

3-D Sequential Integration: A Key Enabling Technology for Heterogeneous Co-Integration of New Function With CMOS

Perrine Batude, Thomas Ernst, Julien Arcamone, Gregory Arndt, Perceval Coudrain, and Pierre-Emmanuel Gaillardon

Abstract—3-D sequential integration stands out from other 3-D schemes as it enables the full use of the third dimension. Indeed, in this approach, 3-D contact density matches with the transistor scale. In this paper, we report on the main advances enabling the demonstration of functional and performant stacked CMOS-FETs; i.e., wafer bonding, low temperature processes (<650 °C) and salicide stabilization achievements. This integration scheme enables fine grain partitioning and thus a gain in performance versus cost ratio linked to separation of heterogeneous technologies on distinct levels. In this work, we will detail examples taking advantage of the unique 3-D contact pitch achieved with sequential 3-D.

Index Terms—Low temperature process, molecular bonding, solid phase epitaxy, three-dimensional (3-D) integration, 3-D monolithic integration, 3-D sequential integration.

I. INTRODUCTION

AMONG 3-D technologies, 3-D sequential integration offers the highest alignment capability and the smallest contact size. Indeed, as observed in Fig. 1, which shows the alignment accuracy between the stacked transistor layers with regards to the 3-D contact width, the parallel integration is far from reaching the contact density required at advanced nodes. Though some parallel architectures enable the reduction of the 3-D contact size (thanks to silicon on insulator (SOI) substrate for example), the 3-D contact pitch remains however limited by the alignment performance, which is limited to 1 μm [1], [2] to guarantee a correct throughput. On the other hand, in the case of sequential integration, it has been demonstrated that 3-D interconnections and planar contact pitches are matching even for advanced nodes. This means that 3-D sequential integration is the only possibility to fully exploit the potential of the third dimension, i.e., to connect two stacked layers at the transistor scale, 3-D parallel integration being indeed limited to coarse-grained connection of a few thousand transistors blocks (Fig. 2).

Manuscript received June 29, 2012; revised August 22, 2012; accepted September 11, 2012. Date of current version December 10, 2012. This paper was recommended by Guest Editor S. Mitra.

P. Batude, T. Ernst, J. Arcamone, and G. Arndt are with the CEA-leti, Minatéc Campus, F-38054 Grenoble, France (e-mail: perrine.batude@cea.fr).

P. Coudrain is with the ST Microelectronics, F-38926 Crolles, France.

P. Coudrain is with the École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland.

Digital Object Identifier 10.1109/JETCAS.2012.2223593

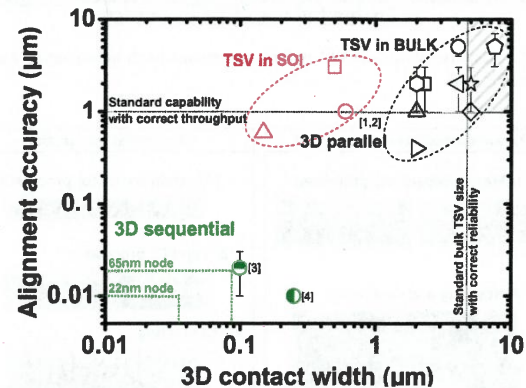


Fig. 1. Alignment capability versus 3-D contact width in parallel and sequential integration schemes.

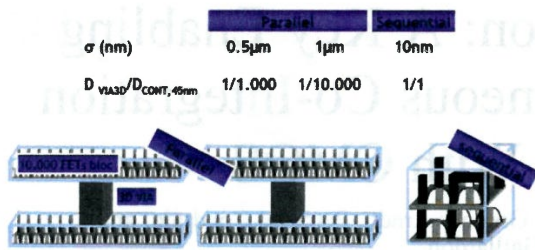
Numerous applications can take advantage of this promising 3-D integration technique. That allows a fine-grained partitioning, i.e., at gate or transistor levels, that cannot be obtained with the large 3-D contact pitch of parallel integration. The high alignment performance of 3-D sequential is due to the fact that, instead of being processed separately, each layer is processed sequentially after having been bounded on the top of the previously processed one. Therefore, the alignment performances for the top devices are in line with the stepper capabilities, which typically present a standard deviation close to 10 nm for 22 nm steppers. These high alignment performances are illustrated by [3], [4] in Fig. 1, which are already compatible with advanced planar contact pitch.

However, 3-D sequential implementation faces the challenge of being able to process a high performance top transistor at low temperature (LT) in order to prevent any deterioration of the previously realized bottom field-effect transistor (FET) (Fig. 3).

The first part of this paper, dedicated to the process technology, presents the challenges of 3-D sequential integration with the solutions proposed to solve them. Then potential applications, taking full advantage of this new technological approach, will be detailed in the second part.

II. 3-D SEQUENTIAL TECHNOLOGY

When setting a sequential integration technology line, scientists are facing several problems that have to be solved. The



Sequential approach: Density_{3D contact} ~ Density_{planar contact}
→ Enable full use of 3rd dimension

Fig. 2. Ratio between 3-D VIA and planar contact pitch in sequential and parallel case.

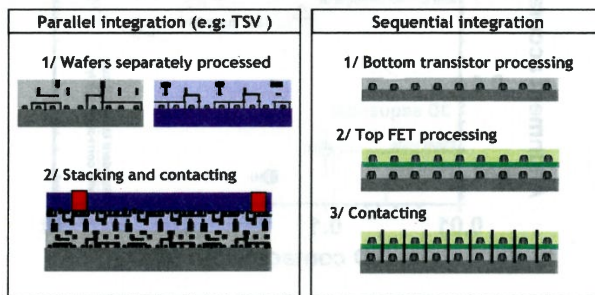


Fig. 3. Parallel and sequential integration general process flows.

main problems are presented in Fig. 4: the preservation of the bottom FET during the top FET processing; the top film realization; and the fabrication of top transistor at LT.

A. Bottom MOSFETs Performance

The first challenge in sequential integration is to preserve the bottom FETs performance during the top FETs processing. Bottom transistors are processed with standard high temperature (HT) budget ($\approx 1050^\circ\text{C}$). However, to avoid additional dopant diffusion or interfacial oxide growth on bottom transistor, LT ($<650^\circ\text{C}$) top FETs process is mandatory. Nevertheless, at these temperatures, standard NiSi salicide is still subject to dewetting. Therefore, the first challenge that had to be solved was the development of a stabilized salicide. The incorporation of platinum together with fluor and tungsten implantation enabled to reach stable sheet resistance values up to a 650°C budget [4].

B. Top Active Fabrication

Fig. 5 surveys the different techniques developed so far to obtain a crystalline semiconductor layer above processed transistors. Among them, molecular bonding clearly stands apart from other techniques thanks to two major advantages. First, it does not require a seed windows (SW) required in recrystallization techniques [5]–[10], allowing hence a higher integration density. Second, the bonding operation benefits from pristine crystalline quality and accurate thickness control.

The Fig. 6(a) illustrates the bonding process flow. Before the bonding steps, a thin interlayer dielectric (ILD) is deposited and

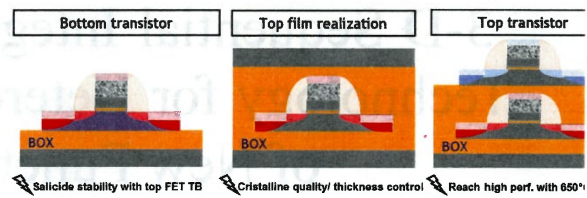


Fig. 4. Description of the 3-D sequential process flow with the emphasis put on the main challenges.

	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.

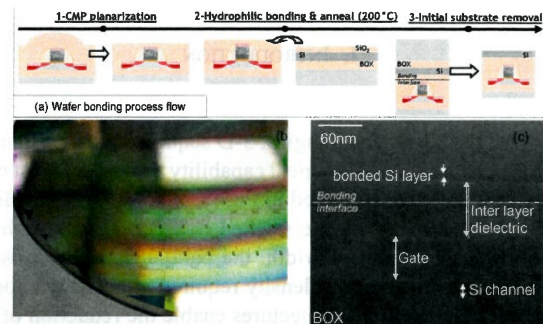


Fig. 6. (a) Wafer bonding flow. (b) Top view of a bonding above a bottom transistor layer showing no bonding defects. (c) SEM cross section of a thin Si film reported on bottom transistor with wafer bonding.

planarized on top of the previously patterned bottom transistors. LT molecular bonding (200°C) of SOI substrate enables the full transfer of a monocrystalline Si layer [Fig. 6(b)]. Si layers down to 10 nm [Fig. 6(b)] and interlayer dielectrics down to 23 nm [16] have been reported.

C. Low Temperature Process (650°C) for Top MOSFET

Top MOSFETs are processed with an overall thermal budget of 650°C . Since the classical high temperature thermal dopant activation ($\approx 1050^\circ\text{C}$) is replaced by a solid phase epitaxy (SPE) operated at 600°C , which allows activating the dopants during the recrystallization of the preamorphized Silicon region. As shown in Fig. 7, the devices activated at LT match the performance of HT activated devices in both n- and p-FETs cases.

Junction leakage increase, usually observed for thick channels devices, does not occur for devices with thin channel (<10 nm). This has been explained by the junction close proximity

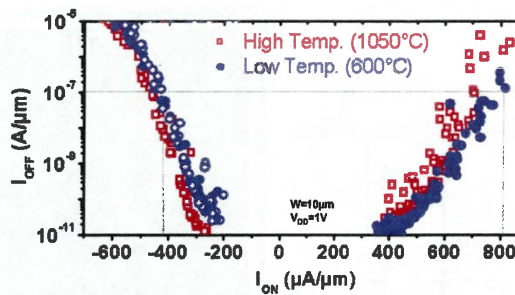


Fig. 7. Comparison of I_{ON} - I_{OFF} trade-off for planar n- and p-FETs with low and high temperature dopant activation anneal.

with the buried oxide [17]. Moreover, LT process also presents intrinsic advantages with respect to HT process. First a better trade-off between gate leakage and *equivalent oxide thickness* can be achieved [15], [18], then a broader range of material options for an effective tuning of the work function in a gate-first process scheme can be used [18].

The *solid phase epitaxy* activation is carried-out with a thermal budget around 500 °C–600 °C. Though this thermal budget permits building transistors on top of each other, it is still too high for being used above copper metal lines since the maximum temperature supported by the back end is actually in the 400 °C range. In order to reduce the thermal budget below this limit, novel device architectures such as surrounding gate transistor (SGT) [19], [20] or recessed channel transistors (RCAT) [21] have been proposed. In these devices, the PN junctions are vertical and obtained by the transfer of a pre-doped substrate. In the context of standard high performance MOSFET, i.e., with an horizontal channel, other techniques such as laser [22] or microwave annealing [23] are under investigation for the dopant activation step.

D. 3-D Structures Demonstration

Fig. 8 displays a TEM cross section of a 3-D sequential structure with two stacked FDSOI transistors with $L_G = 50$ nm. To our knowledge, it is the smallest gate length demonstrated using a 3-D sequential integration approach. Fig. 9 presents the transfer voltage characteristics of a 3-D inverter, i.e., with a bottom n-type transistor and a top p-type transistor.

The technological achievements presented in this section open the path to the realization of multiple stacks of transistors with matched performances and a large density of 3-D contacts. From these results other co-integration schemes can be extrapolated provided that the top devices are fabricated with a 650 °C process. Different applicative examples of co-integration with CMOS will be detailed in the following section, in order to illustrate the opportunities of the technology.

III. APPLICATION AND PERSPECTIVES

Sequential integration offers a 3-D contact pitch close to its planar counterpart (Fig. 1). This enables novel circuit partitioning at several granularities, as depicted in Fig. 10.

Fine-grained partitioning, i.e., at transistor/gate scale, leverages to performance gain that only 3-D sequential integration can afford.

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

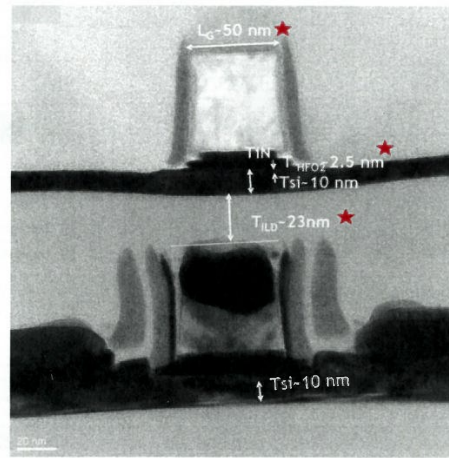


Fig. 8. TEM cross-sections of stacked transistors with record $L_G = 50$ nm and ultra thin interlayer dielectric $T_{ILD} = 23$ nm, $T_{Si} = 10$ nm.

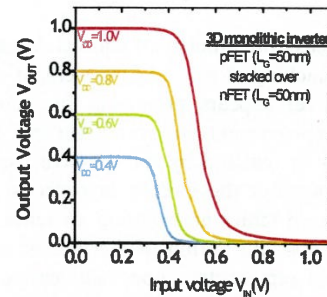


Fig. 9. Inverter transfer voltage characteristic with pFET stacked over nFET ($L_{G,P} = L_{G,N} = 50$ nm).

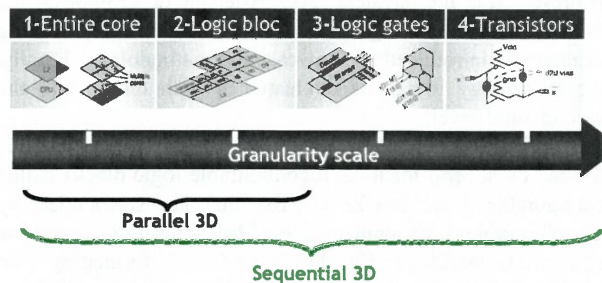


Fig. 10. Coarse grain to fine grain circuit partitioning for 3-D circuits as described in [24]. 1: Memory/core on core. 2: Function unit block. 3: Logic gate distributed across different layers. 4: Transistor scale partitioning.

A performance increase is possible thanks to the closer proximity between functions as well as the opportunity to tune separately the different layers technological process to fit them better with dedicated functions. For example, high-density 3-D contacts are helpful for applications as different as FPGAs, sensor amplification systems, highly miniaturized imagers.

A. Field Programmable Gate Arrays

Field-programmable gate arrays (FPGAs) are currently the most advanced generic and programmable fabric suited for digital applications. While 3-D parallel integration allows to

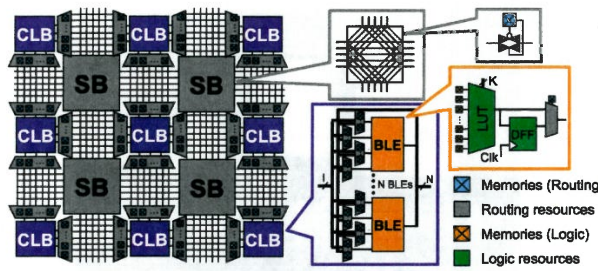


Fig. 11. Baseline FPGA architecture [26].

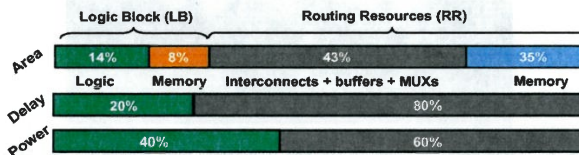


Fig. 12. FPGAs area/delay/power repartition per resources [25].

drastically increase the number of equivalent gates of FPGAs (thanks to the stacking of several FPGA chips), the integration at fine-grain level appears promising for compactness and performance improvements of reconfigurable circuits. Their large versatility is realized by the massive use of memories distributed throughout the circuit, in order to configure the embedded look-up tables and routing switches. Thus, FPGA circuits are a good illustration of the interest of a gate level partitioning, and represents a good applicative field for 3-D sequential technology [25]. In [25], the performance benefits of a sequentially stacked FPGA are surveyed analytically at architectural level. In this section, the design of a 3-D FPGA is presented at the physical level. After the description of the FPGA architectural organization, the design of 3-D routing elements is introduced to showcase the methodology. Finally, the technology impact is evaluated at both circuit and the architectural level.

1) *The FPGA Architecture:* FPGAs are regular circuits formed by several identical reconfigurable logic blocks called configurable logic blocks (CLBs) that are surrounded by reconfigurable interconnect lines [26]. The FPGA generic structure is depicted in Fig. 11. Every CLB is formed by a set of N basic logic elements (BLEs). A BLE is simply a K-input look-up-table (LUT), whose output can be routed to any other LUT input with or without being saved in a flip-flop. Every CLB has I inputs coming from other CLB outputs. All design parameters N, K, and I can be set by the FPGA architect depending on the targeted system granularity.

Programmable interconnections between the different blocks are realized by a massive number of *SwitchBoxes* (SBs) configured by memories cells. Fig. 12 presents the area/delay/power breakdown of the various components of a baseline SRAM-based island-style FPGA. It is noteworthy that the memories and the routing resources play a major role in FPGA performances, contributing to more than 80% of the total area and delay. For this reason, we propose an improvement of the programmable routing resources using 3-D sequential integration, as an interesting example.

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

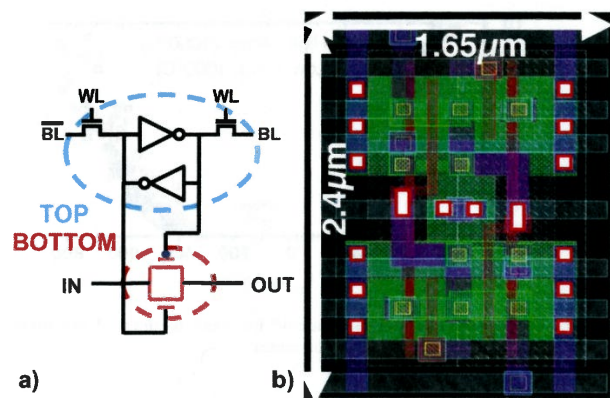


Fig. 13. a) 3-D crosspoint circuit splitting and b) layout (3-D connections are highlighted in red).

2) *3-D Sequential FPGA Cross-Points:* In this section, the design of a sequentially stacked cross-point is presented. Cross-points are the most common blocks found across the structure. Using the unique density of 3-D interconnections offered by the integration process, the configuration memories are realized on top of the logic elements.

Fig. 13(a) illustrates how 3-D sequential could lead to compact FPGA cross point. In the proposed circuit, a 6-transistors memory (6-T SRAM) is stacked over a pass-gate. Such a repartition allows a clear separation between the circuits involved in the configuration storage and the circuits related to the data path. Technological optimization might be further applied. Especially, it is envisaged to have a low-leakage/low-power process for memories, while the active data paths are created using high performance transistors.

The layout associated with the structure is depicted in Fig. 13(b). 3-D connections are highlighted in the figure. Note that the density of interconnection required by the application can only be reached by sequential integration. In this implementation, the memory dominates the dimensions of the cell. Indeed, the pass-gate has low requirements in terms of metal interconnections. It is worth pointing out that the pass-gate might thus be oversized while keeping the cross-point size equal to the SRAM. Large pass-gates demonstrate good electrical performances and are of high interest for routing applications. Note that, as technology shrinks, the memory and the pass-gate will not scale at the same rate. Indeed, for performance purpose, the pass-gate will need to be oversized. Then, even if 3-D sequential integration enable natively the oversizing at large lithographic nodes, this advantage may be less pronounced at hyper scaled nodes, because of the limiting factor will shift between the memory and the pass-gate footprint.

3) *Performance Evaluation:* To showcase the unique gain of the 3-D sequential integration, we characterized the impact of the designed cross-point at both circuit and architectural level.

a) *Cross-Point Electrical Performances:* Performances of the cross point have been first extracted by electrical simulations. The simulations are run using a homemade compact model with HSPICE simulator. To allow a fair comparison with planar technologies, the same design is implemented in

TABLE I
GATE-LEVEL CROSS-POINT PERFORMANCE

65-nm node	Area (μm^2)	Intrinsic delay (ps)	Kload (ps.fF ⁻¹)	Average power at 2 GHz (μW)
2-D	15.83	0.44	0.71	0.12
3-D	3.96	0.14	0.66	0.058
2-D vs. 3-D	x 4	x 3.1	x 1.1	x 2.1

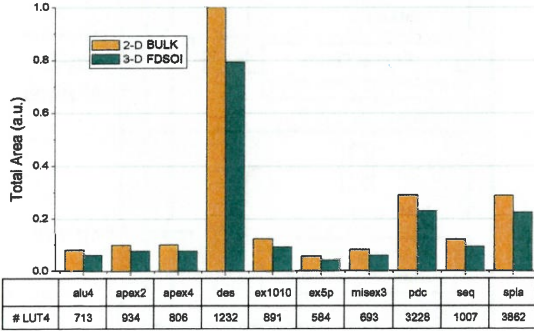


Fig. 14. Area estimation for FPGAs synthesized with standard bulk circuits and sequential 3-D integrated FDSOI circuits.

2-D CMOS. The comparison metrics are the area, the delay (extracted at a fan-out-of-4) and the average power (extracted at 2 GHz). Experimental results are presented in Table I. At the gate level, it has been observed a gain of 2.9x, 3.1x, and 2.1x, respectively, in area, delay and power consumption. While good numbers in terms of area are the direct impact of stacking, it is worth noticing that good electrical performances might be reached thanks to the novel requirements in sizing. Indeed, in the proposed illustration, the 6T-SRAM is driving the area occupancy, then the pass-gate transistors might be oversized without any prejudicial impact on footprint area.

b) *Architecture Level Performances*: Moving towards the architectural level, the expected gain is evaluated by FPGA benchmarking. A set of logic circuits taken from the MCNC benchmark was first synthesized using the ABC tool [27]. The technology mapping is then performed with a library of 4-input LUTs ($K = 4$) using ABC as well. Subsequently, the logic packing of the mapped circuit into CLBs is done with $N = 10$ BLEs per CLB and $I = 22$ external inputs using T-VPACK [28]. Note that, while MCNC benchmark suite tends to use small designs, their diversity allows evaluating the scaling trend over a large set of application cases. Finally, the placement and routing were carried out using VPR [28]. The considered benchmarks were synthesized twice: first on a planar SRAM-based FPGA in the CMOS 65 nm process and then on its 3-D sequential counterpart. The area estimation is normalized and shown in Fig. 14.

The benchmarks show an area reduction ranging from 20% to 25%, with 21% on average. The main benefit of using the 3-D integration technology is the smaller impact on memory circuits on the area. Memories circuits are placed on top of the FPGA structure, while all the data path circuits are placed below. Nevertheless, since the top transistors are directly stacked above

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

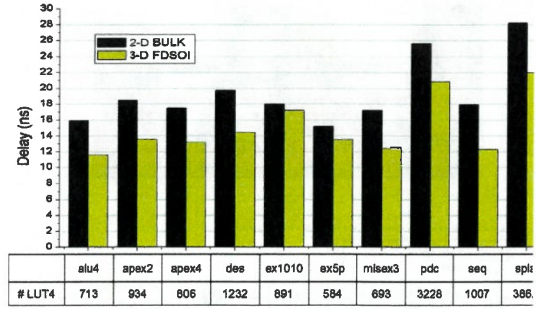


Fig. 15. Delay estimation for FPGAs synthesized with standard bulk circuits and sequential 3-D integrated FDSOI circuits.

the bottom silicon layer, the circuit suffers some loss of performance in terms of routing. Indeed, the top transistors are impeding the connections between bottom transistors and the first metal layer. To improve the quality of routing, the use of internal metal layers is envisaged. The delay estimation with respect to the benchmark circuits is shown in Fig. 15.

We can observe a delay reduction ranging from 10% to 45%, with 22% on average. The main benefits of using the 3-D FDSOI instead of standard planar technologies are twofold. First, the intrinsic performance of FDSOI compared to bulk contributes to speed up the gate delay and thus reduce the computation time. Second, the reduction of routing area, as well as logic area, leads naturally to a reduction of routing wires. Shorter routing wires will be driven more efficiently, besides the fact that they will be controlled by drivers with improved output conductance. This makes the 3-D integrated FPGA potentially faster than its standard Bulk counterparts.

B. 3-D Sequential Integration of Sensors

Many types of silicon-based miniaturized sensors or actuators would presumably benefit from a 3-D co-integration with their CMOS readout and conditioning electronics, as they would gain in signal-to-noise ratio (SNR), immunity to external parasitics, density, compactness, and possibly in cost. In particular, a 3-D sequential integration of nano-electro-mechanical resonators and systems (NEMS) with CMOS 1) would make them drastically gain in SNR, and 2) would enable implementing an unprecedented density of hybrid NEMS-CMOS pixels density (not achievable by NEMS-CMOS TSV-based integration), required for novel NEMS-based applications, such as mass spectrometry or hybrid NEMS-CMOS logic circuits.

Up to now, NEMS have been mostly envisaged as ultra-sensitive mass sensors since they have demonstrated extremely low mass detection levels [29], thanks to their intrinsic low mass, their high resonance frequency, and good frequency stability. Therefore, they represent excellent candidates as detectors for mass spectrometry [30], [31] and gas sensing [32]. Recently, VLSI compatible nano/micro-fabricated, high-performance, portable multi-gas analyzers associating gas chromatography and NEMS resonators [33] were reported, confirming NEMS high potentiality.

The first advantage of integrating NEMS with CMOS in a 3-D sequential configuration is to enhance the NEMS SNR. This

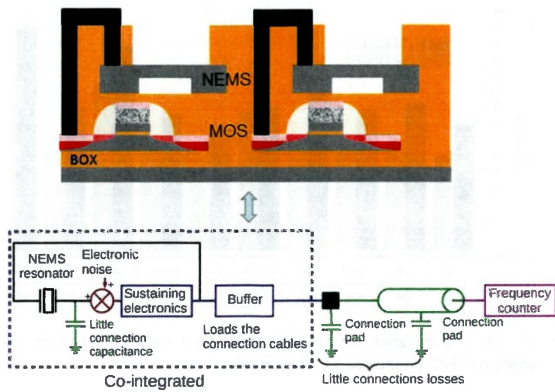


Fig. 16. Equivalent open-loop electrical scheme of a nano-mechanical resonator sequentially integrated in 3-D with its CMOS readout circuitry.

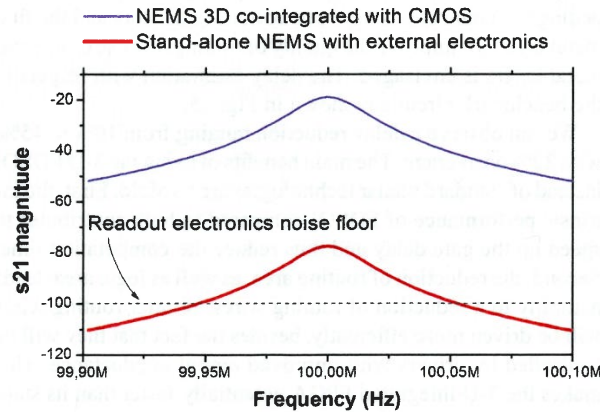


Fig. 17. Calculated S21 (open-loop) magnitude around the resonance frequency of a 100 MHz NEMS resonator. In red, its frequency response in a stand-alone configuration whereby the NEMS chip is connected to an external electronics by wire bonding and coaxial cables. In blue, in a configuration of 3-D sequential integration whereby the NEMS is superposed on top of an unity gain CMOS amplifier.

is achievable with an ultra-compact readout circuitry, additionally the NEMS frequency response is much less contaminated by parasitic peaks and coupling. In fact, by scaling mechanical structures from MEMS down to NEMS, some difficulties appear such as 1) an extremely low detection signal due to their small dimensions, 2) very high impedance due to their small size, and 3) a very low cutoff frequency due to pad and cable capacitances with respect to their resonance frequencies. In this context, the monolithic integration of NEMS resonators with their CMOS readout electronics (Fig. 16) [34]–[36] is the most effective scheme to detect the NEMS resonance signal, i.e., without being attenuated and/or degraded by the effects of external parasitic connections.

Fig. 17 illustrates the gain in dynamic range (resonance signal over noise floor) between two NEMS resonators having the same dimensions and layout, but implemented in two different configurations (stand-alone and 3-D integrated with CMOS). NEMS-CMOS monolithic integration potentially enables the realization of ultra-compact, on-chip NEMS-CMOS self-oscillators ([37] is a first proof-of-concept in 2-D) which allow mon-

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

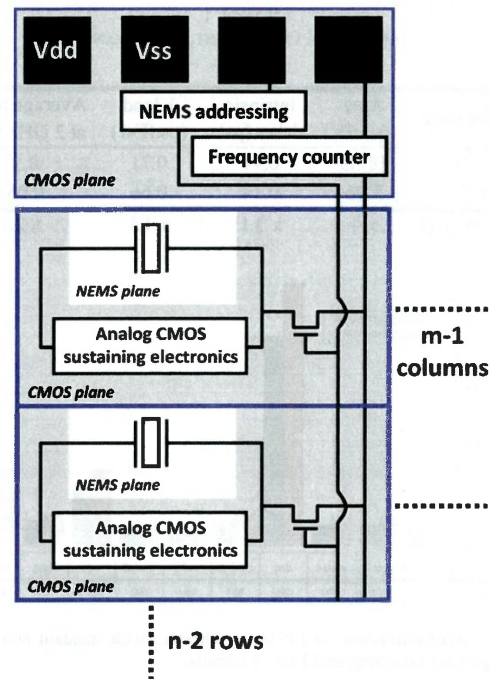


Fig. 18. Schematic representing a $(n \times m)$ array of individually addressable, densely packed, close-loop NEMS-CMOS self-oscillators pixels, whereby NEMS resonators and CMOS circuitry are on two distinct planes according to a 3-D sequential technology.

itoring in real-time the NEMS resonance frequency, as required in most of NEMS-related applications. Additionally, arranging NEMS in arrays of individually addressed resonators seems to be the most promising paradigm to make NEMS-based sensors: such arrays not only provide benefits in terms of robustness and noise reduction (by averaging NEMS individual responses), they may also provide spatial information (like an imager) and increased capture area. In this sense, the possibility to implement an array of NEMS-CMOS self-oscillators pixels [35] would open great possibilities for instance for NEMS-based mass spectrometry.

A 3-D sequential integration of NEMS with CMOS is probably the best way to combine all the aforementioned requirements and to practically realize NEMS-CMOS arrays (Fig. 18). In this way, nanoresonators can be arranged with ultra-high density for an increased capture area, each resonator being individually connected to its own CMOS readout circuit. In the architecture of Fig. 18 (see also [35]), every pixel contains a NEMS-CMOS oscillator which is individually addressable, therefore the resonance frequency of every NEMS can be read sequentially by one single frequency counter. Another interest is that NEMS resonators can be manufactured and packaged independently of the CMOS part. Their structural layer can for instance be single-crystal Si in order to obtain large quality factors and enhanced piezoresistive response (with an optimized doping level and an adequate crystalline orientation, independently of the CMOS transistors).

In conclusion, 3-D sequential integration of NEMS with CMOS would enable implementing new architectures, like

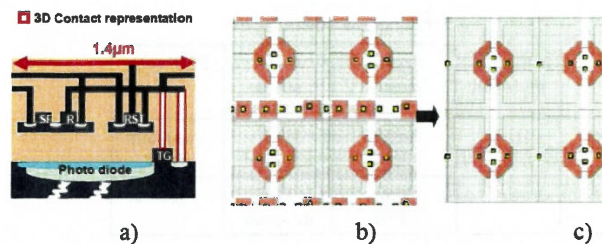


Fig. 19. Highly miniaturized pixel [26]. (a) 3-D pixel schematic. (b) 2-D pixel layout with red = gates/yellow = contacts. (c) 3-D pixel layout at bottom level (readout transistors are moved away to the top level).

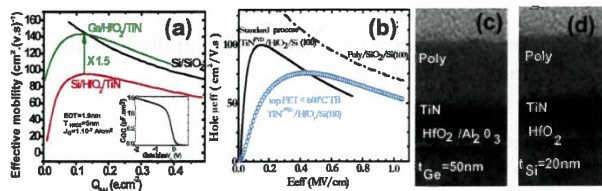


Fig. 20. (a) p-Ge FET or (b) p-(110)Si-pFET stacked above (100)Si-n-FET. (c) Adapted gate stack for Ge top FET. (d) Adapted gate stack for Si [15].

the one of Fig. 18, which offers unprecedented perspectives for NEMS-related, frequency-based sensing applications. It would also be a key technology for other applications, such as NEMS-based logic [38], [39], a growing field of research in the past few years, whereby the possibility to implement a high density of NEM switches, closely connected to CMOS circuits would pave the way for mixed electrical/mechanical logic computation at very large scale.

Sequential integration has some advantages compared to parallel integration: it offers the possibility to address each NEMS independently thanks to the possibility to connect layers at the transistor scale. Also the co-integration of NEMs and CMOS is particularly easy to achieve: a single additional thermal budget is needed (bonding anneal at 200 °C, 1 h) and few process steps are added to stack the NEMS above the transistors. The 3-D contacts in a 3-D sequential integration are similar to the standard contacts in a planar integration, i.e., the contacts are etched in an oxide with small aspect ratios. While, TSV are more challenging to process (new equipment and processes are needed to etch high aspect ratio 3-D VIA in Si).

C. CMOS Image Sensors—Miniaturized Pixels and Diversification

Similarly to Moore's law scaling, CMOS image sensors (CIS) have experienced continuous miniaturization; industrial pixels now reaching 1.1 μm pitch. This trend is favorable in term of resolution but its consequences are somehow detrimental to the image quality. Major improvements have led to high performance back-illuminated pixels with 4-transistors architecture (4 T), but they might be insufficient for further scaling, due to limited full well capacity and exacerbated crosstalk. To overcome these drawbacks, we introduced a 4 T 3-D-integrated miniaturized pixel, where a back-illuminated photodiode layer comes with a second active layer for the readout transistors (Fig. 19) [40].

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

CIS counting amongst the precursors in the 3-D exploration, the idea of a 3-D pixel has been first reported as early as 1987 [41], but has never been applied to 4 T pixels with a pitch of less than 8 μm. State-of-the-art wafer bonding capabilities providing best layer-to-layer overlays of 0.3 μm, the minimum 3-D VIA diameter is found to be 1.25 μm when stacking individually processed layers [42]. This value is far from the mandatory design rules in a miniaturized pixel (80 nm). Miniaturized 3-T pixels, with built-in above-IC photodiodes made of a-Si or Ge [43] have also been proposed, but they remain incompatible with low noise standards, due to the inability of processing a transfer gate on the photodiode layer. Furthermore, they can not be considered as effective 3-D circuits as they incorporate a unique FET layer.

These two limitations vanish in a sequential 3-D process, with an overlay of 10 nm between the two active layers [4], allowing 3-D contact diameters in line with 2-D values. Implementation of such integration on a 1.4 μm pixel shows a 44% increase of the photodiode area, with associated full well enhancement [40]. The photodiode layer process also becomes dedicated, allowing reduction of dark current sources around the diodes. The upper level low-temperature HfO₂/TiN MOS technology does not appear detrimental to the source follower low-frequency noise level performance, and is even further improved by the potential use of relaxed geometries in the free space above the photodiode.

The sequential approach hence becomes the only 3-D scheme extendable to the whole pixel pitch range, from highly miniaturized pixels to large ones with complex in-pixel readout. Material, resistivity and thickness of the substrate on the photodiode layer can be tailor-made without affecting the readout FDSOI MOS technology on the top layer. This technology offers perspectives for CIS diversification, far beyond classical applications, like parallel signal processing. Differing from classical multiplexing operation, this latter allows detection, conversion and digital treatment of the pixels signal in a massively parallel manner. This paves the way to applications with both high dynamic range and high speed, like low consumption real time vision, retinal neuromorphic sensors and particle detection sensors for high energy physics. Finally, the co-integration of heterogeneous semiconductors is being feasible, for instance with a III-V base substrate for IR imaging purposes.

D. Vertical CMOS Cells

Designing vertical CMOS cell, for example stacking the nFET above the pFET, can be considered as the ultimate way to optimize independently each type of transistor. Indeed it allows choosing for each type of transistor, not only the channel material, its crystalline orientation, the strain (via substrate choice and/or CESL and/or specific SD process), but also the gate stack as well as the transistor architecture itself. Such an example of decorrelated transistor optimization is displayed on Fig. 20. Fig. 20(a) and (b) corresponds to the effective mobility curves of top transistors in 3-D sequential integrations in which the bottom transistor is a Silicon nFET with standard (100) orientation. In the inset (a) the top FET is a Germanium p-MOSFET, whereas in the inset (b) the top FET is fabricated on (110) oriented surface. These demonstrations illustrate the

« 20nm node » scenario				« 14nm node » scenario				
type	Channel material	Strain booster		type	Channel material	Strain booster	Gate stack	FET type
		N°1	N°2					
nMOS	Si	sSOI	Tensile CESL	nMOS	III-V	Tensile CESL	III-V adapted	Quantum well
pMOS	Si with SiGe cap	Ge/RSD	Comp CESL	pMOS	Ge	Comp CESL	Ge adapted	FDSOI

Fig. 21. Examples of plausible scenario for 20 and 14 nm node in term of n and pFET independent process adaptation.

ability to co-integrate different channel material and substrate orientation for p and n-MOSFETs in a 3-D sequential integration.

The possibility of optimizing easily the n and pFET independently is a crucial asset. Indeed, as the gate length scaling road map is slowing down, new performance boosters are needed. This introduces new differences in p and n MOSFET process. For example, Fig. 21 presents plausible scenario envisaged to target sub-20 nm and 14 nm node performance, respectively.

In a 20 nm node scenario, a thin SiGe capping layer is implemented for reducing the threshold voltage as well as increasing the channel mobility. Mobility enhancement can also be obtained using strained SOI (sSOI) substrates for nFET. For the pFET performance improvement, SiGe raised or embedded source and drain enable creating compressive strain in the channel. Additional boost in performance can be achieved by using tensile and compressive contact etch stop layers (CESL) for n and pFET, respectively. In a 14 nm node configuration, the differentiation of the n and pFET process is even more complex as some new channel materials might be needed to match the performance targets. III-V materials are considered for the n-MOSFET, whereas Germanium appears to be a good candidate for p-MOSFET. Channel material modification leads also to gate stack and possibly to device architecture modification.

The integration of all these boosters in a planar configuration is very complex and thus costly. At the contrary, vertical CMOS is intrinsically adapted to independent optimization of the different transistor type and will besides lead to density increase. This design is, once again only possible in the 3-D sequential scheme due to its high 3-D contact pitch.

The examples selected in this paper focus intentionally on heterogeneous co-integrations. Still, 3-D sequential integration has also been shown promising for complex digital ASIC [44] and for memory stacks [45].

IV. CONCLUSION

Thanks to its ability to offer fine-grain circuit partitioning at the transistor scale, 3-D 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 3-D transistors matching the targets of advanced nodes.

Authorized licensed use limited to: University of Texas at Austin. Downloaded on December 04, 2023 at 21:01:22 UTC from IEEE Xplore. Restrictions apply.

ACKNOWLEDGMENT

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

REFERENCES

- [1] P. Garrou, C. Bower, and P. Ramm, *Handbook of 3-D Integration: Technology and Applications of 3-D Integrated Circuits*, 1st ed. Hoboken, NJ: Wiley, 2008.
- [2] P. Garrou, C. Bower, and P. Ramm, *Handbook of 3-D Integration: Technology and Applications of 3-D Integrated Circuits*. Hoboken, NJ: Wiley, 2008, vol. 2.
- [3] S.-M. Jung, Y. Rah, T. Ha, H. Park, C. Chang, S. Lee, J. Yun, W. Cho, H. Lim, J. Park, J. Jeong, B. Son, J. Jang, B. Choi, H. Cho, and K. Kim, "Highly cost effective and high performance 65 nm S3 (stacked single-crystal Si) SRAM technology with 25F2, 0.16 μm^2 cell and doubly stacked SSTFT cell transistors for ultra-high density and high speed applications," in *Dig. Tech. Papers Symp. VLSI Technol.*, 2005, p. 220.
- [4] P. Batude, M. Vinet, A. Pouydebasque, L. Clavelier, C. LeRoy, C. Tabone, B. Previtali, L. Sanchez, L. Baud, A. Roman, V. Carron, F. Nemouchi, S. Pocas, C. Comboroure, V. Mazzocchi, H. Grampeix, F. Aussenac, and S. Deleonibus, "Enabling 3-D monolithic integration," in *Proc. Electro-Chemical Soc. (ECS) Spring Meet.*, 2008, vol. 16, p. 47.
- [5] Y.-H. Son, J.-W. Lee, P. Kang, M.-G. Kang, J.-B. Kim, S.-H. Lee, Y.-P. Kim, I.-S. Jung, B.-C. Lee, S.-Y. Choi, U.-I. Chung, J.-T. Moon, and B.-I. Ryu, "Laser-induced epitaxial growth (LEG) technology for high density 3-D stacked memory with high productivity," in *Dig. Tech. Papers Symp. VLSI Technol.*, 2007, p. 80.
- [6] S.-M. Jung, H. Lim, C. Yeo, K. Kwak, B. Son, H. Park, J. Na, J.-J. Shim, C. Hon, and K. Kim, "High speed and highly cost effective 72 M bit density S3 SRAM technology with doubly stacked Si layers, peripheral only CoSi_x layers and tungsten shunt W/L scheme for standalone and embedded memory," in *Dig. Tech. Papers Symp. VLSI Technol.*, 2007, p. 82.
- [7] J. Feng, Y. Liu, P.-B. Griffin, and J.-D. Plummer, "Integration of germanium-on-insulator and silicon MOSFETs on a silicon substrate," *IEEE Electron Device Lett.*, vol. 27, no. 11, p. 911, Nov. 2006.
- [8] H. Liu, M. Kumar, and J.-K.-O. Sin, "A novel BiCMOS technology using selective epitaxy growth and lateral solid phase epitaxial," in *IEEE Int. Electron Device Meet. (IEDM)*, 2001, p. 729.
- [9] T. Chen, R. Ishihara, J. van der Cingel, B. Alessandro, M.-R.-T. Mofrad, H. Schellevis, and K. Beenakker, "Integrated high performance (100) and (110) oriented single-grain Si TFTs without seed substrate," in *IEEE Int. Electron Device Meet. (IEDM)*, 2009, p. 179.
- [10] T. Chen, R. Ishihara, M. R. T. Mofrad, S. Vollebregt, J. van der Cingel, M. van der Zwan, H. Schellevis, and K. Beenakker, "High performance single-grain Ge TFTs without seed substrate," in *IEEE Int. Electron Device Meet. (IEDM)*, 2010, p. 496.
- [11] T. Naito, T. Ishida, T. Onoduka, M. Nishigoori, T. Nakayama, Y. Ueno, Y. Ishimoto, A. Suzuki, W. Chung, R. Madurawe, S. Wu, S. Ikeda, and H. Oyamatsu, "World's first monolithic 3-D-FPGA with TFT SRAM over 90 nm 9 layer Cu CMOS," in *Dig. Tech. Papers Symp. VLSI Technol.*, 2010, p. 219.
- [12] L. Xue, C.-C. Liu, H.-S. Kim, S.-K. Kim, and S. Tiwari, "Three-dimensional integration: Technology, use, and issues for mixed-signal applications," *IEEE Trans. Electron Devices*, vol. 50, no. 3, p. 601, Feb. 2003.

- [13] D.-S. Yu, A. Chin, C.-C. Laio, C.-F. Lee, C.-F. Cheng, W.-J. Chen, C. Zhu, M.-F. Li, W.-J. Yoo, S.-P. McAlister, and D.-L. Kwong, "3-D GOI CMOSFETs with novel IrO_2 (Hf) dual gates and high-k dielectric on $1\text{P6M-}0.18\text{ }\mu\text{m-CMOS}$," in *Tech. Dig. Int. Electron Devices Meet. (IEDM)*, 2004, p. 181.
- [14] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali, C. Tabone, L. Clavelier, S. Michaud, A. Valentian, O. Thomas, O. Rozeau, P. Coudrain, C. Leyris, K. Romanjek, X. Garros, L. Sanchez, L. Baud, A. Roman, V. Carron, H. Grampeix, E. Augendre, A. Toffoli, F. Allain, P. Grosgeorges, V. Mazzocchi, L. Tosti, F. Andrieu, J.-M. Hartmann, D. Lafond, S. Deleonibus, and O. Faynot, "GeOI and SOI 3-D monolithic cell integrations for high density applications," in *Tech. Dig. Symp. Very Large Scale Integration (VLSI)*, 2009, p. 166.
- [15] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali, C. Tabone, J.-M. Hartmann, L. Sanchez, L. Baud, V. Carron, A. Toffoli, F. Allain, V. Mazzocchi, D. Lafond, N. Bouzaida, O. Thomas, O. Cueto, A. Amara, S. Deleonibus, and O. Faynot, "Advances in 3-D CMOS sequential integration," in *IEEE Int. Electron Device Meet. (IEDM)*, Dec. 2009, pp. 1–4.
- [16] P. Batude, M. Vinet, C. Xu, B. Previtali, C. Tabone, C. Le Royer, L. Sanchez, L. Baud, L. Brunet, A. Toffoli, F. Allain, D. Lafond, F. Aussenac, O. Thomas, T. Poiroux, and O. Faynot, "Demonstration of low temperature 3-D sequential FDSOI integration down to 50 nm gate length," in *Tech. Dig. Symp. Very Large Scale Integrat. (VLSI)*, 2011, p. 158.
- [17] B. Sklénard, C. Xu, P. Batude, B. Previtali, C. Tabone, Q. Rafhay, B. Colombeau, F.-A. Khaja, I. Martin-Bragado, J. Berthoz, F. Allain, A. Toffoli, R. Kies, M.-A. Jaud, P. Rivallin, S. Cristoloveanu, C. Tavernier, O. Faynot, and T. Poiroux, "FDSOI devices: A solution to achieve low junction leakage with low temperature processes ($\leq 650^\circ\text{C}$)," in *13th Int. Conf. Ultimate Integrat. Silicon*, pp. 169–172.
- [18] A. Veloso, L.-A. Ragnarsson, Ragnarsson, M.-J. Cho, K. Devriendt, K. Kellens, F. Sebaai, S. Suhard, S. Brus, Y. Crabbe, T. Schram, E. Rohr, V. Paraschiv, G. Eneman, T. Kauerauf, M. Dehan, S.-H. Hong, S. Yamaguchi, S. Takeoka, Y. Higuchi, H. Tielens, A. Van Ammel, P. Favia, H. Bender, A. Franquet, T. Conard, X. Li, K.-L. Pey, H. Struyf, P. Mertens, P.-P. Absil, N. Horiguchi, and T. Hoffmann, "Gate-last vs. gate-first technology for aggressively scaled EOT logic/RF CMOS," in *Tech. Dig. Symp. Very Large Scale Integrat. (VLSI)*, 2011, p. 34.
- [19] S. Y. Lee and D.-K. Schroder, "3-D IC architecture for high density memories," in *IEEE Int. Memory Workshops*, 2010, pp. 1–6.
- [20] S. Y. Lee and J. Park, "Architecture of 3-D memory cell array on 3-D IC," in *IEEE Int. Memory Workshops*, 2012, pp. 1–3.
- [21] D. C. Sekar and Z. Or-Bach, "Monolithic 3-D-ICs with single crystal silicon layers," in *IEEE 3-D Syst. Integrat. Conf.*, 2012, pp. 1–2.
- [22] B. Rajendran, R.-S. Shenoy, D.-J. Witte, N.-S. Chokshi, R.-L. De Leon, G.-S. Tompa, and R. Fabian, "Low thermal budget processing for sequential 3-D IC fabrication," *IEEE Trans. Electron Devices*, vol. 54, no. 4, pp. 707–714, Apr. 2007.
- [23] F.-K. Hsueh, Y.-J. Lee, K.-L. Lin, M.-I. Current, C.-Y. Wu, and T.-S. Chao, "Amorphous-layer regrowth and activation of P and As implanted Si by low-temperature microwave annealing," *IEEE Trans. Electron Devices*, vol. 58, no. 7, pp. 2088–2093, Jul. 2011.
- [24] G.-H. Loh, Y. Xie, and B. Black, "Processor design in 3-D die-stacking technologies," *IEEE Micro*, vol. 27, no. 3, pp. 31–48, May/Jun. 2007.
- [25] M. Lin, A. El Gamal, Y.-C. Lu, and S. Wong, "Performance benefits of monolithically stacked 3-D FPGA," *IEEE Trans. Computer-Aided Design Integr. Circuits Syst.*, vol. 26, no. 2, pp. 216–229, Feb. 2007.
- [26] V. Betz, A. Marquardt, and J. Rose, *Architecture and CAD for Deep-Submicron FPGAs*. Norwell, MA: Kluwer, 1999.
- [27] ABC: A system for sequential synthesis and verification [Online]. Available: <http://www.eecs.berkeley.edu/~alanmi/abc/>
- [28] VPR and T-VPack 5.0.2 [Online]. Available: <http://www.eecg.utoronto.ca/vpr/>
- [29] Y. T. Yang, C. Callegari, X. L. Feng, K. L. Ekinici, and M. L. Roukes, "Zeptogram-scale nanomechanical mass sensing," *Nano Lett.*, vol. 6, no. 4, pp. 583–586, 2006.
- [30] A. K. Naik, M. S. Hanay, W. K. Hiebert, X. L. Feng, and M. L. Roukes, "Towards single-molecule nanomechanical mass spectrometry," *Nature Nanotechnol.*, vol. 4, no. 7, pp. 445–450, 2009.
- [31] M. S. Hanay *et al.*, "Single-protein nanomechanical mass spectrometry in real time," *Nature Nanotech.*, vol. 7, no. 9, pp. 602–608, 2012.
- [32] M. Li, E. B. Myers, H. X. Tang, S. J. Aldridge, H. C. McCaig, J. J. Whiting, R. J. Simonson, N. S. Lewis, and M. L. Roukes, *Nanoletters*, vol. 10, pp. 3899–3903, 2010.
- [33] J. Arcamone, A. Niel, V. Gouttenoire, M. Petitjean, N. David, R. Barattin, M. Matheron, F. Ricoul, T. Bordy, H. Blanc, J. Ruellan, D. Mercier, N. Pereira-Rodrigues, G. Costa, V. Agache, S. Hentz, J. Gabriel, F. Baleras, C. Marcoux, T. Ernst, L. Duraffourg, E. Colinet, E.-B. Myers, M.-L. Roukes, P. Andreucci, E. Ollier, and P. Puget, "VLSI silicon multi-gas analyzer coupling gas chromatography and NEMS detectors," in *Proc. IEEE Int. Electron Devices Meet.*, 2011, pp. 669–672.
- [34] J. Arcamone, M. A. F. van den Boogaart, F. Serra-Graells, J. Fraxedas, J. Brugger, and F. Perez-Murano, "Full-wafer fabrication by nanoscale lithography of micro/nanomechanical mass sensors monolithically integrated with CMOS," *Nanotechnology*, vol. 19, p. 305302, 2008.
- [35] G. Arndt, C. Dupré, J. Arcamone, G. Cibrario, O. Rozeau, L. Duraffourg, E. Ollier, and E. Colinet, "Towards ultra-dense arrays of VHF NEMS with FDSOI-CMOS active pixels for sensing applications," in *Proc. IEEE Solid-State Circuits Conf. Dig. Tech. Papers*, 2012, pp. 320–321.
- [36] J. Arcamone, M. Savoye, and G. Arndt, "VLSI platform for the monolithic integration of single-crystal Si NEMS capacitive resonators with low-cost CMOS," in *IEEE Int. Electron Devices Meet.*, 2012, to be published.
- [37] J. Verd, A. Uranga, G. Abadal, J. L. Teva, F. Torres, J. L. Lopez, F. Perez-Murano, J. Esteve, and N. Barniol, "Monolithic CMOS MEMS oscillator circuit for sensing in the attogram range," *IEEE Electron Device Lett.*, vol. 29, no. 2, pp. 146–148, Feb. 2008.
- [38] M. Spencer, F. Chen, C.-C. Wang, R. Nathanael, H. Fariborzi, A. Gupta, H. Kam, V. Pott, J. Jeon, T.-J. K. Liu, D. Marković, E. Alon, and V. Stojanovic, "Demonstration of integrated micro-electro-mechanical relay circuits for VLSI applications," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 308–320, Jan. 2011.
- [39] O. Y. Loh and H. D. Espinosa, "Nanoelectromechanical contact switches," *Nature Nanotechnol.*, pp. 283–295, 2012.
- [40] P. Coudrain, P. Batude, X. Gagnard, C. Leyris, S. Ricq, M. Vinet, A. Pouydebasque, N. Moussy, Y. Cazaux, B. Giffard, P. Magnan, and P. Ancey, "Setting up 3-D sequential integration for back-illuminated CMOS image sensors with highly miniaturized pixels with low temperature fully depleted SOI transistors," in *IEEE Int. Electron Device Meet. (IEDM)*, 2008, p. 271.
- [41] T. Nishimura, Y. Inoue, and K. Sugahara, "Three dimensional IC for high performance image signal processor," in *IEEE Int. Electron Device Meet. (IEDM)*, 1987, pp. 111–114.
- [42] C. L. Chen, D.-R. Yost, J.-M. Knecht, D.-C. Chapman, D.-C. Oakley, D. L.-J. Mahoney, J.-P. Donnelly, A.-M. Soares, V. Suntharalingam, R. Berger, V. Bolkhovsky, W. Hu, B.-D. Wheeler, C.-L. Keast, and D.-C. Shaver, "Wafer-scale 3-D integration of InGaAs image sensors with Si readout circuits," in *Proc. IEEE Int. Conf. 3-D Syst. Integrat.*, 2009, p. 252.
- [43] S. Kim, T. Kim, M. Kim, D. Park, H. Shim, J. Hwang, S. Jun, O. Jung, J. Yoon, M. Lee, C. Park, J. Han, and J. Hwang, "New integration technology of CIS for high sensitivity small pixel," presented at the Int. Image Sensor Workshop, Bergen, Norway, 2009.
- [44] S. Bobba, A. Chakraborty, O. Thomas, P. Batude, T. Ernst, O. Faynot, D.-Z. Pan, and G. De Micheli, "CELONCEL: Effective design technique for 3-D monolithic integration targeting high performance integrated circuits," in *Proc. 16th Asia South Pacific, Design Automat. Conf.*, 2011, p. 336.
- [45] S.-M. Jung, J. Jang, W. Cho, H. Cho, J. Jeong, Y. Chang, J. Kim, Y. Rah, Y. Son, J. Park, M.-S. Song, K.-H. Kim, J.-S. Lim, and K. Kim, "Three dimensionally stacked NAND flash memory technology using stacking single crystal Si layers on ILD and TANOS structure for beyond 30 nm node," in *Tech. Dig. Int. Electron Devices Meet.*, 2006.

Authors' photographs and biographies not available at the time of publication.