) have excellent performance characteristics, allowing a reduction in the supply voltage, decreasing the power consumption and were adopted by the industry as solution for scaling [3]. In 5nm node, the stacked gate all around is expected to be introduced, increasing further the gate electrostatic control. The back-end of the line (BEOL) has to follow the transistor shrink, or in the other words, the interconnections have to be scaled in the same ratio of the transistor. The BEOL scaling trend is shown in Table I as minimum metal pitch. The metal pitch is composed by the metal line width plus the dielectric width separating two conductors. As the node name no longer represents a featured dimension in the transistor, Table I data is illustrated in Fig.2 showing the CPP x M1 pitch metric [4]–[10]. In this work, we simulate the BEOL for future nodes using a common setup layout, and then by parasitic extractions (PEX), we evaluate the BEOL impact in performance via SPICE simulations.

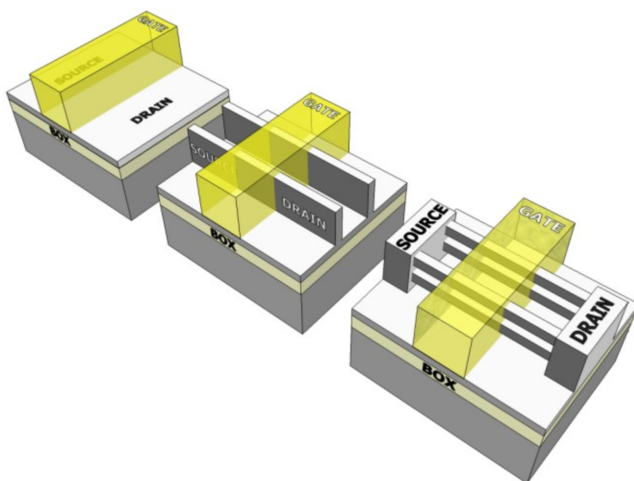

Fig. 1. Artist conception of transistor architectures for advanced nodes. From left to right: FD-SOI, SOI-FinFET and Stacked Gate All Around.

This work is partly funded by the French Public Authorities through NANO 2017 program and EQUIPEX FDSOI11, ST-LETI Alliance program and by Qualcomm.

TABLE I
SCALING IN ADVANCED NODES

Node	CPP [nm]	Metal Pitch [nm]	Supply Voltage [V]
14	90	64	0.8
10	64	48	0.7
7	46	36	0.64
5	32	24	0.6

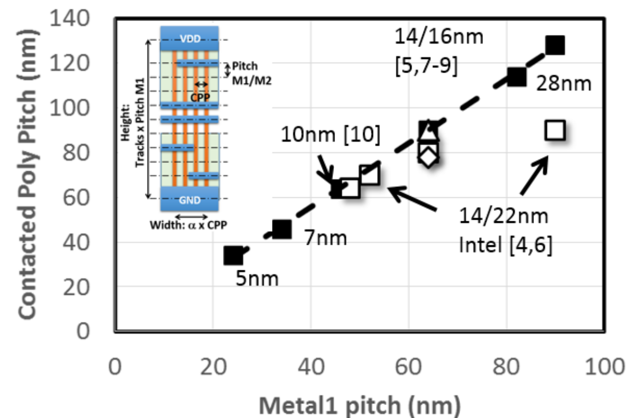

Fig. 2. Contacted Poly Pitch versus Metal 1 pitch.

II. SPICE Model Assesment

SPICE models have been developed to benchmark the technology nodes at circuit level. The 14nm model is based in the Leti-UTSOI2 [11], [12], which is already used in industrial environment. This model can be modified and used for benchmarking of FF and GAA. The GAA compact model is adapted from nanowire compact model for 7nm [13]. The models are calibrated using TCAD and experimental data. Some properties for the GAA model are deduced from previous published works, which includes wafer orientation and quantum effects. In this section, the technology

© 2017 The Japan Society of Applied Physics

performance is assessed using Ring Oscillators (ROs) in pre-layout, meaning that FEOL parasitic elements are included. The transistor performance for each technology is illustrated in Fig.3. The gate electrostatic control is greatly improved due to the architecture geometry. The aggressive FinFET scaling on the 7nm degrades the electrostatic control compared to the 10nm node due to fin width around 7nm.

	10nm		7nm	5nm
	FDSOI	FinFET	FinFET	GAA
DIBL (mV/V)	100	62	78	70
Slope (mV/dec)	84	72	79	71

Fig. 3. Transistor DIBL and SS compared for different architectures. The electrostatic control increases for advanced transistor architectures.

ROs performance are compared in Fig.4 for constant static power or constant dynamic power, using the chain delay versus the node CPP. A general trend of delay reduction of 35% per node is observed in both cases, the delay reduces with the scaling, due to increased transistor performance. Regarding the FEOL capabilities, the scaling should not face problems to deliver performances by introducing new transistor architectures.

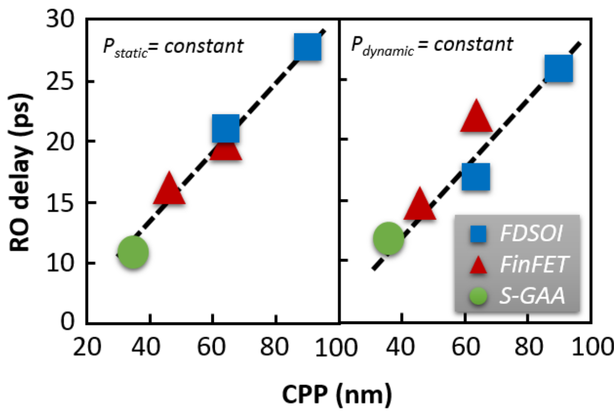

Fig. 4. Pre-layout RO delay versus nodes for all architectures. Symbols are RO simulations, dashed line is the trend of +35% on speed per node.

III. Back-end Scaling

The BEOL shrinking augments the connection resistance and capacitance for a given length, considering the same BEOL integration for different nodes. This problem is has a limited impact in the circuit performance, as the scaling reduces the transistor CPP, or the distance between gates. This translates in a short interconnection length; thus compensating, or even

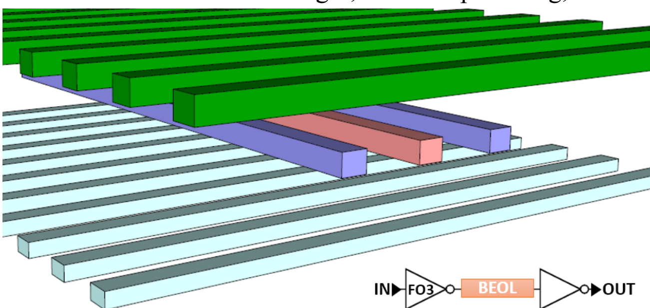

Fig. 5. Layout for delay benchmarking. M2 in green, M1 in dark blue and M0 in cyan. The connection under evaluation is highlighted in reddish.

increasing the interconnection performance despite the increased normalized parasitic elements per length. A layout has been designed in a PDK as follows in Fig.5, in order to simulate interconnection parasitic elements. The back-end connects an inverter logic gate in a FO3 configuration. Metal 0 is placed under the interconnection connected to supply source, as well M2 above connected to ground. Parallel to M1 signal, two lines are connected to supply and ground respectively. The connection length is normalized to node CPP, as hundred times the CPP for typical interconnections and ten thousands the CPP for hypothetical critical cases. In real circuits, the 10KCPP connections would be done in less dense metal and have buffers to not degrade the delay timing. This condition is used to illustrate a case where the BEOL delay is dominant, and does not depend on the transistor capabilities. The metal width and spacing follows the Table I for each node, along with the metal thickness adjusted x0.7 of previous node. After the layout parasitic extraction, the circuit was simulated in SPICE for several nodes and the back-end delay evaluated as in Fig.6.

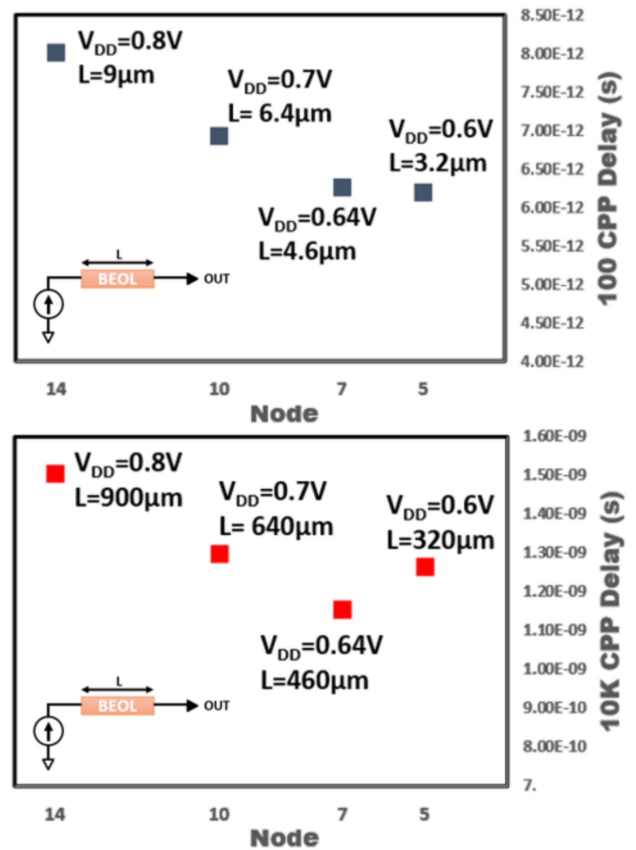

Fig. 6. Back-end delay for at fixed current for node given voltage. In the top, 6(a) evaluation for a connection of 100 CPP in length. In the bottom, 6(b) same illustration for ten thousand CPP in length.

In this simulation, all the cases use the same fixed current for the node given voltage, hence only the BEOL performance is evaluated. With the same BEOL composition (Cu/ULK), a trend reversal is observed in the 5nm node. Due to BEOL geometries, the interconnection length reduction is not anymore enough to compensate the parasitic elements. In Fig.7 the same netlist with the parasitic extraction is used, and

simulated using SPICE compact models for FD-SOI, FinFET and GAA. The total delay illustrated being composed by transistor delay and BEOL delay. For 100CPP the BEOL delay represents approximately 60% of the total delay. In Fig.7a, the delay reduction still occurs for the transition from 7nm to 5nm node. This outcome is explained by the 5nm transistor better current drive, compensating the degraded BEOL delay described in Fig.6. However, for very long wires, in the range of 10K CPP, the BEOL delay is dominant as seen in Fig.7b; and for 5nm, the performance is extremely impacted, confirming the results from Fig.6. As the BEOL performance does not scale from 7nm to 5nm, some solutions process solutions can be employed. An additional case for 5nm BEOL was extracted; employing air-gaps in the M1/M2, which are already implemented in upper metal levels of devices in mass production [4]. The air-gaps are implemented in the PEX files, simulating a relative permittivity (ε_r=2.2) for M1/M2 dielectrics. For 100CPP, the benefits are minimal, nonetheless a major improvement is seen for very long connections in 5nm, placing the node back in the general trend of Fig.7b.

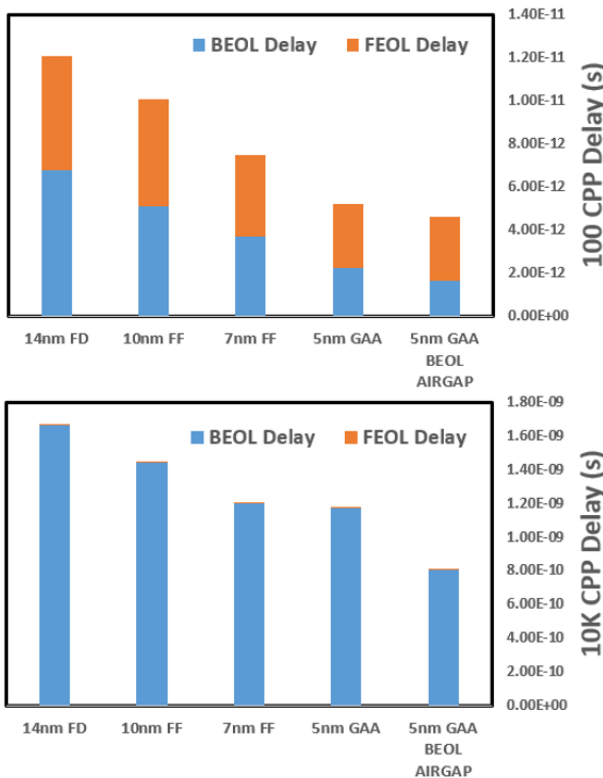

Fig. 7. Total delay considering previous layout setup and using SPICE compact model for each node. 7(a) evaluation for a connection of 100 CPP in length while 7(b) for ten thousand CPP in length.

The simulations done in Fig.7 consider the metal resistivity (ρ) of 4µΩ.cm for all nodes. Nonetheless, copper wire resistivity for minimum width increases in each node due to electron surface scattering and grain-boundary scattering. As the dimensions scale down, the metal widths are in the order of electron mean-free path, augmenting those effects and increasing the resistivity [7]. The setup was redone, at this time, considering the resistivity increase for advanced nodes.

The cooper resistivity is extracted from the evaluations of [15]. For example, the minimum metal width in 7nm node has a resistivity of 5.5µΩ.cm, while in the 5nm node 8µΩ.cm. The simulations are illustrated in Fig.8.

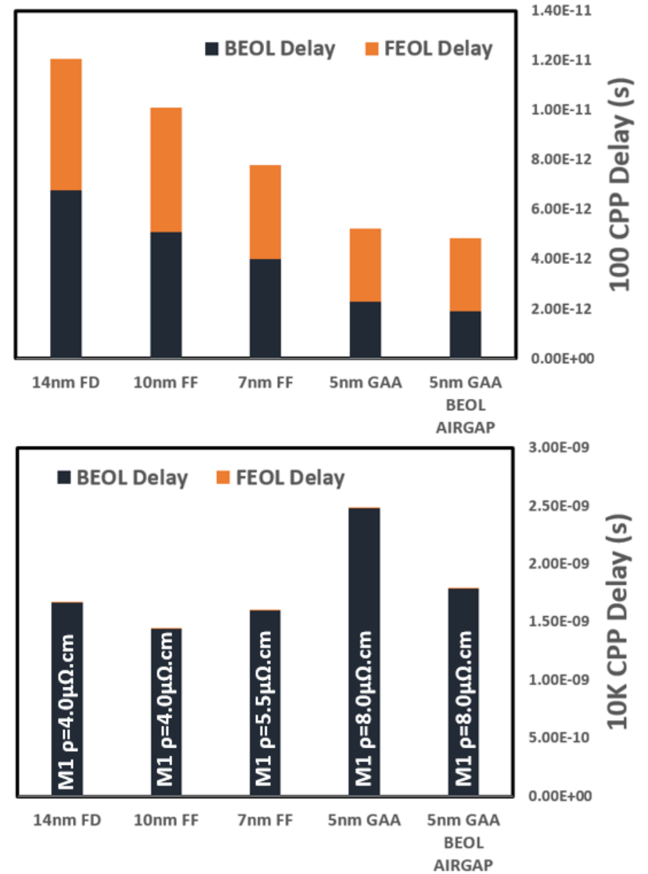

Fig. 8. Total delay considering previous layout setup and using SPICE compact model for each node and adjusted metal resistivity. 8(a) evaluation for a connection of 100 CPP in length while 8(b) for ten thousand CPP in length.

The general trend for previous simulation of 100CPP average interconnection length is still conserved. The increased resistivity in 7nm and 5nm nodes are compensated by the transistor performance. However, for the hypothetical case of very long wires, where the BEOL delay is dominant, the 5nm BEOL is performance impacting. This can result in more buffers used to transmit the signal through long wires or critical paths in real circuits, in order to avoid the BEOL parasitic elements.

IV. 3DVLSI AS BEOL SCALING ALTERNATIVE

Another possible way of BEOL scaling, is the use of 3D sequential integration. In this integration, the circuits are positioned in different levels, namely tiers. The CoolCube^{TM} has been developed with 3D VLSI logic circuits in mind [16], [17]. A presence of an intermediate back-end level (iBEOL) is necessary to increase the routability for dense circuits. The great advantage of this process is the high-alignment between tiers enabling 3D contact vias between tiers (3DCO) with a small size and pitch, as illustrated in Fig.9 for 14nm BEOL rules. As the 3DCO have similar size to M1 vias, the area

overhead is minimal compared to Through Silicon Vias (TSV) 3D parallel integration. The number of iBEOL layers in each tier is an input from designer, and mainly represents the

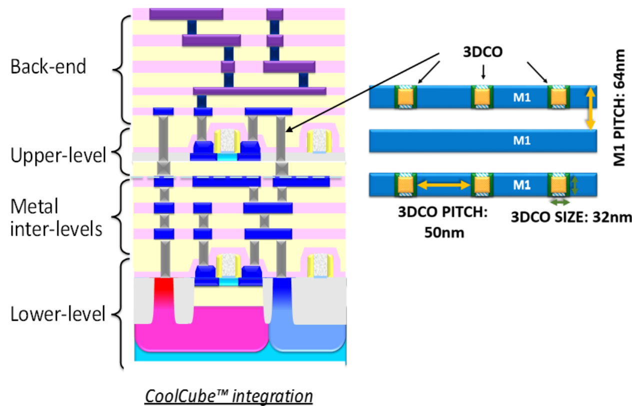

Fig. 9. On the left, 9(a) A schematic for a 3D sequential integration featuring two-tiers. On the left 9(b), the top view of 3D sequential vias (3DCO) with typical dimensions for 14nm design rules.

tradeoff between routability and the cost. The main goal of 3DVLSI is to allow an optimization at routing level, using the stacking of transistor to make them closer physically. This approach reduces the interconnections wirelength, thus the circuit performance can be increased. In Fig.10 a simulation compares the delay of 1K and 100 CPP in 5nm BEOL with

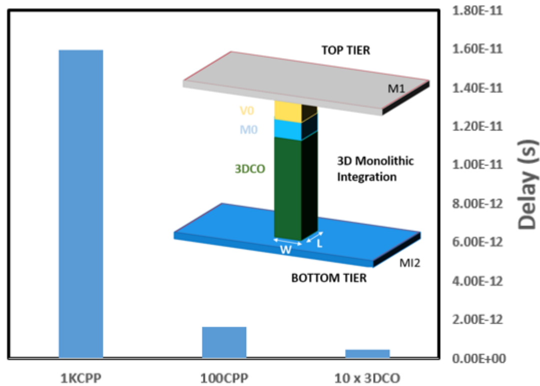

Fig. 10. Comparison of BEOL delay in 5nm node with air-gap. The 3D sequential integration 3D via (3DCO) is shown as an alternative to 2D BEOL scaling.

air-gaps to the delay of ten 3D contacts. The delay impact from 3DCOs connecting different tiers is negligible. Thus, the 3D sequential integration is ideal for dense logic circuits requiring routing through the tiers. Here, it is shown that if no technological solution is found for BEOL scaling in advanced nodes, such air-gaps in M1/M2, the 3D sequential integration can be a suitable candidate, as expanding to the Z direction can further reduce the BEOL delay.

V. Conclusion

In this paper, simulations using parasitic extracted layout are done for several advanced nodes. Then compact models are employed, and SPICE simulations revels a back-end limitation for the 5nm for very long wires. Finally, we show that if no technological solution is found to increase interconnect performance, 3D sequential integration is a key contender, avoiding BEOL limitations in advanced nodes.

References

[1] C. H. Jan *et al.*, "A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications," in *2012 International Electron Devices Meeting*, 2012, p. 3.1.1-3.1.4.

[2] F. Boeuf, M. Sellier, A. Farcy, and T. Skotnicki, "An Evaluation of the CMOS Technology Roadmap From the Point of View of Variability, Interconnects, and Power Dissipation," *IEEE Trans. Electron Devices*, vol. 55, no. 6, pp. 1433–1440, Jun. 2008.

[3] T. Skotnicki *et al.*, "Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 96–130, Jan. 2008.

[4] S. Natarajan *et al.*, "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 #x00B5;m2 SRAM cell size," in *2014 IEEE International Electron Devices Meeting*, 2014, p. 3.7.1-3.7.3.

[5] O. Weber *et al.*, "14nm FDSOI technology for high speed and energy efficient applications," in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1–2.

[6] C. Auth *et al.*, "A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012, pp. 131–132.

[7] C. H. Lin *et al.*, "High performance 14nm SOI FinFET CMOS technology with 0.0174 #x00B5;m2 embedded DRAM and 15 levels of Cu metallization," in *2014 IEEE International Electron Devices Meeting*, 2014, p. 3.8.1-3.8.3.

[8] T. Song *et al.*, "13.2 A 14nm FinFET 128Mb 6T SRAM with VMIN-enhancement techniques for low-power applications," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 232–233.

[9] S. Y. Wu *et al.*, "An enhanced 16nm CMOS technology featuring 2nd generation FinFET transistors and advanced Cu/low-k interconnect for low power and high performance applications," in *2014 IEEE International Electron Devices Meeting*, 2014, p. 3.1.1-3.1.4.

[10] K. I. Seo *et al.*, "A 10nm platform technology for low power and high performance application featuring FINFET devices with multi workfunction gate stack on bulk and SOI," in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1–2.

[11] T. Poiroux *et al.*, "Leti-UTSOI2.1: A Compact Model for UTBB-FDSOI Technologies #x2014;Part I: Interface Potentials Analytical Model," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2751–2759, Sep. 2015.

[12] T. Poiroux *et al.*, "Leti-UTSOI2.1: A Compact Model for UTBB-FDSOI Technologies #x2014;Part II: DC and AC Model Description," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2760–2768, Sep. 2015.

[13] O. Rozeau *et al.*, "NSP: Physical compact model for stacked-planar and vertical Gate-All-Around MOSFETs," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, p. 7.5.1-7.5.4.

[14] F. Chen and D. Gardner, "Influence of line dimensions on the resistance of Cu interconnections," *IEEE Electron Device Lett.*, vol. 19, no. 12, pp. 508–510, Dec. 1998.

[15] T. Huynh-Bao *et al.*, "Statistical Timing Analysis Considering Device and Interconnect Variability for BEOL Requirements in the 5-nm Node and Beyond," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. PP, no. 99, pp. 1–12, 2017.

[16] L. Brunet *et al.*, "First demonstration of a CMOS over CMOS 3D VLSI CoolCube #x2122; integration on 300mm wafers," in *2016 IEEE Symposium on VLSI Technology*, 2016, pp. 1–2.

[17] P. Batude *et al.*, "3DVLSI with CoolCube process: An alternative path to scaling," in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015, pp. T48–T49.