

3D sequential integration:

an alternative path towards CMOS scalability

leti



Outline

What is 3D sequential integration?

Why 3D sequential integration?

Key technological modules





Outline

What is 3D sequential integration?

Why 3D monolithic integration?

Key technological modules





3D sequential integration flow

1/ Bottom Layer process



2/ Top active creation





Thermal budget limitation is needed

3/ Top FET process



4/3D contact formation







Difference with 3D packaging

Packaging integration (e.g: TSV)

1/ Wafers processed separately



2/ Stacking and contacting



Sequential integration

1/ Bottom Layer process



2/ Top layer process



3/ 3D contact formation



Packaging integration: stacked MOSFETs processed in separatly Sequential integration: stacked MOSFETs processed sequentially

3D sequential integration: advantage & challenge

Major asset: Highest 3D contact density

Packaging integration (e.g: TSV)



3D TSV contact pitch 3-8μm 3D contact density 10⁴ -10⁵ /mm²

Two reasons for the high 3D contact pitch:

- Alignment performance
- 3D contact process

Major challenge: Process top MOSFET at low thermal budget

Sequential integration



3D contact pitch <100nm 3D contact density > 10⁸ /mm²

Alignment performance with sequential 3D

SEQUENTIAL ^{3D}



PACKAGING^{3D}



Bonding of patterned films Alignment & bonding at the same time

Sequential 3D:

alignement obtained by lithography. Litho stepper capability ~ 3σ =5nm Packaging 3D:

alignment made during bonding Bonding stepper capability 3σ~ 1μm



3D contact integration scheme



3D contact process similar to a standard planar W plug process

Contact in an oxide with a slightly higher depth No keep out of zone





3D contact density



[3,4]: P. Garrou et al., Handbook of 3D integration, Vol 1,2 (Wiley ed) / [5]: B. Banijamali, ECTC2011 [6]: S-M. Jung et al., VLSI 2005 pp220 / [7]: P. Batude et al, ECS journal 2008, VO16,pp47





Partitioning levels







Outline

What is 3D sequential integration

Why 3D sequential integration

- 1-Road map driven by FET performance
- 2-Road map driven by interconnection delay reduction
- 3-Opportunities of heterogeneous cointegration

Key technological modules



1- Roadmap driven by FET performance





1- Roadmap driven by FET performance



Performance boosters are different for N & PFETs

Independant N&P optimization is easier with P/N stacking



N/P configuration: boosting FET performance



pFET Ge

nMOS SOI

Ge/HfO /TiN

.0.0 (h F.c m 2) .0.0 (h F.c m 2)

Gate bias V_ (V)

0.4

0.3

X 1.5

Si/HfO_/TIN

EOT=1.9nm T Hto2=5nm

0.1

0.0

J_=1.10-7 A/cm2



0.2

Q_{INV} (e.cm⁻²)

leti

Eeff (MV/cm)

The ultimate cointegration III-V and (Si)Ge

nFET III-V and pFET Ge process are highly different:

Dual active etching/ Dual Gate stack/ Dual RSD/ different optimum architecture/ Dual salicidation/Substrate fabrication: dual material (epi with ART or III-V bonding on GeOI)...

Processing independantly n&pFET on distinct levels enable to save a lot of lithography levels and process co-integration challenges.



Additionnaly, III-V and Ge transistors requires lower thermal budget process. These devices are well adapted to 3D sequential integration.

T. Irisawa et al., VLSI 2013 (AIST)





Choice of architectures



For bottom level, every existing technology can be used On top level, all the thin film technology can be used





Partitioning levels **Sequential 3D 4-Transistors 1-Entire core** 2-Logic bloc **3-Logic gates** Vdd L2 d2d vias Multiple gnd CPU Granularity scale Packaging 3D

N/P or P/ N stacking → IC Gain obtained by boosting FET performance

2- Roadmap driven by interconnection delay





2- Roadmap driven by interconnection delay

IC's performance are strongly limited by interconnection delay



Wirelength must be reduced to benefit from gate delay reduction



leti

Partitioning levels Sequential 3D





CMOS/ CMOS stacking -> IC Gain obtained by decreasing wirelength



CMOS/ CMOS option: Reducing wirelength

The 3D contact links blocks of MOSFETs

To avoid routing congestion, introduction of intermediate lines is mandatory





PPA gain thanks to wirelength reduction

Case of FPGA circuit 14nm planar FDSOI versus 2 stacked 14 nm FDSOI levels

Study taking into account parasitics / 3 metal line between the 2 stacked layers Partitioning SRAM memory on bottom level, logic on top





Analysis on a specific application: FPGA

- Power Performance Area (PPA) benchmark for 3D versus planar
 - Area gain=55%
 - Perf gain = 23%
 - Power gain = 12%
- Decreased delay and power due to shorter wirelengths
 - reduced wire capacitance
 - less signal buffering requirement

1,5 node gain without scaling Stacking 14nm/ 14nm leads to PPA below 10nm



Specific caution for IC performance evaluation

N°1: Evaluation gain will depend on the considered node and on technology hypothesis (Design Rule Manual)

PPA on FPGA for <u>65nm²</u> vs 65nm [1]: P. Gaillardon *et al.,* CASFET 2012

No intermetal lines

Area gain: 21% in average Delay gain: 22% in average

PPA on FPGA for <u>14nm²</u>vs 14nm

[2]: O. Turkyilmaz et al., DAC 2014

3 level of intermetal lines W lines , no low-k dielectrics

Area gain: 55% in average Delay gain: 23% in average

Evaluation gain will depend on the considered node



Specific caution for IC performance evaluation

- N°2: No place & route (P&R) tool adapted for 3D
- → Results provided only for FPGA thanks to it's regular architecture that can be achieved with full custom layout
- → P&R tool enable to optimize the position of the cells to obtain the best gain (in term of area, timing or power)
- \rightarrow Preliminar evaluation via « DIY » P&R tools



Reliable conclusion can't be provided without dedicated 3D P&R tool

leti

N°3: PPA Gain will highly depend on the application

Evaluation for ASiCs 45nm node/ 1 intermediate level only / with 2D modified P&R tool



Gain depends on the application



2- Roadmap driven by interconnection delay : Conclusion

Evaluation on FPGA is feasible thanks to its regular design (Full custom layout possible).

FPGA with two 14 nm stacked levels should enable to outperform the 10nm

Stacking more efficient than scaling?

No modification of transistor technology (tool reuse)

Gain obtained only due to wirelength reduction





2- Roadmap driven by interconnection delay : conclusion

Specific cautions to make a proper PPA benchmark:

1: Evaluation for other ICs cannot be properly performed due to the lack of 3D P & R tool \rightarrow work with EDA providers is necessary

#2: The results will greatly depend on the IC (predominance of delay in interconnection delay in the full IC performance)

#3: The result will greatly depend on the considered node (predominance of delay in advanced nodes)

#4: Technology assumptions must be analyzed carefully, particularly the number of intermediate lines authorized as well as the metal and dielectrics chosen.





3- A key technology for heterogeneous co-integration

Advantages of 3D

 \rightarrow Independant optimisation of each level

 \rightarrow Proximity between stacked functions

Only 3D sequential technology available if 3D contact pitch is smaller than $1\mu m$





3D integrated CMOS Image Sensors (3D-CIS)

Concept: Vertical dissociation of basic pixel operations



Advantages of 3D configuration:

Photodetection \rightarrow dedicated material and processes, 100% fill factor Readout transistors \rightarrow optimized process, relaxed geometries Signal Processing \rightarrow Massively parallel treatment, dense IC integration



30

Miniaturized 3D CMOS Image Sensors

Innovative miniaturized 4T pixels with backside illumination (BSI)^[24]

- Bottom layer: pinned photodiode + Transfer Gate
- Top layer with 3 transistors



Multiple benefits

Ceatech

- BSI integration \rightarrow high quantum efficiency
- Photodiode area +44% for 1.4µm pitch pixel

Only sequential integration can address these dimensions



3D sequential integration of NEMS with CMOS

NEMs = ultra sensitive mass sensors used for gas sensing & mass spectroscopy

MEMS \rightarrow NEMS enable to increase the sensivity but resonance detection is more complex

The challenge of detecting NEMS resonance: NEMS-CMOS integration schemes

Stand-alone NEMS + off-chip CMOS No density (pads number limitation) Very strong signal attenuation (LP filter)



3D sequential NEMS-CMOS No density limitation no signal attenuation



3D sequential NEMS-CMOS co-integration can solve the NEMS density and detection limitation



3D sequential integration of sensors with CMOS

Moving to NEMS array: i.e each NEMS is individually addressed.

 \rightarrow

→ Multiple benefits: increased robustness (redundancy), better SNR (averaging NEMS individual response, provide a spatial response (like an imager), ultra-high density to improve capture efficiency (smaller concentration detection)

> NEMS array enables novel NEMS-based applications such as mass spectrometry



At this scale, only 3D sequential integration enable to reach the 3D contact pitch required (NEMS dimensions converge towards CMOS transistors)

Easy integration of NEMS thanks to it low thermal budget.

3D sequential is well adapted for NEMS array

Rmk: same benefits applicable to other domains such as NEMS-based logic



Outline

What is 3D sequential integration

Why 3D sequential

- 1-Road map driven by FET performance
 - 2-Road map driven by interconnection delay reduction
 - 3-Opportunities of heterogeneous cointegration

Key technological modules

- 1-Maximum top layer thermal budget determination
- 2-Top active creation
- 3-Top FET at low temperature process



General Integration Flow



Which thermal budget?





Part III - Key technological modules

1- Maximum top FET thermal budget determination




Making 3D sequential integration possible







Making 3D sequential integration possible







• Simple method: Annealing of a transistor and observe what is the critical thermal budget for preserving its performance

• This critical thermal budget will depend of the technology (BULK, FINFET, FDSOI) and the node analyzed



FDSOI with implanted Si RSD stability



FDSOI with implanted RSD: max thermal budget = 500°C



FDSOI with implanted Si RSD stability

• No impact of the different anneals shown on multiple key electrical parameters such as:

• DIBL, short channel mobility evidencing no additional dopant diffusion

• EOT regrowth and reliability (NBTI & PBTI) evidencing good stability of the gate stack



FDSOI with implanted Si RSD stability NMOS PMOS MOS PMOS



Ref without annealing



After annealing

• 550°C 1h30 : anneal degrades the silicide on NMOS and PMOS

Need to improve NiPt 10% silicide thermal stability





NiPt silicidation associated with W offers higher thermal stability





Techno 1: FDSOI [1]

-Si channel -Si RSD and implanted source & drain -NiPt 10% salicidation \rightarrow Performance degradation above 500°C,5h

Techno 2: FDSOI with additional boosters [2]

-Si channel for NFET, SiGe for pFET -SiGe:B & SiC:P in situ doped source & drain -NiPt 15% salicidation \rightarrow No Performance degradation up to 550° 2h & 500°C 5h

Weak point of MOSFET technologies is salicide stability

Improving salicide stability could relax top FET maximum thermal budget

Most secure criteria: max thermal budget= 500°C (5h)



NiPt 10%

nFET



Ceatech [1] C. Fenouillet- Beranger et al., ESSDERC 2014 [eff] [2] C. Fenouillet- Beranger et al., IEDM 2014

Part III - Key technological modules

2- Low temperature top active layer creation



Objective:

high cristalline quality/ controlled thickness/ Max TB =500°C



Cristallization of amorphous Silicon



Finding solutions to increase the size of the grains at temperature compatible with bottom MOSFET integrity





Cristallisation of amorphous Silicon

Example of laser annealing



Rmk: Not a lot of application can tolerate the performance dispersion brought by the poly-Si

Ceatech

"Seed window" techniques

To obtain the control of the grain boundaries position and grain orientation







"Seed window" techniques



Remark: Too high thermal budgets applied to the bottom MOSFET





"Seed window" techniques



Liquid phase recristallization on Si with laser

Important challenges needed to be adressed:

Thickness control is mandatory Supression of cristalline defects Reduction of seed window surface penalty

Liquid phase recristallization on Ge with RTP



J. Feng et al., EDL 06, Stanford



Y-H. Son et al, VLSI 2007 & ECS 2013, Samsung

µ-Czochralski

Enable to grow a monocrystalline semiconductor seed on an oxide Control of grain position



Could solve the thermal budget issue, but some points need to be improved:

- -The window is large : important loss in density
- -The window is deep: very high aspect ratio for 3D contact
- -The thickness of the film is highly variable





Carbon nanotubes

Decoupling high temperature CNT growth and 3D sequential integration



BULK direct bonding and ion slicing





Examples of ion slicing reports on bottom MOSFET level



D-S. Yu et al., IEDM 2004, Nat'l Chiao-Tung Univ, Taiwan



L. Xue et al., TED 2003, Cornell university



Patterned reports might be due to:

-Non perfect CMP

-Bonding done with deposited oxide (which contains H_20), patterning avoid the defect due to degassing

ightarrow Solutions described in the following slides

F. Crnogorac et al., JVSTB 2010, Stanford



SOI direct bonding









High quality top film

Blanket Si film on top of a bottom transistor layer:

Stack cross section

Top view:

Acoustic characterization:



Full transfer on processed 300 mm wafers



	Seed window (SW)	Poly-Si	Wafer bonding
Description	SW PMD oxide	PMD oxide	PMD oxide
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
Thermal budget	Seems incompatible with bottom max TB	Ok with ns laser	<400°C



	Seed window (SW)	Poly-Si	Wafer bonding
Description	SW PMD oxide	PMD oxide	PMD oxide
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 Not compa	tible with high per	formance devices
layer orientation	same orientation	random orientation for top substrate	different orientation possible
Thermal budget	Seems incompatible with bottom max TB	Ok with ns laser	<400°C



	Seed window (SW)	Poly-Si	Wafer bonding
Description Densit Too hi	y limitation gh thermal budget	PMD oxide	PMD oxide
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	10¢ nm range	um range	Å range
layer orientation	same orientation	random orientation for top substrate	different orientation possible
Thermal budget	Seems incompatible with bottom max TB	Ok with ns laser	<400°C

	Seed window (SW)	Poly-Si	Wafer bonding
Description Densit	y limitation gh thermal budget	PMD oxide	PMD oxide
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	10¢ nm range	um range	Å range
layer orientation	same orientation	random orientation for top substrate	different orientation possible
Thermal budget	Seems incompatible with bottom max TB	Ok with ns laser	<400°C

More details on direct bonding flow



Constraints brought by 3D sequential integration vs std direct bonding:

Deposited oxides Planarization of topography Unusual MEOL→FEOL flow



Importance of perfect CMP and particle free bonding surface



Perfect CMP





Solution for bonding with deposited oxides at Low Temp



Degassing phenomenon:

Ceatech

Solution for bonding with deposited oxides at Low Temp

Solution proposed \rightarrow Si₃N₄ H₂O barrier to avoid Si oxidation



leti



L. Brunet et al, ECS fall meeting 2014

Part III - Key technological modules

3-Top FET process at low temperature





Low temperature MOSFET main challenges





Low temperature top transistor







Low temperature dopant activation



6nm Si channel SiO₂/HfSiON/TiN patterning First Si₃N₄ spacer Raised S/D epitaxy (Si: nFET/ SiGe: pFET) Low Doped Drain implantation Second spacer High Doped Drain implantation LT=600 °C / HT=1050 °C anneal Salicidation





Ceatech

SPER activated FDSOI devices



SPER process leads to similar performance than High T activation \rightarrow High dopant activation level with 600°C process



leti

SPER activated FDSOI devices

- → Standard drawbacks observed on bulk devices:
 Junction leakage increase
 - Deactivation
- → SPER FDSOI device do not present same drawbacks, why?



Solid Phase Epitaxial Regrowth

Amorphizing implant (e.g. As or Ge + B)



EOR defects formation below former amorphous-cristalline interface


End of Range (EOR) formation

Two phenomena during anneal:

Ceatech

1- Defects growth through an Oswald ripening process



2- EOR defects dissolution via recombination at the interfaces



End of Range defects in FDSOI vs BULK



Junction leakage



For thin channel, no EOR defects at the channel entrance
→ No junction leakage increase



Boron deactivation for thin SOI and BULK

Rsheet evolution with post activation anneal (spacer, BEOL..)



Boron deactivation can be suppressed for Tseed = 5nm



Dopant deactivation in SOI

EOR evolution with post anneals

Original situation after recristallisation anneal (original EOR reservoir is smaller in thin SOI)



During post anneal: emission of interstitial Si atoms from the EOR joining the closest sink



Scaling T_{seed} enable to: - reduce original EOR concentration

- reduce the Si_i flux crossing the highly activated region

avoiding formation of inative Boron interstitials clusters



Optimizing the sheet resistance



Ceatech

- LT : 2min 600°C SPER



□ Applying HT POR implant conditions leads to severe resistance degradation

❑ Reducing implanted dose, R_{sheet} reduced for both As and P

Optimizing the sheet resistance



Dopant Clustering: above a certain dopant concentration, for a given T, part of the dopant forms inactive clusters.



Optimizing the sheet resistance



C: active concentration, $\boldsymbol{\mu}$ mobility of carriers

extraction by Hall measurements



Implantation with concentration higher than clustering limit degrades either the active level and/ or carrier mobility

Dopant concentration must not exceeds the clusterisation limit



Main learnings from SPER junctions optimization

• FDSOI is a key asset for low temperature junction

Indeed EOR concentration reduction enable to supress:

- Junction leakage issues
- Boron deactivation phenomenon
- Optimization of dopant concentration is needed in order to avoid important clusterization phenomena
- SPER activated devices have similar performance than HT spike activated devices



Solutions for activation in 3D sequential integration



Solutions for activation in 3D sequential integration



83

Laser activation

Efficiency of laser ns anneal to activate dopants proven



Protection of bottom MOSFET during laser anneal validated



ightarrow Promising techniques for dopant activation

- → Additionnal work is needed to evaluate its' interest for 3D sequential integration: Find the laser anneal conditions to have the best gate/source-drain selectivity Evaluation of 3D pattern effect on scaled design rules and versus density Evaluation on thin SOI devices Compatibility with scaled ILD thicknesses
- ightarrow Some answers to be found in C. Fenouillet IEDM 2014 paper

Microwave annealing

✓ Equivalent activation level at 150°C lower than SPER



- ightarrow Promising techniques for dopant activation
- → Additionnal work is needed to evaluate its' interest for 3D sequential integration: Evaluation of metal interconnections (temperature depending on conductivity)















Goal: Obtain a selective epitaxy on the source & drain

Standard process: dichlorosilane and HCl flow together

dichlorosilane : deposition of Si and etch on oxide isolation HCL: etching on nitride spacer



Reasonable deposition rate cannot be obtained at temperature below 750°C



Low temperature epitaxy

Cyclic Deposition Etch (CDE) process



Step 2-selective etch steps



→ Use disilane instead of dichlorosilane enable to increase Si deposition rate



 \rightarrow Addition of GeH4 to HCl enable to increase the etching speed







Low temperature epitaxy

Dramatic decrease of Epitaxy thermal budget is achievable via CDE process



Si epitaxy @600°C J-M. Hartmann et al., Semicond. Sci. Technol 2013



SiGe 35% epitaxy @500°C J-M. Hartmann et al., ECS journal 2014

Use of new precursor for deposition such as Si_3H_8 and Cl_2 for etching are promising for further thermal budget reduction [M. Bauer et al., Thin Solid fim 2012]

Work is still needed to lower the epitaxy down to 500°C

BOTTOM FET PROCESS





Low temperature spacer

Requirements:

- Resistance against HF chemistry (epitaxy pre-clean)
- Stability against subsequent epitaxy thermal budget
- Conformal

Some Low-k spacers seem well adapted : Low temperature (<500°C) and lowk (<6) for parasitic capacitance reduction and delay improvments</p>





Fig. 3. TEM (left) and SEM (right) of SiCOH spacers (NFET) $% \left({{\rm NFET}} \right)$



BOTTOM FET PROCESS





BOTTOM FET PROCESS





Some examples of 3D sequential demonstrations



S³ SRAM technology from Samsung

Dopant activation by high temperature spike anneal *

Top active creation by Selective Epitaxy Growth or laser liquid phase recrystallisation**

*Impossible with intermetal lines

** Incompatible with bottom MOSFET stability at advanced nodes and high loss in density due to seed window area penalty

Some examples of 3D structures



Ultimate CMOS technology from AIST (InGaAsOI nFET on SGOI pFET) SBH adaptation with appropriate metallization Top activate creation by In GaAs Direct Bonding

Transistor and Memories with vertical channel

leti







Surrounding gate transistor from Besang Top activate creation by direct bonding Dopant activation made before wafer report

Technology licensed by Hynix





Thin Film Transistor (Poly-Si)



Ceatech

Characteristics	With CMP	Without CMP
SS 2 σ , mV/dec.	117 ± 32	168±59
Vth 2σ, volt	0.82±0.23	1.02±0.35
lon 2σ, u A/um	89 ±23	40±24

High variability due to grain variable orientation and grain boundaries



Low performance compared to cristalline channel

« Epi-like Si FET » from NDL
 Dopant activation by thermal activation at unknown temperature
 Top activate creation by crystallization of amorphous deposition by laser annealing

Carbon nanotubes







Carbon nanotubes transistors from Stanford Top activate creation obtained by report of carbone nanotube



100

Some examples of 3D sequential demonstrations



Cool Cube[™] technology from Leti Dopant activation by SPER Top active creation by SOI direct bonding

Conclusion

3D sequential integration leads to ultra high 3D contact density $1x10^{6}$ /mm² is demonstrated/ > $1x10^{8}$ /mm² is achievable with 14 nm technology

Roadmap driven by transistor performance :3D monolithic N/P

Main advantage: easier and cheaper way to optimize the transistors' performance a path for III-V cointegration with Ge

Roadmap driven by IC performance: 3D mono CMOS/ CMOS

a path for reducing interconnection delay penalty Stacking instead of scaling: Reaching n+1 node with n node technology

Cointegration of heterogeneous functions requiring small grain partitionning Highly miniaturized CMOS image sensors/ NEM with CMOS



Conclusion

Bottom transistor preservation:

Preserved at 500°C for hours (FDSOI with NiPt salicide)

Solution might be available to increase salicide stability above to 500°C

Top transistor process:

- Direct bonding enable to achieve top active layer equivalent to bottom substrate

- Dopant activation with SPE activation leads to similar performance than the high temperature standard process

- FDSOI is a crucial asset to achieve high performance LT CMOS (EOR reduction)

-Laser (nanosecond) is a promising option for dopant activation thanks to local annealing of top layer

<u>Ceatech</u>

Thank you to all co-authors and colleagues that have been working on Cool Cube[™]

B. Sklenard, C. Fenouillet-Beranger, L. Pasini, B. Sklenard, B. Matthieu, L. Brunet, C. Xu, B. Previtali, C. Tabone, F. Ponthenier, N. Rambal, F. Deprat, L. Tosti, L. Hortemel, M-P. Samson, O. Rozeau, O. Billoint, O. Turkyilmaz, H. Sarhan, G. Cibrario, A. Pouydebasque, F. Fournel, L. Benaissa, T. Signamarcheix, A. Seignard, C. Euvrard-Colnat, M. Rivoire, F. Nemouchi, V. Carron, F. Piegas Luce, F. Mazen, P. Besson, A. Royer, C. Agrafeil, P. Coudrain, Julien Arcamone, P-E. Gaillardon, S. Bobba, T. Ernst, C. Deguet, F. Geiger, J-E. Michallet, C. Reita, F. Clermidy, O. Faynot and M. Vinet

Thank YOU for your attention



This work is partly funded by the ST/IBM/LETI Alliance program and by Qualcomm