Sequential 3D Integration - Design Methodologies and Circuit Techniques

PANAGIOTIS CHAOURANI

Doctoral Thesis in Information and Communication Technology
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology
Stockholm, Sweden, 2019

KTH School of Electrical Engineering
and Computer Science (EECS)
Division of Electronics
Kistagången 16 SE-164 40 Kista
SWEDEN

TRITA-EECS-AVL-2019:54
ISBN 978-91-7873-228-9

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framläggs till offentlig granskning för avläggande av doktorsexamen i Informations- och Kommunikationsteknik fredagen den 30 augusti 2019 klockan 13.00 i Sal C, Kungl Tekniska högskolan, Kistagången 16, Kista, Stockholm.

Tryck: Universitetsservice US AB

To my family

Abstract

Sequential 3D (S3D) integration has been identified as a potential candidate for area efficient ICs. It entails the sequential processing of tiers of devices, one on top the other. The sequential nature of this processing allows the inter-tier vias to be processed like any other inter-metal vias, resulting in an unprecedented increase in the density of vertical interconnects. A lot of scientific attention has been directed towards the processing aspects of this 3-D integration approach, and in particular producing high-performance top-tier transistors without damaging the bottom tier devices and interconnects. As far as the applications of S3D integration are concerned, a lot of focus has been placed on digital circuits. However, the advent of Internet-of-Things applications has motivated the investigation of other circuits as well.

As a first step, two S3D design platforms for custom ICs have been developed, one to facilitate the development of the in-house S3D process and the other to enable the exploration of S3D applications. Both contain device models and physical verification scripts. A novel parasitic extraction flow for S3D ICs has been also developed for the study of tier-to-tier parasitic coupling.

The potential of S3D RF/AMS circuits has been explored and identified using these design platforms. A frequency-based partition scheme has been proposed, with high frequency blocks placed in the top-tier and low-frequency ones in the bottom. As a proof of concept, a receiver front-end for the ZigBee standard has been designed and a 35% area reduction with no performance trade-offs has been demonstrated.

To highlight the prospects of S3D RF/AMS circuits, a study of S3D inductors has been carried out. Planar coils have been identified as the most optimal configuration for S3D inductors and ways to improve their quality factors have been explored. Furthermore, a set of guidelines has been proposed to allow the placement of bottom tier blocks under top-tier inductors towards very compact S3D integration. These guidelines take into consideration the operating frequencies and type of components placed in the bottom tier.

Lastly, the prospects of S3D heterogeneous integration for circuit design have been analyzed with the focus lying on a Ge-over-Si approach. Based on the results of this analysis, track-and-hold circuits and digital cells have been identified as potential circuits that could benefit the most from a Ge-over-Si S3D integration scheme, thanks to the low on-resistance of Ge transistors in the triode region. To improve the performance of top-tier Ge transistors, a processing flow that enables the control of their back-gates has been also proposed, which allows controlling the threshold voltage of top-tier transistors at runtime.

Keywords: Sequential 3D integration, monolithic inter-tier vias, design platforms, parasitic extraction flows, RF/AMS circuits, inductors, heterogeneous integration, germanium transistors.

Sammanfattning

Sekventiell 3D (S3D) teknologi är en potentiell kandidat för att tillverka area effektiva integrerade kretsar (IC). Sekventiell 3D teknologi innebär sekventiell tillverkning av transistorer i lager ovanpå varandra. Den sekventiella tillverkningen möjliggör elektriska kontakter mellan transistorlagren på samma sätt som elektriska kontakter normalt tillverkas mellan metallager, vilket ger exceptionellt hög densitet av vertikala kontakter mellan transistorlagren. Forskningen har fokuserat på att utveckla processteknologin för S3D och i synnerhet etablera högpresterande transistorer i högre lager utan att degradera prestandan hos transistorer i de undre lagren. Vad gäller applikationerna för S3D-integration har mycket fokus varit inriktat på digitala kretsar. Med ökningen av Internet-of-Things applikationer så motiveras forskning på andra kretsar också.

Som ett första steg har två S3D designplattformar utvecklats, en för att stödja utvecklingen av KTH's egna S3D process och den andra för att utforska potentialen för olika S3D applikationer. Båda designplattformarna innehåller komponentmodeller och fysikaliska verifikationsverktyg. Ett nytt extraktionsflöde för parasit resistanser, kapacitanser och induktanser i S3D ICs har också utvecklats för att studera kopplingen mellan transistorlagren.

Potentialen för S3D RF/AMS kretsar har undersökts och identifierats med hjälp av dessa designplattformar. En frekvensbaserad uppdelning har föreslagits, med högfrekventa block placerade i topplagret och lågfrekvens block i bottenlagret. Som "proof-of-concept" designades en mottagare för ZigBee standarden och en 35% areareduktion erhölls utan att prestanda degraderade.

För att ytterligare utforska S3D RF/AMS-kretsar, så utfördes en studie av induktanser i S3D teknologin. Planara spolar identifierades som den mest optimal konfiguration för S3D-induktanser och olika sätt att förbättra deras prestanda undersöktes. Dessutom förslås en uppsättning riktlinjer för att kunna placera kretsblock i bottenlagret, under en induktor i topplagret, för att erhålla en kompakt S3D integration. Dessa riktlinjer tar hänsyn till frekvensen och typen av komponenter som placeras i botten lagret.

Slutligen har möjligheten för kretsdesign med heterogen S3D integration analyserats med fokus på Ge transistorer i ett övre lager och Si transistorer i ett undre lager. Baserat på resultaten av denna analys, så har "track-and-hold" kretsar och digitala celler identifieras som potentiella kretsar som kan få mest nytta av ett Ge över Si S3D-integrationflöde. Detta tack vare Ge-transistorernas låga resistans i triodregionen. För att förbättra prestanda hos Ge transistorerna i topplagret, så har ett processflöde föreslagits som möjliggör elektrisk kontroll av Ge kanalen från en elektrod under Ge kanalen, vilket tillåter att tröskelspänningen hos Ge transistorerna kan kontrolleras under drift.

Nyckelord: Sekventiell 3D integration, design plattformar, parasitiska extraktionsflöden, RF/AMS-kretsar, induktorer, heterogen integration, germanium transistorer

Acknowledgements

First and foremost, I would like to express my deepest gratitude for my main supervisor, Professor Ana Rusu, for entrusting me with a position at her research group at KTH. The past five years, it has been a real honor to collaborate with her. Her supervision has left a clear impact on the way I perceive things, not only in the academic world, but also in life. She has tirelessly provided me with her help and guidance, and inspired me to always strife for the best. I will be always indebted to her.

My gratitude extends also to my co-supervisor, Assistant Professor Saul Rodriguez. His advices on circuit-design aspects have been of great value to me. I am also grateful for his continuous positive attitude that has instilled to me the desire to never give up.

Special thanks to my other co-supervisor, Docent Per-Erik Hellström for his precious help in topics related to process integration and device physics. Everything I know on these two fields, I owe them to him. I am also grateful for our collaboration is setting up the design platform for the in-house process. My appreciation extends to Professor Mikael Östling for leading the Ge3D research project I was working for during my PhD studies. I am also thankful to Professor Gunnar Malm for his precious help in topics related to device physics and TCAD simulations.

This thesis would have been impossible without the financial support of the Swedish Foundation for Strategic Research (Stiftelsen för Strategisk Forskning, SSF).

My appreciation goes to Liviu Popa and Satyabrata Ghosh from Cadence® for their invaluable help in setting up a S3D parasitic extraction flow. I am also thankful to Lars Riekehr from Ångström Laboratory for the cross-section photos of the in-house vias and devices. Laura Žurauskaitė from KTH is acknowledged for providing me with the measurement data of the in-house devices, as well as for the top-view photos of them.

I am deeply thankful to Professor Spyridon Nikolaidis, from Aristotle University of Thessaloniki, Greece for the opportunities he provided me with during my Bachelor and Master studies, and his guidance during my master thesis. I would also like to thank Professors Stylianos Siskos and Theodore Laopoulos for introducing me to the world of electronics.

My gratitude extends to my colleagues from the Integrated Circuits group at KTH for the great moments we passed together. Dr Sha Tao, Dr Tingsu Chen, Dr Janko Katic, Dr Nikola Ivanisevic, Muhammad Waqar Hussain, Dagur Ingi Albertsson, I will be always grateful for our discussions and our joyful breaks from the everyday routine. Special thanks to Dr Raul Onet. His eagerness to help and assist others has been really inspiring for me.

I am also thankful to all my colleagues from the division of Electronics at KTH, Dr Ali Asadollahi (Pooria), Ahmad Abedin, Kostas Garidis, Laura Žurauskaitė, Dr Hossein Elahipanah, Dr Ganesh Jayakumar, Dr Shuoben Hou, Viktoriia Mishukova, Szymon Sollami Delekta, Mattias Ekström, Muhammad Shakir, Corrado Capriata,

Babak Taghavi, Dimitris Stathis and Yang Yu. I will always cherish the time we passed together.

It is a huge honor for me to have Professor Georges Gielen as the opponent at my defense. I am looking forward to discussing the results of my thesis with him. I am also grateful to Professor Christian Fager from Chalmers University, Associate Professor Adriana Serban from Linköping University, and Associate Professor Joachim Rodrigues from Lund University for accepting the invitation to participate in my grading committee. My gratitude extends to Professor Carl-Mikael Zetterling from KTH for reviewing my thesis. His feedback and constructive suggestions have contributed to the improvement of this work.

I could not stress more my gratitude to my family: my mother Maria, my father Ilias, my sister Adlida, my grandmother Alexandra, my aunt Sophia and my uncle Yiannis. They have been always there for me and they have provided me with their unconditional love and support. I am deeply grateful for the sacrifices they have endured to provide me with everything I needed. Of course, I would also like to thank my good friends Mezes, Oliver, Vito, Buddy, Daphne and Clara for always cheering me up.

Ultimately, I would like to thank my dear wife Pinelopi. Her support and love were crucial for me to complete this thesis. She has always helped me to overcome every obstacle I encountered, not only during my PhD studies, but at every other aspect of our common life. I cannot imagine my life without her.

Panagiotis Chaourani,
Stockholm, August 2019

Contents

Contents **ix**

List of Figures **xii**

List of Tables **xvii**

List of Acronyms **xix**

1 Introduction **1**
1.1 New Directions for Moore's Law 1
1.2 3D Integration Technologies 4
1.3 Motivation 6
1.4 Research Objectives 7
1.5 Research Contributions 8
1.6 List of Publications 9
1.7 Thesis Organization 10

2 Insights into the processing of 3D ICs **13**
2.1 TSV processing options 13
2.1.1 Via-first TSVs 13
2.1.2 Via-last TSVs 14
2.1.3 Via-middle TSVs 16
2.1.4 Comparison between via-middle and via-last TSVs 17
2.2 Issues with TSV processing 18
2.2.1 Copper contamination 18
2.2.2 Stress-related issues 19
2.3 Stacking Techniques with TSVs 22
2.3.1 Die-to-Die bonding 23
2.3.2 Direct Wafer bonding or Wafer-to-Wafer bonding 23
2.3.3 Die-To-Wafer bonding 24
2.4 Sequential 3D Integration 25
2.4.1 Si over Si 26
2.4.2 Heterogeneous S3D Integration 29

2.4.3 S3D integration of Ge FETs over Si-FETs 32
2.5 Heat Dissipation in 3D ICs 33
2.6 Summary 35

3 Sequential 3D Design Platform for Custom Integrated Circuits 37
3.1 Sequential 3D Design Platforms 37
3.1.1 Overview 37
3.1.2 Custom IC Design Platforms 38
3.2 KTH-S3D PDK 39
3.2.1 KTH-S3D process stack 40
3.2.2 KTH-S3D PCELLs 41
3.2.3 KTH-S3D Device Models 43
3.2.4 KTH-S3D Physical Verification 44
3.2.5 KTH-S3D parasitic extraction flow 47
3.2.6 Digital Cells Library 51
3.3 S3D Predictive PDK 54
3.3.1 S3D PPDK Process Stack 56
3.3.2 S3D-150 nm PPDK Devices and PCells 57
3.3.3 S3D-150nm PPDK Device Models 59
3.3.4 S3D PPDK Physical Verification 62
3.3.5 S3D-150 nm PPDK parasitic extraction flow 62
3.4 Summary 64

4 Sequential 3D Receiver Front-End 65
4.1 Applications of S3D integration 65
4.1.1 State-Of-The-Art 65
4.1.2 Motivation for S3D RF/AMS circuits 66
4.2 Design of the S3D receiver front-end 69
4.2.1 High frequency blocks in the top tier 69
4.2.2 Baseband blocks in the bottom tier 74
4.3 S3D layout of the receiver front-end 75
4.4 Performance of the S3D receiver front-end 76
4.5 Summary 79

5 Inductors in a S3D process 81
5.1 Inductor Topologies for S3D Integration 81
5.2 Impact of S3D Integration on Planar Inductors 84
5.3 Impact of shields 87
5.4 Multiple-Metal S3D Inductors 90
5.5 Placement of bottom-tier blocks under top-tier inductors 93
5.5.1 Analog Blocks 93
5.5.2 Digital Blocks 98
5.6 Summary 103

6 Heterogeneous S3D Integration: Ge over Si **105**
6.1 Current Status of Heterogeneous S3D Integration 105
6.2 In-house Ge transistors 107
6.3 Short-channel Ge transistors 108
6.4 Applications Ge-Over-Si S3D Integration 113
6.4.1 Track-and-Hold Circuit 113
6.4.2 Digital Cells 115
6.5 Back-gate control of top-tier FETs 116
6.6 Summary 118

7 Conclusions and Future Work **121**

Bibliography **125**

List of Figures

1.1 Current trends in the semiconductor industry (adopted from [1]). 3
1.2 Example of 3D integration with 3 stacked layers 3
1.3 Examples of 3D ICs (a) bond-wire stacking (b) TSV based ICs (c) S3D integration 4
1.4 Interconnected devices for Internet Of Things (IoTs) applications 7
1.5 Example of heterogeneous integration 8

2.1 Via-first TSV consisting of polysilicon rings 14
2.2 (a) Wafer thinning (b) patterning of through-silicon holes (c) deposition of liner and its etching away from the M1 landing pad (d) TSV filling and planarization 15
2.3 (a) Wafer thinning (b) patterning of through-silicon holes (c) deposition of liner and its etching away from the M1 landing pad (d) TSV filling and planarization 16
2.4 TSV-reveal steps: (a) Wafer grinding (b) selective Si etch (c) deposition of passivation layer to protect the Si substrate from copper contamination (d) TSV CMP to reveal the TSV 18
2.5 (a) Build-up of compressive stress in Cu during heating (b) Cu pumping to relax compressive stress (c) impact of TSV-induced stress in Si on neighboring FETs 20
2.6 (a) Die-to-Die (D2D) bonding (b) Wafer-to-Wafer (W2W) bonding (c) Die-to-Wafer (D2W) bonding 22
2.7 High density of TSVs, enabled by the use of a RDL. 23
2.8 CMOS over CMOS S3D integration process steps: (a) Bottom tier with tungsten metal lines (b) Deposition of the ILD and transfer of a thin crystalline Si layer (c) Top tier FEOL and patterning of MIVs (d) Top tier BEOL 26
2.9 Structure of a n-type junctionless FET 28
2.10 S3D integration of CNTFETs over Si FETs. The inset shows the top-view of a CNTFET 30
2.11 Formation of a high quality, thin and crystalline Ge layer over a Si wafer, as proposed in [2] 32

2.12 (a) Classic packaging solution for high power applications (b) New packaging approach for mobile applications 35

3.1 (a) Custom IC design flow (b) Digital IC design flow 38
3.2 Process stack of the in-house KTH-S3D process 40
3.3 (a) Automatically generated layout for a pFET with L= 1 μm and W= 2 μm (b) Manually generated layout of a pFET with L= 100 nm and W= 2 μm (c) Manually generated layout of a nano-wire pFET with L= 100 nm. The device consists of 10 nano-wires, each 20 nm wide. . . 42
3.4 Layout of a KTH-S3D diode with W= 2μm. 43
3.5 UTSOI2.1 model and measurement results for (a)-(b) a Si nFET in the bottom-tier, (c)-(d) a Si pFET in the bottom tier and (e)-(f) a Ge pFET in the top tier. The plots in the left show the transfer characteristics and the ones in the right the output characteristics. All devices have equal dimensions, W=L= 2 μm. 45
3.6 Level-1 diode models versus measurement results for the I-V charcteristics of two diodes fabricated by the in-house process. 46
3.7 MIVs parallel to top-tier contacts (Cont1), poly-silicon (Gate1) and active regions (Active1). The minimum distance between these layers is d= 1 μm. 47
3.8 (a) TEM image of a via processed with the in-house process (b) Allowed stacking of non-successive vias (Via2 over MIV in this case). 47
3.9 Layout structures to extract the parasitics of: (a) a single MIV (b) the coupling between two neighboring MIVs (c) the coupling between a MIV and a top-tier contact (Contact1) (d) the coupling between a M0 line and a top tier active (Active1) region (e) the capacitance between a M0 and a M1 line. 49
3.10 Schematic of a 3-input NAND-based D flip-flop. 51
3.11 Layout view of the KTH-S3D DFFSR cell in (a) the bottom tier and (b) the top tier. (c) 3D view of the KTH-S3D DFFSR. The NAND based module is shown in the inset. 53
3.12 Layout view of the ideal KTH-S3D DFFSR cell (with allowed stacking of successive vias) in (a) the bottom tier and (b) the top tier. 54
3.13 Time-domain operation of a KTH-S3D DFFSR cell 55
3.14 Process stack of the S3D PPDK. 56
3.15 3-D layout view of a S3D PPDK MIM capacitor. 59
3.16 Calibrated UTSOI2.1 model and TCAD simulations for the transistor (a) transfer and (b) output characteristics. The plots refer to a n-type FET with L= 250 nm and W= 1 μm 60
3.17 Input referred noise voltage, V_{n-in} of a nFET in the top and bottom tier. Both devices have the same dimensions and biasing: L= 150 nm, W= 10 μm, I_D=200 μA and V_{DS}= 0.5 V 61

3.18 Layout structures to extract the parasitics of: (a) a single Met5 -> Met1_top via stack (b) the coupling between two neighboring Met5 -> Met1_top via stacks (c) the coupling between a Met1_top and a Met5 line (d) the coupling between a MIV and Cont_top (e) the coupling between a Met5 line and an Active_top region. 63

4.1 Id versus Vgs for (a) $V_{DS} = 0.1$ V and (b) $V_{DS} = 1.8$ V. (c) Transconductance-gm versus Vgs for $V_{DS} = 0.1$ V and 1.8 V. (d) Structures of the top-tier FDSOI (top structure) and bottom tier bulk (bottom structure) FETs. Note that for visual purposes, the dimensions of the various device regions are not scaled proportionally 67
4.2 (a) Block-level representation of the considered receiver front-end, based on the LMV cell. (b) Schematic of the LMV cell, adopted from [3]. . . . 69
4.3 (a) CS LNA with simple resistive input termination. (b) CS LNA with a passive impedance transformation network. 70
4.4 Quadrature generation through C0 and M1. 72
4.5 Mixer sub-block in the LMV cell. 73
4.6 (a) Schematic of the TIA. (b) Schematic of the CMFB circuit. 74
4.7 (a) Layout of the designed S3D receiver in (a) the top and (b) the bottom tier. The location of MIVs are highlighted with yellow circles 75
4.8 3-D view of the S3D receiver front-end. Top tier blocks are shown with opaque colors and bottom tier with transparent. 76
4.9 (a) S_{11} , (b) Gain of the receiver chain, (c) Phase Noise and (d) double-sideband NF versus frequency. 78

5.1 3D solenoid implemented with TSVs 82
5.2 (a) Structure of the S3D solenoid. (b) Inductance and (c) Quality factor of the S3D solenoid. 83
5.3 (a) Asymmetric and (b) Symmetric inductor topology. 84
5.4 (a) Lumped network employed for inductors. (b) Simplification of the inductor lumped network, assuming Port 2 is grounded. 85
5.5 (a) Inductance and (b) quality factor for an inductor built in the base process and the same topology transferred to the S3D PPDK. 87
5.6 Inductor with a PGS. The shield contains metal stripes, each 5 μm wide. The spacing between adjacent stripes is 1 μm. 88
5.7 (a) Considered structure to validate the effectiveness of the PGS in suppressing substrate noise (b) Isolation between the inductor and the substrate, quantified by the S_{13} parameter. 89
5.8 (a) Base process inductor (b) S3D inductor formed by shunting together the top tier metals MThick ->M3_top. 90
5.9 (a) Multi-metal S3D inductor. The red rectangles are the vias that shunt together the multiple metal layers (b) 3D view of the multi-metal S3D inductor 91
5.10 (a) Inductance and (b) quality factor of the multi-metal S3D inductor. . 92

5.11 (a) Resistance and (b) low-frequency inductance of the multi-metal S3D inductor versus frequency. 93
5.12 Considered structure to study the impact of the location of bottom tier blocks under the inductor on the inter-tier EM isolation. 94
5.13 Inter-tier EM isolation when the resistive network is placed under the inductor at position (a) A (b) B and (c) D. (d) Resistance between ports P3 and P5 before and after placing the resistive network under the inductor center (location A). 95
5.14 (a) Studied topology to emulate larger blocks and (b) inter-tier EM isolation. 96
5.15 4x4 resistor array under a top-tier inductor 97
5.16 S3D process stack used to study the EM coupling between bottom-tier digital blocks and top-tier inductors. The process data of the 20 nm FD-SOI process are provided on an "as is" basis through ASCENT, funded by the EU H2020 Infrastructure Programme (H2020-INFRAIA-2014-2015) under Grant Agreement 654384 99
5.17 (a) Studied topology to identify the most optimal PGS layer (b) Impact of the PGS on inductor's Q_{max} and wire's *Cap.* (c) Impact of PGS on f_{SR} and L_{diff}. 100
5.18 (a) Top tier inductor and bottom tier power ring.(b) Proposed broken power ring to suppress eddy currents. (c) Inter-tier EM isolation between the inductor and the power ring, and quality factor for the inductor. 101
5.19 A 3-bit adder is placed in the bottom tier under a top-tier inductor. Details of the adder implementation are shown in the inset. 102

6.1 (a)Top-view and (b) cross-section of a measured in-house Ge-pFET with L= 800 nm 107
6.2 Measurement results for an in-house Ge transistor and corresponding TCAD simulations. I_D versus V_{GS} in (a) linear and (b) semi-logarithmic axis and (c) I_D versus V_{DS}. 109
6.3 Net concentration of activated dopants in the channel 110
6.4 Dependence on DIBL and $G_{m,max}$ on the Ge layer's thickness, T_{Ge}. . . 110
6.5 (a) Transfer characteristics of Ge and Si pFETs. TCAD and spice simulations for (b) Ge pFETs and (c) Si pFETs. (d) Output characteristics of the Ge and Si pFETs. (e) Transfer and (f) output characteristics of the Si nFET device. 111
6.6 Intrinsic gain A_v of a Si nFET, a Si pFET and a Ge pFET. The dimensions of all three devices were L = 150 nm and W = 1 μm. 113
6.7 CMOS TH circuit with C_{hold} = 500 fF. 114
6.8 (a) R_{TH} for a Ge and a Si pFET (b) BW_T for a Ge and Si pFET. . . . 115

6.9 (a) Deposition of ILD over the last bottom tier metal, MB. (b) Etching of the ILD to form back-gate contact-hole. (c) Metal Filling of the contact-hole. (d) Etching and CMP of the metal from all the areas outside the contact-hole. (e) Formation of a 10 nm thick oxide, that will form the BOX of the top-tier FETs. (f) Processing of the top-tier FETs. 118
6.10 Transfer characteristics of a 150 nm Ge pFET for various back gate voltages V_{BG}. The back-gate biasing has been enabled through the devised process flow. 119

List of Tables

2.1 Heterogeneous S3D integration: Comparison between Ge and InGaAs top-tier FETs. 33

3.1 Characteristics of the in-house KTH-S3D process stack 41
3.2 Sheet resistance for the KTH-S3D resistors 43
3.3 Minimum width and spacing for the layers in the KTH-S3D PDK 46
3.4 Comparison between the parasitic capacitance obtained with the KTH-S3D parasitic extraction flow and with Momentum EM simulations (by Keysight®) 50
3.5 Truth Table for the KTH-S3D DFFSR cell 52
3.6 Performance comparison between the KTH-S3D inverter cell and its 2-D in-house implementation 54
3.7 Characteristics of the S3D PPDK considered process stack 58
3.8 Basic design rules for S3D-150nm PPDK layers 61
3.9 Comparison between the parasitic capacitance obtained with the S3D-150 nm PPDK's parasitic extraction flow and Momentum EM simulations (by Keysight®) 62

4.1 Simulated performance of the S3D receiver front-end 77

5.1 Impact of the PGS on a S3D inductor 88
5.2 Inter-tier EM isolation 97
5.3 Characteristics of the passive devices before and after stacking 98
5.4 Analysis of the placement of a 3-bit adder under a top-tier inductor . . 102

6.1 Comparison between the Ge pFET, the Si pFET and the Si nFET . . . 112
6.2 Performance comparison between a TH with a Ge and with a Si pFET . 116
6.3 Characteristics of the passive devices before and after stacking 117

List of Acronyms

ABB	Adaptive Back-gate Biasing
ADC	Analog-to-Digital Converter
AMS	Analog/Mixed-Signal
AR	Aspect Ratio
ASIC	Application Specific Integrated Circuit
BEOL	Back-End-Of-Line
BOX	Buried Oxide
BSIM	Berkley Short-channel Independent gate field effect transistors Model
BSIM-IMG	BSIM Independent Multiple Gates
CMFB	Common Mode Feedback
CMP	Chemical Mechanical Polishing
CNTs	Carbon Nano-Tubes
CNTFETs	Carbon Nano-Tube Field Effect Transistors
CS	Common Source
CTE	Coefficient of Thermal-Expansion
CVD	Chemical Vapor Deposition
D2D	Die-To-Die bonding
D2W	Die-To-Wafer bonding
DDR3	Double Data Rate Type Three
DFF	D flip flop
DIBL	Drain Induced Barrier Lowering
DRAM	Dynamic Random-Access Memory
DRC	Design Rule Checks

EBL	Electron Beam Lithography
EDA	Electronic Design Automation
EM	Electromagnetic
EOT	Equivalent Oxide Thickness
EUV	Extreme Ultra-Violet
FET	Field-Effect Transistor
FD-SOI	Fully Depleted - Silicon On Insulator
FEOL	Front-End-Of-Line
HVM	High Volume Manufacturing
IC	Integrated Circuit
IEOL	Intermediate End Of Line
IF	Intermediate Frequency
IO	Input/Output
KGD	Known Good Die
KOZ	Keep Out Zone
LED	Light Emitting Diodes
LER	Line Edge Roughness
LMV	Low noise amplifier - Mixer - Voltage Controlled Oscillator
LNA	Low Noise Amplifier
LVS	Layout Versus Schematic
M3D	Monolithic 3-D
MIM	Metal-Insulator-Metal
MIV	Monolithic Inter-tier Via
MOM	Metal-Oxide-Metal
MOSCAP	Metal-Oxide-Semiconductor Capacitor
NF	Noise Figure
PCells	Parameterized Cells
PDK	Process Design Kit
PDN	Pull Down Network
PGS	Patterned Ground Shield
PiN diodes	p+/intrinsic Si/n+ diode
PITN	Passive Impedance Transformation Network
PoP	Package-On-Package

PPDK	Predictive PDK
PR	Power Ring
PUN	Pull Up Network
RF	Radio Frequency
RAM	Random Access Memory
RDF	Random Dopant Fluctuation
RSD	Raised Source/Drain
RTA	Rapid Thermal Annealing
RDL	Redistribution Layer
ReRAMs	Resistive Random Access Memories
S3D	Sequential 3-D
SADP	Self-Aligned Double Patterning
SCE	Short Channel Effects
SiP	System in Package
SoC	System on Chip
SPE	Solid-Phase Epitaxy
SR	Slew Rate
SRB	Strain Relaxed Buffer
SS	Sub-threshold Slope
TCAD	Technology Computer Aided Design
THC	Track and Hold Circuits
UTBB	Ultra Thin Body and BOX
VCO	Voltage Controlled Oscillator
VGA	Variable Gain Amplifier
W2W	Wafer-To-Wafer bonding
WPAN	Wireless Personal Area Networks

Chapter 1

Introduction

1.1 New Directions for Moore's Law

The observations and predictions made by Gordon Moore about the density of integrated components in a chip [4], which later came to be known as "Moore's Law", have driven the semiconductor industry since its first steps. Gains in the integration density of chips have been typically achieved through the continuous dimensional scaling of the integrated components, a path that became known as "More Moore". Although Moore's predictions were based on financial grounds and in particular on reducing the cost per components of integrated circuits (ICs), dimensional scaling has led also to a reduction in gate delays, and for the case of constant field scaling, in the total power consumption as well [5].

However, the continuous dimensional scaling has brought many challenges. As transistor lengths moved to the sub-100 nm regime, more complex processing has emerged to improve transistor performance and at the same time guarantee high production yields. For instance, strained silicon transistors were used by Intel® in their 90 nm node to improve carrier mobility [6]. High-k gate dielectrics and metal gates have been employed in Intel's 45 nm to reduce gate leakage [7]. Air-gapped interconnects were introduced in Intel's 14 nm node to reduce the coupling capacitance between adjacent metal lines, as well as self-aligned double patterning (SADP) for patterning critical layers [8]. SADP along with multiple patterning have been employed to pattern features like transistor gates, fins or metal wires, whose dimensions are smaller than the resolution of the lithography tools. They involve multiple cycles of lithography exposure and etching to pattern a single layer. In multiple patterning, all features of a layer are split into groups (colors) and each group is patterned by a different mask. Color-based patterning imposes additional design rules, further complicating both the physical implementation and physical verification flows. For instance, local color-density deviations can result in coupling capacitance variations, which in turn may require timing and power re-characterization [9].

Extreme Ultra-Violet (EUV) lithography could mitigate these issues since it offers higher resolution compared to today's 193 nm immersion lithography. Additionally, it negates the need for multiple-exposure patterning. However, current EUV sources lack enough power (larger than 250 W) for High Volume Manufacturing (HVM), while there is still room for improvements in the sources' lifetime [10]. Another critical issue that impedes the commercial adoption of EUV lithography is the absence of pellicles that could protect the mask from contamination during the exposure stage, which at the same time can withstand the source power levels required for HVM [11].

Apart from the increased process complexity, another issue that rises with the continuous scaling is the increased resistance of the routing resources. At each new node, the metal pitch is also scaled to allow an increase in the integration density. This causes the metal thickness to drop as well, which in turn, accentuates the impact of the highly resistive diffusion barrier material (i.e. TaN) and causes an exponential increase in the wires resistance [12]. Another important issue with scaled interconnects lies with the pronounced reliability issues, caused by the multiple processing during their patterning. Each of these steps induces variation in a design: Chemical-Mechanical Polishing (CMP), etching, overlay error for multiple patterning and for the case of SADP, core and spacer variations. All these variation sources have a negative impact on the performance and reliability of a design [13]. The phenomenon is further exacerbated by the small dimensions of the routing wires. All in all, the interconnect bottleneck has posed a major roadblock in the "More Moore" path.

As transistor dimensions kept dropping, the gate control over the channel has dropped, leading to severe short channel and Drain-Induced Barrier Lowering (DIBL) effects. As a result, FinFETs have been adopted for the Front-End-Of-Line (FEOL) in the Intel's 22 nm node [14], thanks to their superior gate control over the channel, as compared to planar transistors. Although this adoption has been beneficial for digital designs, it has placed an additional burden to Analog and Mixed-Signal (AMS) designers who have to cope also with an increased number of design rules with rather limited assistance from Electronic Design Automation (EDA) tools. FinFETs suffer from high gate, source and drain contact resistances because of the limited silicidation of the active regions [12]. Furthermore, the 3D structure of FinFETs cause higher device capacitances with respect to planar transistors, which in turn limit significantly the performance of these devices at high frequencies. Therefore, advanced process nodes hinder the design of high performance radio frequency (RF) and AMS circuits.

The aforementioned issues (i.e. complex and more expensive processing, interconnect bottleneck, performance degradation for RF/AMS circuits) have led to rising concerns that the dimensional scaling trend is reaching a dead-end. Recently, GlobalFoundries® has decided to drop out from the pursuit of the 7 nm node claiming financial reasons and reduced customer interest. Instead, the company will focus on adding further capabilities to their existing FinFET and FD-SOI processes. Most of the semiconductor industry appears to re-direct their business

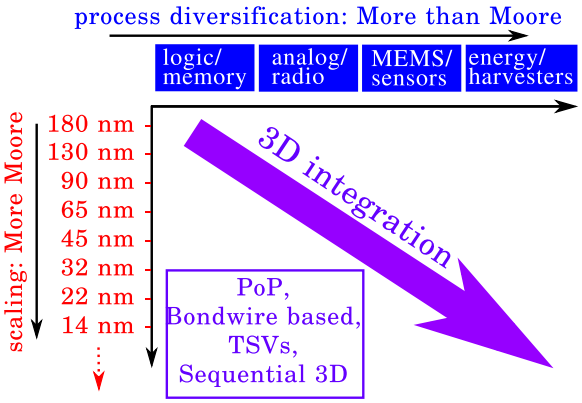


Figure 1.1: Current trends in the semiconductor industry (adopted from [1]).

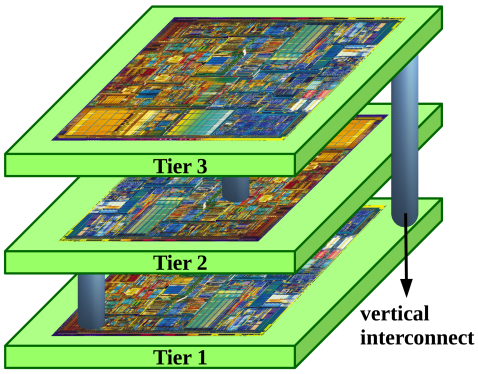


Figure 1.2: Example of 3D integration with 3 stacked layers

model from the traditional supply-driven approach to emphasizing diversification with optimized processes for specific applications (memories, logic, RF, sensors, etc), a path that became known as "More than Moore" [15] and which is shown in Fig. 1.1. This path employs new materials, as well as innovations in both the semiconductor processing and packaging. It also calls for a close synergy between process engineers and circuit designers to obtain optimal integration solutions.

3D integration, with the vertical stacking of tiers of devices, as it is shown in Fig. 1.2, could bridge the "More Moore" and the "More than Moore" paths. Vertical stacking increases the integration density by adding more functionality to an IC with no penalty on area. Moreover, the small distance between the stacked tiers enables short-distance and low-parasitics vertical interconnects, which in turn can reduce wirelength. Denser vertical interconnects result in more wirelength

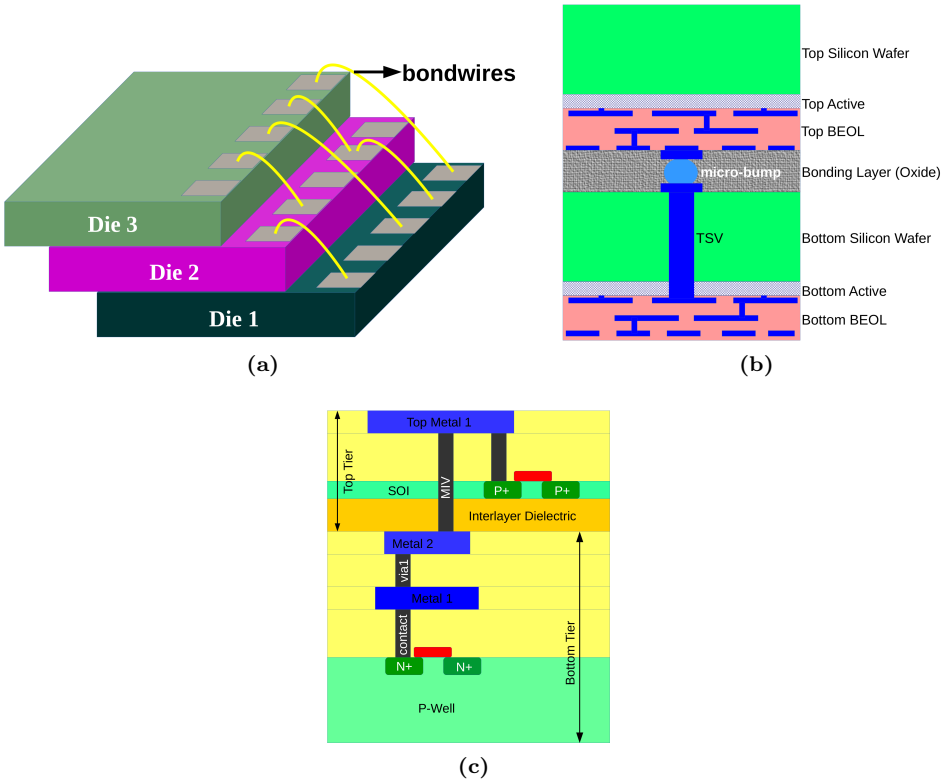


Figure 1.3: Examples of 3D ICs (a) bond-wire stacking (b) TSV based ICs (c) S3D integration

reduction and consequently in performance and power improvements, alleviating thus the interconnect bottleneck in advanced nodes. 3D integration enables also heterogeneous integration by stacking different types of device tiers, with each tier being optimized for a specific function. Thus, 3D integration could lead to the development of power and area efficient System-On-Chip (SoC) applications.

1.2 3D Integration Technologies

The current 3D integration technologies can be divided into two types: package-on-package stacking (PoP) [16] and single package. For the single-package approach, three technologies have been developed: (a) bond-wire based die stacking, (b) Through-Silicon Vias (TSV) based 3D ICs and (c) Sequential 3D (S3D) integration (known also as Monolithic 3D - M3D).

Bond-wire based 3D die-stacking, as shown in Fig. 1.3(a), refers to the vertical

stacking of dies, such as processors, DRAMs and NAND flash memories, that are interconnected with bond-wires and form a System-In-Package (SiP). The stacking of sixteen dies, has been reported in [17]. Although this stacking approach is widely used in mobile devices for area-efficient solutions, it suffers from the lowest density of vertical (inter-die) interconnects among the other alternatives, which in turn leads to trivial, if any, wirelength improvements. Thus, when it comes to performance, the benefits of bond-wire based 3D stacking are insignificant. Instead, this integration approach has mainly financial benefits, thanks to the possibilities of cost reductions, enabled by the availability of multiple supply sources [18].

On the other hand, TSV-based 3D ICs feature a higher density of vertical interconnects than wire-bond based die stacking. Various implementation techniques have been proposed for TSV-based 3D ICs [19, 20]. Typically, they start with at least two patterned dies, with TSVs being etched and filled in at either one or both of them (back-to-back stacking). The two dies are then stacked and bonded together through micro-bumps, as shown in Fig. 1.3(b). TSV-based 3D ICs have already made their way to commercial products. For instance, a TSV-based 8 Gb 3D DDR3 (Double Data Rate Type 3) DRAM (Dynamic Random Access Memory) has been presented in [21], consisting of four stacked 2 Gb DRAM chips and interconnected through TSVs, each with a 30 μm diameter. The TSV pitch equals 80μm and each chip is fabricated in a 50 nm DRAM process resulting in a total footprint of 10.9 x 9 mm. The 8 Gb DRAM demonstrates power gains and improvement of the Input/Output (IO) speed compared to wire-bond based die-stacking implementations. Smaller TSVs with typical values of 5-10 μm for their diameters and 50-100 μm for their heights have been demonstrated for increasing the density of vertical interconnects [18]. Even with these improvements however, TSV-based approaches fall short from the required vertical interconnect pitch (smaller than 5μm) for future 3D SoCs [18]. Further reduction of the TSV dimensions will not yield higher vertical interconnect densities, which are limited by the pitch of the microbumps [18].

S3D integration, also described as monolithic 3D (M3D), offers the highest density of vertical interconnects. Unlike the other two options, S3D integration does not rely on the bonding of two pre-patterned device tiers; instead a second device tier is processed over a pre-patterned one. An example of a S3D process stack is shown in Fig. 1.3(c). More specifically, a thin active layer is transferred (typically through means of wafer bonding) over a patterned wafer and processed to form the top-tier's devices (Front-End-Of-Line, FEOL) [22–24]. The application of wafer bonding negates the use of microbumps, with their large area overhead. Following the top-tier's FEOL, inter-tier vias (Monolithic Inter-tier Vias, MIVs) are etched and filled to establish connectivity between the two tiers. Due to S3D's sequential processing, the alignment precision between stacked tiers depends solely on the lithography stepper. Thus, MIVs can be processed like any other metal-to-metal via, opening the path to ultra high density vertical interconnects [25]. Assuming that all tiers in a S3D implementation are processed in the 14 nm FinFET process from [8], the MIV pitch could be as low as 70 nm. Thanks to the small

MIV pitch, S3D designs are expected to offer smaller footprints and shorter wirelengths than TSV-based ones, with consequent improvements in speed and power [26]. Another area that could benefit from S3D integration is brain-inspired architectures for power efficient computing. S3D ICs, with their vertical interconnects, could better emulate brain-inspired architectures with thousands of synapses per neuron, as opposed to conventional planar implementations [27], becoming thus a key-enabling technology for future computing. Despite the fact that S3D processing technologies are still under development, an insight into potential applications of this technology, design methodologies and EDA tools is required. Based on the concepts of design-technology co-optimization, a close collaboration between designers, EDA tool developers and process engineers is necessary to ensure high production yields, preferably without sacrifices in "designability" and also to allow faster time to market [28].

So far, research on S3D design methodologies has led to three main approaches for the partitioning scheme between the stacked tiers: block-level, gate-level and transistor-level partitioning. In block-level partitioning, the functional blocks of a digital design are split between two or more tiers [29]. Clearly, this approach does not take full potential of the smaller MIV-pitch and consequently, it results in low vertical interconnect densities. Nonetheless, it can better cope with performance variations between the stacked tiers [29]. On the other hand, in gate-level partitioning, digital cells are placed one on top the other [30–33] leading to denser vertical interconnects. The highest density of vertical interconnects is achieved through the transistor-level partitioning scheme [34, 35]. In this approach, each cell of a digital library is split between two tiers, with the NMOS transistors placed in one and the PMOS in another. A hybrid S3D floorplanner that combines the benefits of these three approaches and specifies the optimal partitioning option for each block (no S3D partitioning, transistor-level and gate-level partitioning) has been presented in [36]. All these partitioning schemes aim at equal footprints between stacked tiers. Alternatively, limiting 3D stacking to only long nets would trade-off area reduction with improvements in both speed and power [37]. Memories, for instance SRAM designs, is another application that could benefit from the small pitch of MIVs [38–40]. S3D integration, however, does not limit to CMOS over CMOS applications only. The use of materials other than silicon for top tier devices enables a high-degree of heterogeneous integration. For instance, in [41], the integration of CMOS devices with tiers of resistive RAMs (RRAMs) and Carbon Nano-Tubes (CNTs) Field Effect Transistors (FETs) has led to very tight integration between logic and memory. Ge or III-V materials with their superior carrier mobilities can be also employed as channel materials for the top tier [42, 43].

1.3 Motivation

Recently, a major paradigm shift has been observed from processing data at a central node to distributing data processing among various interconnected devices,

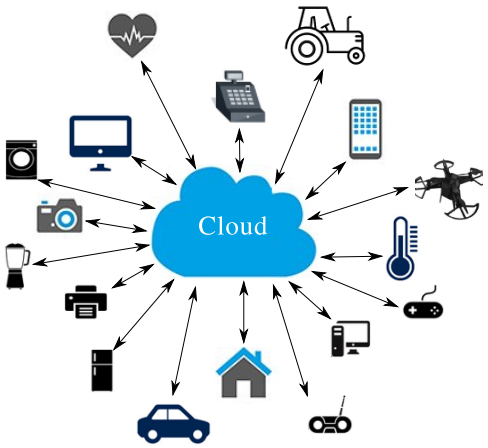


Figure 1.4: Interconnected devices for Internet Of Things (IoTs) applications

allowing thus Internet-Of-Things (IoTs) applications [27], as shown conceptually in Fig. 1.4. The need for distributed processing rises from the large count of such interconnected devices, which cause severe communication traffic between them and the central node, which in turn affects the performance [27]. Furthermore, central processing puts safety into question, owing to the increased probabilities of communication disruptions [27]. For the distributed processing paradigm, apart from logic and memory, radio circuits are also required to communicate data, in addition to AMS circuits for read-out operations. Co-integration of all these circuits with sensors would be also desirable to improve signal-to-noise performance. S3D integration, with its large density of vertical interconnects and its inherent heterogeneous features could prove beneficial for IoTs applications. A S3D IC could consist of various device tiers, each one optimized for a specific function, as shown in Fig. 1.5. Thus, S3D design methodologies for applications other than logic and memories need to be investigated and developed. Towards this, a S3D design platform for custom integrated circuits (ICs) is essential. The S3D design platforms proposed so far, are restricted only to digital circuits [44, 45]. Furthermore, to take full advantage of the heterogeneous integration capabilities of S3D technology, the impact of Ge or III-V materials on the performance of S3D circuits needs also to be explored. For instance, the superior mobility of III-V materials or Ge could allow the downsizing of the top-tier devices, leading to further area gains.

1.4 Research Objectives

The main goal of this thesis is to demonstrate the potential of S3D integration for future circuits and systems. More specifically, this thesis aims to investigate design

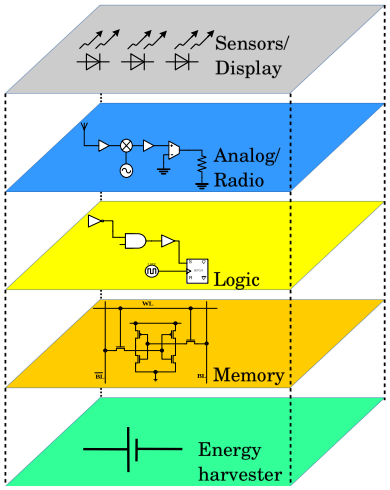


Figure 1.5: Example of heterogeneous integration

methodologies for high performance and area efficient S3D ICs. Towards this, and based on the motivations described previously, a set of research objectives has been set:

- **Objective 1:** Develop a S3D design methodology and flow for customs ICs.
- **Objective 2:** Investigate circuits and applications that could benefit from S3D integration.
- **Objective 3:** Propose solutions to counteract the impact of S3D processing on the circuit performance.
- **Objective 4:** Study the impact of materials other than silicon in the top tier, on the performance of S3D circuits and systems.

1.5 Research Contributions

In relation to the objectives stated above, the research contributions of the present thesis are the following:

- **Contribution 1:** A S3D design platform has been developed through means of a Process Design Kit (PDK). The PDK is compatible with commercial CAD and EDA tools to facilitate migration to S3D technologies. Accurate transistor models are included for both bottom and top tier devices. To enable physical verification, sets of Design Rule Checks (DRC) and Layout Versus Schematic (LVS)

checks have been also included in the PDK. A parasitic extraction flow has been developed to analyze the inter-tier coupling. This S3D PDK is used to explore circuits that could benefit from S3D integration, develop design techniques and analyze the S3D circuits performance (Papers I and II). An additional S3D PDK has been also developed to facilitate the development of the KTH in-house S3D process.

• **Contribution 2:** Considering the main characteristics of S3D integration technology, the potential of S3D RF/AMS circuits and systems has been identified. A frequency-based partition scheme has been devised, in which high frequency blocks were placed in the top tier and low-frequency ones in the bottom tier. As a proof of concept, a S3D receiver front-end for Wireless Personal Area Network (WPAN) applications has been designed and simulated in the developed S3D PDK. (Paper II)

• **Contribution 3:** To further advance the previous study on S3D RF/AMS circuits and systems, the placement of inductors in a S3D technology has been explored (Paper III). It has been found that a planar coil in a top tier thick metal is the most optimal configuration for inductors in a S3D process. To further improve the area efficiency of S3D RF/AMS circuits and systems, guidelines have been proposed for the effective placement of bottom tier blocks underneath top tier inductors to ensure high electromagnetic isolation between them (Paper III and IV).

• **Contribution 4:** The design prospects of a S3D integration technology with Ge in the top-tier have been investigated. Circuits that do not operate continuously in the saturation region, such as track-and-hold and digital cells have been shown to benefit the most from such a Ge-over-Si S3D integration technology.

1.6 List of Publications

• [**Paper I**] **P. Chaourani**, P.-E. Hellström, G. Malm, S. Rodriguez, A. Rusu "Towards Monolithic 3D Integration: A Design Flow," in CDNLive2016, Cadence User Conference EMEA, Munich, Germany, May 2-4, 2016, available: "http://kth.diva-portal.org/smash/record.jsf?pid=diva2%3A1082029&dswid=8595".

Author's Contribution: The author developed the parameterized cells, physical verification rules, parasitic extraction flow and the device models based on TCAD simulations. He also wrote the manuscript.

• [**Paper II**] **P. Chaourani**, P. Hellström, S. Rodriguez, R. Onet and A. Rusu, "Enabling area efficient RF ICs through monolithic 3D integration," Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017, Lausanne, pp.

610-613.

Author's Contribution: The author performed the design-space exploration and devised the frequency-based partition scheme. He also designed the S3D receiver front-end and wrote the manuscript.

• **[Paper III] P. Chaourani**, S. Rodriguez, P. Hellström and A. Rusu, "Inductors in a Monolithic 3D Process: Performance Analysis and Design Guidelines," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, no. 2, pp. 468-480, Feb. 2019.

Author's Contribution: The author analyzed the performance of S3D inductors and devised basic rules to allow the placement of bottom tier blocks under top-tier inductors in a S3D process. He also wrote the manuscript.

• **[Paper IV] P. Chaourani**, D. Stathis, S. Rodriguez, P. Hellström and A. Rusu, "A Study on Monolithic 3D RF/AMS ICs: Placing Digital Blocks Under Inductors," 2018 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), Burlingame, CA, USA, 2018, pp. 1-3.

Author's Contribution: The author developed the design technology co-optimization flow to enable digital blocks under top-tier inductors and wrote the manuscript.

• **[Paper V] P. Chaourani**, P. Hellström, S. Rodriguez and A. Rusu, "A study on heterogeneous S3D integration: The impact of Ge transistors on S3D circuit design," *in manuscript*

Author's Contribution: The author has calibrated the TCAD models to match in-house measurement results, and he has used TCAD to predict the performance of short-channel Ge transistors. He has also explored circuits that could benefit from a Ge-over-Si S3D integration and he has written the manuscript.

1.7 Thesis Organization

The thesis is organized in seven chapters as follows:

• **Chapter 1** provides a brief introduction to the topic of this thesis along with the state-of-the-art approaches in the field. The main motivations behind this work, as well as its main contributions are also presented.

• **Chapter 2** introduces the main features of TSV-based and S3D ICs and a comparative study between them is carried out. It then discusses the processing

issues related to S3D integration and solutions to overcome them.

• **Chapter 3** describes the development of a S3D predictive PDK, built around a conventional 2D bulk process (S3D PPDK). A detailed description of the device models and the establishment of a parasitic extraction flow is given. A PDK for KTH's in house Ge-based S3D process (KTH-S3D PDK) has been also developed. The motivations behind the employed design rules are discussed in relation to the in-house process capabilities.

• **Chapter 4** investigates the possible applications of S3D integration and it identifies the potential of S3D RF/AMS circuits. As a proof of concept, the design of a receiver front-end in the S3D PPDK is described. The simulation results are then compared against the performance of the conventional 2D implementation of the same receiver front-end topology.

• **Chapter 5** focuses on S3D inductors. It starts with a study of the most optimal inductor topologies in a S3D process. The impact of S3D integration on the inductance and quality factors of the inductors is analyzed and shield structures are investigated. Next, the potential of placing bottom tier RF/AMS blocks under top tier inductors is identified and design guidelines are proposed to handle this placement with best performance trade-offs. A methodology for the placement of digital blocks underneath top-tier inductors is finally presented.

• **Chapter 6** studies the impact, performance-wise, of Ge as a channel material for the top tier devices. In addition, it investigates circuit topologies that can benefit from the use of Ge top tier devices.

• **Chapter 7** concludes the thesis and suggests directions for future research on the field.

Chapter 2

Insights into the processing of 3D ICs

TSV-based and sequential 3D ICs offer the highest density of 3D interconnects among other 3D integration options and consequently the potential for higher performance gains. They differ in their processing, and yield different trade-offs in terms of performance and manufacturability. This chapter describes the main features of TSV-based and sequential 3D ICs from a processing point of view and their impact on design methodologies, as described by the most recent research trends and state-of-the-art literature in the field.

2.1 TSV processing options

TSVs are a key enabling technology for 3D ICs allowing connections between the front- and back-side of a chip with relatively high densities. These connections are essential for the establishment of 3D integration. Based on the process used for their fabrication, TSV processing can be divided into three groups:

- via-first TSVs
- via-last TSVs
- via-middle TSVs

2.1.1 Via-first TSVs

Via-first TSVs, as their name implies, are fabricated at the very beginning of the wafer processing, before any of the FEOL steps. Thus, they are also known as pre-process TSVs. The precedence of the TSV formation over the FEOL processing

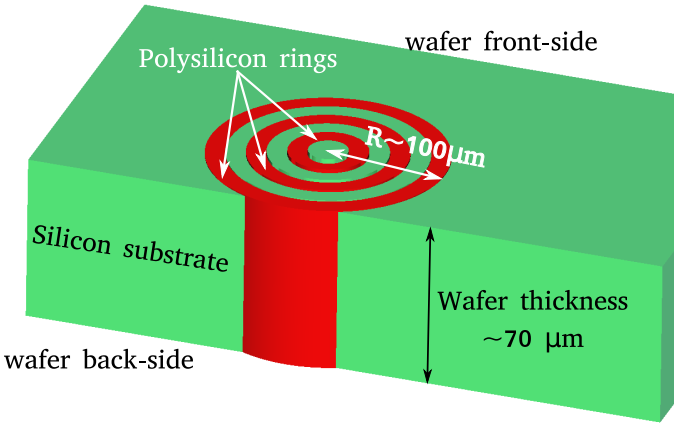


Figure 2.1: Via-first TSV consisting of polysilicon rings

limits the possible materials that can be used for their filling to only heavily doped poly-silicon. To counter the relatively large resistivity of poly-silicon, the diameter of via-first TSVs needs to be enlarged, limiting in turn the 3D interconnect densities. Via-first TSVs with diameters of 80-100 μm have been reported in [46]. There, a ring-shaped approach was adopted for their implementation, as it provided the best trade-off between via-resistance, production yield and induced stress (see Fig. 2.1). A 70 μm deep via-first TSV, with four rings, each 6 μm wide, and a total diameter of 100 μm resulted in 60 mΩ/TSV. However, despite its high resistivity, the use of poly-silicon is ideal for high-voltage ICs, since the latter require high termal budget processes [47]. Another popular application for via-first TSVs is the stacking of a MEMS die and an ASIC [18].

2.1.2 Via-last TSVs

Contrary to via-first TSVs, via-last TSVs are formed last, after the processing of the interconnects (Back End Of Line - BEOL), providing a connection path between the back-side of a wafer and its first metal layer (M1) in the front-side. As a result, copper or any other metals can be used as a filling material, allowing substantially lower resistivity than via-first TSVs. However, the thermal budget for the TSV processing must be kept low, so as to avoid any damage to the existing BEOL and FEOL. The processing of via-last TSVs is illustrated in Fig. 2.2 and it can be summarized into 4 steps [20]:

- **Step 1:** A processed wafer with complete BEOL and FEOL is bonded to a handle wafer and thinned, typically through means of wafer grinding.

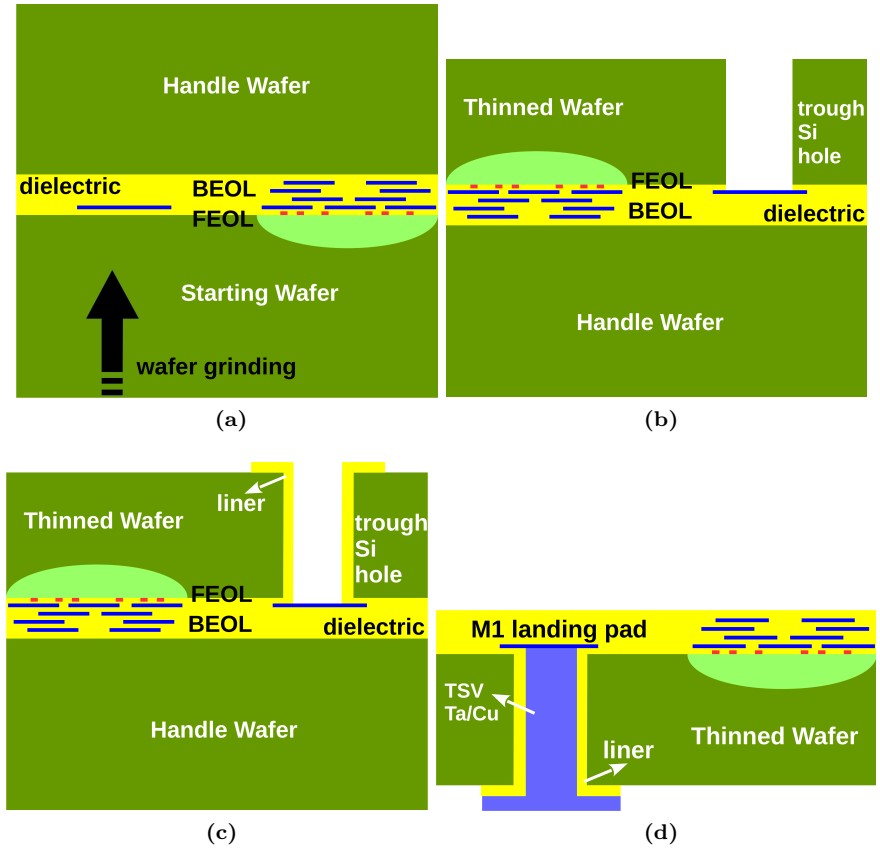


Figure 2.2: (a) Wafer thinning (b) patterning of through-silicon holes (c) deposition of liner and its etching away from the M1 landing pad (d) TSV filling and planarization

- **Step 2:** The through-silicon holes are defined through lithography exposure and deep silicon etching from the back-side of the wafer.
- **Step 3:** Dielectric is deposited in the through-silicon hole to ensure electric isolation between the TSV filling metal and the surrounding silicon substrate. This dielectric layer is known as the TSV liner. The liner then needs to be etched away from the bottom side of the TSV to ensure connection to M1.
- **Step 4:** The through-silicon hole is filled with a metal, typically copper to ensure low-resistivity. To avoid diffusion of copper in the surrounding silicon, the through silicon hole is first coated with a barrier layer, like tantalum (Ta).

One of the main challenges of this TSV implementation approach is the need to selectively remove the liner from the bottom of the through-silicon hole with

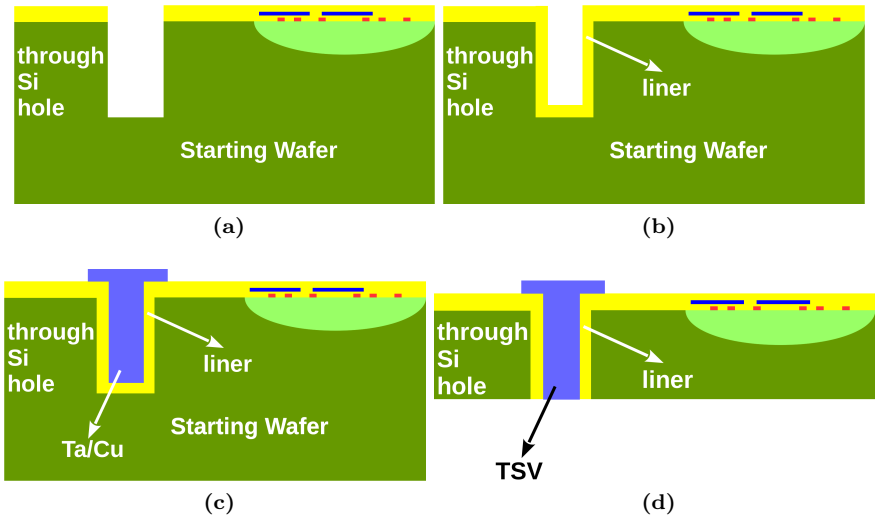


Figure 2.3: (a) Wafer thinning (b) patterning of through-silicon holes (c) deposition of liner and its etching away from the M1 landing pad (d) TSV filling and planarization

minimal impact on the sidewall liner (see Fig. 2.2(c)). In addition, sufficient coating of the back-side wafer is required during TSV filling to minimize the risk of copper contamination of the silicon substrate. The resistance of via-last TSVs is substantially smaller than the via-first ones, thanks to their metal filling. In [20], TSVs with an aspect ratio of 10 (5 μm wide, 50 μm deep) yielded 60-70 mΩ, approximately the same value with the via-first option in [46], but with 20 times smaller diameters. As for their applications, via-last TSVs with a diameter and pitch of 30 and 80 μm respectively have been employed for the stacking of four DDRAM dies, achieving a total memory of 8 Gb and 1600 Mb/s data access rates [21].

2.1.3 Via-middle TSVs

The processing of via-middle TSVs takes place after the FEOL but before the BEOL. Their fabrication, which is illustrated in Fig. 2.3, shares many common steps with the via-last approach:

- **Step 1:** Through-silicon holes are formed through lithography and etching steps from the wafer front-side. Note that unlike the case of via-last TSVs, the through silicon holes do not extend from one side of the wafer to the other, since the wafer has not been thinned yet.

- **Step 2:** An oxide layer (TSV liner) is deposited over the through silicon hole in a conformal way, to electrically isolate the TSV conducting material from the surrounding silicon. However, contrary to the case of via-last TSVs, no removal of the liner from the bottom of the through silicon holes is required.
- **Step 3:** The through-silicon hole is filled with copper. To prevent diffusion of copper in the surrounding silicon, a barrier layer is also used, as it was done for via-last TSVs.
- **Step 4:** Wafer thinning is performed to "reveal" the TSVs. During this thinning step, the liner from the TSV's bottom is also etched to allow electrical connection to the wafer back-side.

Via-middle TSVs with 5 μm diameter and an aspect ratio of 10 (50 μm deep) have been shown in [48] and the scaling of the TSV diameter to 3 μm has been reported in [19].

2.1.4 Comparison between via-middle and via-last TSVs

So far, it is evident that via-last and via-middle TSVs show the strongest potential when it comes to the density of 3D interconnects. Choosing between them is a trade-off between cost and performance: overall, via-middle TSVs offer better performance and higher integration densities than via-last ones, but with increased implementation costs. The patterning of via-last TSVs after the BEOL requires larger dimensions, as well as larger landing pads to ensure high alignment precision between the back and front-side of a wafer [18]. For instance, although via-last TSVs have been shown to scale down to 5 μm in diameter, their landing pads cannot be smaller than 7.5 μm. Apart from the obvious limitations that large-diameter TSVs place on the density of 3D interconnects, they also tend to induce higher stress in the surrounding silicon. Stress-related effects that are associated with TSVs will be described in more detail in Section 2.2.2.

The patterning of via-last TSVs at the end of the wafer processing requires special attention not to degrade the wafer's interconnect lines. Towards this, no signal lines are allowed over them [21, 49, 50]. In other words, via-last TSVs form routing obstacles, which degrade routability and limit the TSV count in a design. Furthermore, no "reliability" anneal is allowed after the TSV filling, which would otherwise improve the stability of copper, by densifying it and removing any impurities or defects caused during TSV filling. On the other hand, via-middle TSVs are formed before the BEOL, so they do not impose such limitations.

However, the TSV's "reveal" step, inherent in via-middle TSVs, is rather complex [21], which may lead to increased fabrication costs. The simpler processing of via-last TSVs could enable their fabrication in a single facility/foundry [20], leading to additional benefits, cost- and production-wise. Based on the cost model proposed in [51], via-last TSVs are expected to reach 10 % cost reduction compared to via-middle ones.

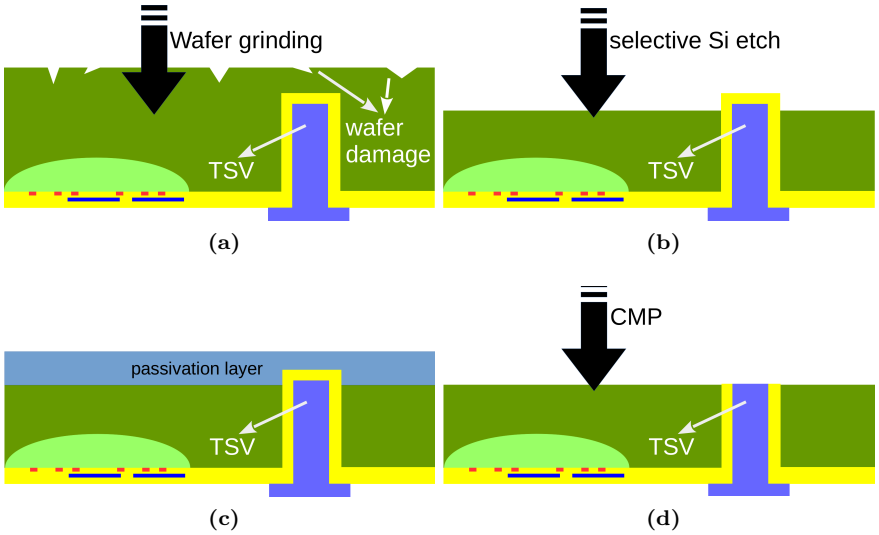


Figure 2.4: TSV-reveal steps: (a) Wafer grinding (b) selective Si etch (c) deposition of passivation layer to protect the Si substrate from copper contamination (d) TSV CMP to reveal the TSV

2.2 Issues with TSV processing

The processing of TSVs entails a number of yield-related risks that need to be addressed appropriately, most importantly copper contamination and stress-related issues. The focus of this section lies in these two challenges and the proposed approaches to overcome them.

2.2.1 Copper contamination

The risk of copper contamination of the silicon substrate has been already mentioned. However, even the use of a barrier layer does not guarantee full protection against copper contamination. In the via-last scheme, residual of the TSV copper filling process may appear on the backside of the processed wafers, which necessitates the efficient coating of the silicon wafer prior to TSV filling. On the other hand, in the via-middle scheme, copper particles may appear on the wafer backside during the TSV "reveal" process. Due to copper's high diffusivity in silicon, these particles can reach the device layers during the subsequent thermal cycles, severely degrading their performance. Thus, the TSV "reveal" process requires special handling and is often carried out in multiple steps [52]. An example of a TSV "reveal" process is illustrated in Fig. 2.4 and it is as follows:

- **Step 1.** Grinding is typically employed first for coarse wafer thinning. Wafer

grinding incurs substantial damage to the silicon substrate in the form of cracks and dislocations. To prevent this, grinding halts a few micrometers above the TSV bottom edge [52].

- **Step 2.** The remaining Si is selectively etched with respect to the TSV liner. Usually, Si over-etching is deliberately carried out to facilitate the TSV exposure.
- **Step 3.** The whole structure at the wafer back-side is covered with a thick passivation layer, to prevent copper contamination during the TSV-reveal process.
- **Step 4.** CMP is performed to remove the passivation layer, etch the liner and reveal the TSV.

2.2.2 Stress-related issues

The large discrepancy between the thermal expansion coefficients (CTEs) of silicon and copper can induce substantial stress in both the TSVs' filling material and the silicon substrate. This should be attributed to the thermal-cooling cycles that typically follow the TSV filling. The following example from the processing of via-middle TSVs serves to explain the emergence of stress as well as its evolution with temperature. During the TSV "reliability" anneal, the copper inside the TSV tries to expand ($CTE_{Cu} > CTE_{Si}$), however, as it is confined on all sides by silicon and dielectric, it experiences compressive stress. As the stress builds up, it can lead to deformation and copper pumping, as shown in Fig. 2.5(a)-(b). To suppress copper deformation, the TSV "reliability" anneal should be followed by a CMP step. Copper deformation can also occur after any similar thermal cycle, for instance during the subsequent BEOL processing. Increasing the duration and temperature of the TSV "reliability" anneal is known to negate deformation in such case [53].

During the subsequent cooling steps, the compressive stress experienced by copper turns into tensile [48], reaching relatively high values (100-800 MPa in [54]). This in turn, induces tensile stress in the surrounding silicon substrate that could impact the performance of neighboring transistors. The stress component that is vertical to the device plane is typically smaller than the in-plane component and thus it can be neglected [55]. Assuming cylindrical coordinates, the in-plane stress consists of a radial (σ_r) and a circumferential (σ_θ) component, as shown in Fig. 2.5(c). These components are equal in size but opposite in direction. They can be approximated by [48]:

$$\sigma_r \approx \sigma_{Cu}(\frac{d_{TSV}}{2r})^2 \tag{2.1}$$

$$\sigma_\theta \approx - \sigma_r \tag{2.2}$$

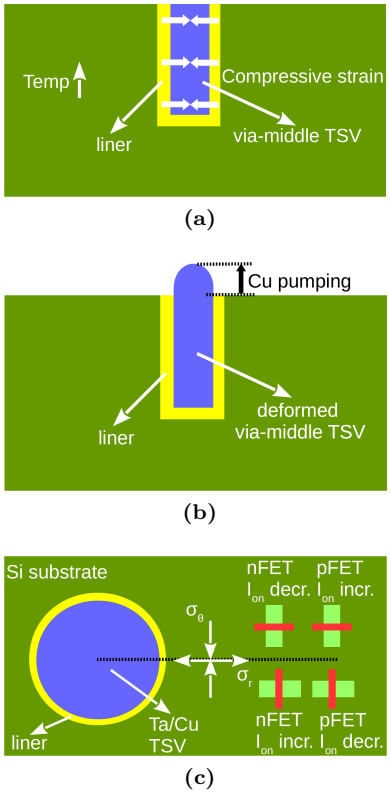


Figure 2.5: (a) Build-up of compressive stress in Cu during heating (b) Cu pumping to relax compressive stress (c) impact of TSV-induced stress in Si on neighboring FETs

where σ_{Cu} is the radial stress experienced by the copper filling, d_{TSV} is the TSV diameter and r the distance from the TSV center at which stress is calculated. The radial component σ_r is tensile, whereas the circumferential σ_θ is compressive as illustrated in Fig. 2.5(c). The impact of the TSV induced stress on FETs is in general complicated and depends on the device orientation relative to TSVs, their type (n- or p-type) and the device geometry (FinFET or planar). The situation is further aggravated by the application of stress-engineering in short-channel FETs to enhance carrier mobility in these devices.

Before the analysis, it is important to identify first the impact of stress on carrier mobilities. Compressive strain reduces the distance between lattice atoms, increasing thus the interactions between them and the electrons. Tensile strain on the other hand, reduces the electron-lattice atoms interactions. For holes, the situations is reversed: compressive strain facilitates the flow of holes, whereas tensile

impedes it. All in all, compressive strain enhances the mobility of holes, μ_h, and degrades the one of electrons, μ_e, whereas tensile strain improves μ_e and degrades μ_h. When it comes to the position of the devices relative to the TSVs, two cases can be distinguished [56]:

(a) Their current flow is parallel to the axis connecting them to the TSV.

(b) Their current flow is perpendicular to the axis connecting them to the TSV.

Both of these cases are illustrated in Fig. 2.5(c). For the first case, the device channel experiences tensile strain in the direction of the current flow, causing an increase in the on-current, I_{on}, for the NMOS and a reduction for the PMOS. Similarly, the second case leads to a reduction of I_{on} for the NMOS and an increase for the PMOS. The effects of TSV-induced stress on the transistors' performance are more pronounced for short-channel devices than long channel ones, as well as for PMOS compared to NMOS [56]. As for the device type, FinFETs appear to be less sensitive to TSV-induced stress than equally sized planar devices, thanks to the 3D topography of their channels [57].

To negate the variation of the transistors' I_{on}, caused by their proximity to TSVs, no devices should be placed close to TSVs, forming essentially a Keep Out Zone (KOZ). Eq. (2.1)-(2.2) indicate that the TSV-induced stress reduces with increasing the distance from the TSV center. Thus, the size of the KOZ is set to ensure that the percentage of on-current variation, $\frac{\Delta I_{on}}{I_{on}}$, does not exceed a specific threshold. Since analog circuits are more sensitive to variations in I_{on} than digital ones, this threshold is typically 0.5% for analog and 5% for digital circuits [56]. It is clear that the existence of a KOZ limits the density of 3D interconnects and so, its size should be minimized. A number of solutions have been proposed towards this:

- FinFET technologies could enable higher densities of TSVs than planar ones, since FinFETs are more immune to TSV-induced stress, which in turn leads to scaled down KOZs.
- Thinner TSVs induce less stress to the surrounding silicon, as indicated by (2.1)-(2.2). Thus, the KOZ dimensions could decrease with reducing the TSV diameter. A 37.5 % reduction in the width of the KOZ has been reported in [58] when scaling the TSV diameter from 5 μm to 3 μm.
- Further reduction in the TSV KOZ is possible by introducing air-gaps around the TSVs that could allow the copper TSV to expand and contract during thermal cycles freely, without interacting with the surrounding substrate [59]. Furthermore, this solution has the additional benefit of reducing the TSV capacitance, however, it requires complex processing and increased fabrication costs.

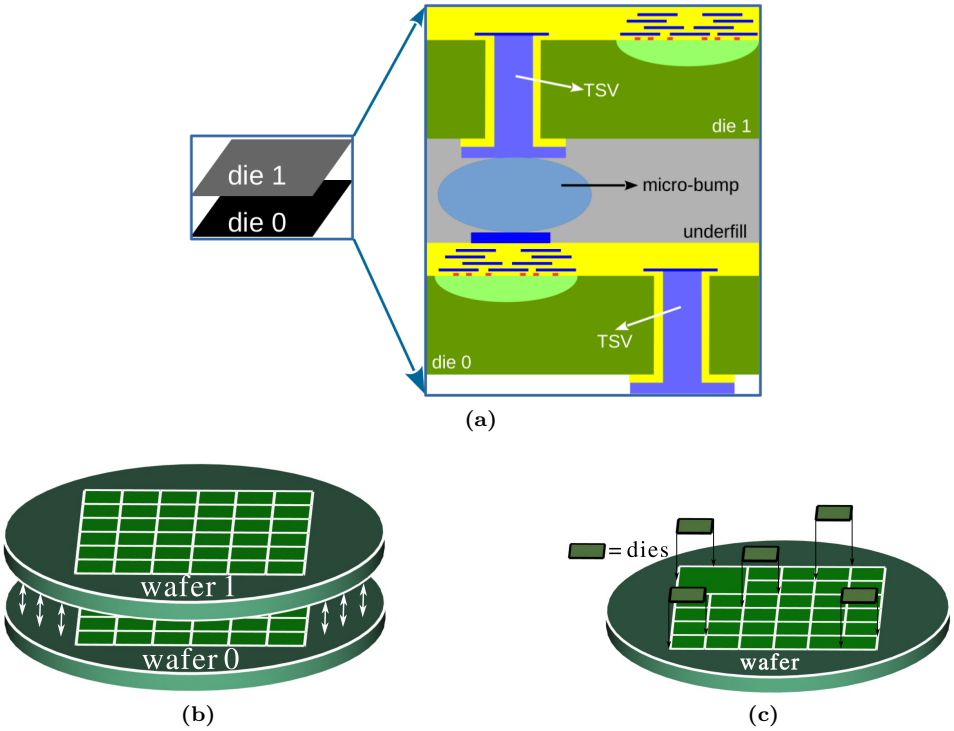


Figure 2.6: (a) Die-to-Die (D2D) bonding (b) Wafer-to-Wafer (W2W) bonding (c) Die-to-Wafer (D2W) bonding

2.3 Stacking Techniques with TSVs

In the previous section, the current status of TSV technologies has been introduced, along with some of their processing difficulties and the solutions proposed to tackle them. However, apart from the ability to connect the back-side of a die to its front, 3D integration calls also for solutions to enable the stacking (bonding) of two or more chips. The focus of this section lies on the latest trends in bonding technologies, their benefits and drawbacks, as well as their applications. The following three alternatives, illustrated in Fig. 2.6, are among the most popular ones for IC stacking:

- Die-to-Die bonding (D2D)
- Direct wafer bonding, or Wafer-to-Wafer bonding (W2W)
- Die-to-Wafer bonding (D2W)

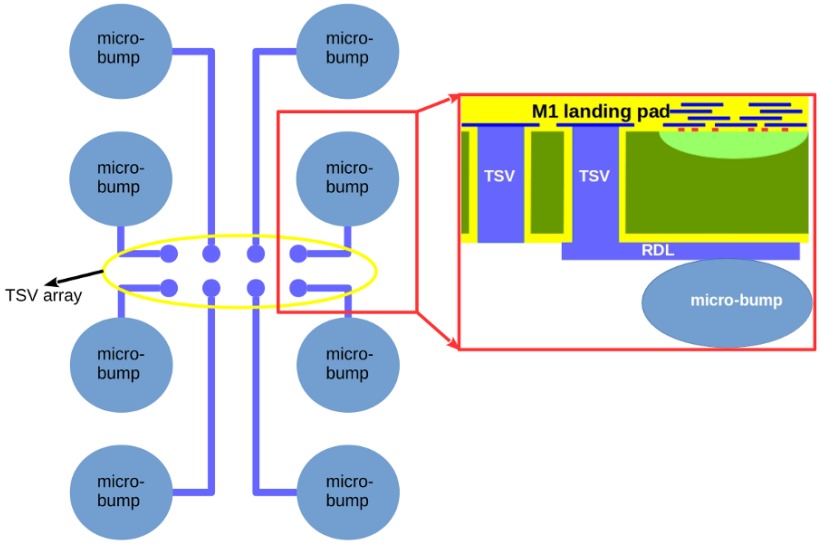


Figure 2.7: High density of TSVs, enabled by the use of a RDL.

2.3.1 Die-to-Die bonding

Die-to-die bonding, as the name implies, refers to the vertical integration of complete dies. It follows the formation of TSVs and the wafer dicing. The dies are connected through conductive micro-bumps, as shown in the inset of Fig. 2.6(a) [60]. To ensure high alignment precision between the stacked dies, the pitch of micro-bumps must remain large enough. In state-of-the-art processes used in production today, micro-bumps can be scaled down to 40 μm, significantly larger than the TSV-pitch (3-10 μm) [18]. This clearly limits the density of TSVs to the detriment of applications like DRAMs, which call for arrays of densely-packed 3D interconnects towards smaller area and larger bandwidth [61]. To accommodate the need for tightly packed TSVs with existing micro-bump technologies, a redistribution layer (RDL) is employed in the wafer back-side to route each TSV to a specific micro-bump, as shown conceptually in Fig. 2.7 [61]. One of the main benefits of D2D bonding is the resulting high yield, thanks to the stacking of only Known Good Dies (KGDs)

2.3.2 Direct Wafer bonding or Wafer-to-Wafer bonding

It is clear that the presence of micro-bumps inhibits high TSV densities. However, substantial improvement will be possible, if the use of micro-bumps is avoided. This can be achieved through the application of wafer-to-wafer bonding, in which two or more wafers are bonded together before dicing. To achieve high alignment

accuracy, the wafers need to be processed in the same foundry and by the same tools. Two cases for wafer-to-wafer (W2W) bonding are distinguished, based on the type of the bond that is formed between them:

- **Dielectric bonding.** A dielectric layer with low roughness is deposited over each wafer. Bringing the two wafers into contact leads to the formation of a temporary bond between them, which becomes permanent after annealing [18, 62]. To achieve electrical connectivity between the two wafers, via-last TSVs are patterned after the formation of the bond.

- **Copper-to-copper bonding.** Following the formation of Cu metal lines and/or TSVs, the two wafers are aligned and brought into contact. In other words, no dielectric layer is deposited over the wafer. After annealing, a Cu-Cu bond is formed in addition to the dielectric-dielectric bond in the locations with no copper [18, 62].

Apart from increasing the density of 3D interconnects, the absence of micro-bumps in the W2W case leads to significant reductions in the parasitic capacitance associated with the 3D interconnects. The adoption of W2W bonding has led to 75 % reduction in the capacitance of the 3D interconnects, as compared to the case of D2D bonding [63]. However, since W2W bonding occurs only before dicing, it is limited to the stacking of chips with the same sizes and consequently it has limited applications in heterogeneous integration. This is because in heterogeneous integration wafers of different technologies (i.e. Si and GaN) may differ in dimensions. Furthermore, W2W bonding suffers from lower yield compared to D2D, due to the lack of a priori knowledge of KGDs.

2.3.3 Die-To-Wafer bonding

Die-to-Wafer bonding combines the advantages of both W2W and D2D bonding. It consists of the stacking of KGDs over specific wafer places, just before dicing. Therefore, it allows better versatility in the sizes of the stacked ICs, as compared to W2W bonding. The yield achieved with D2W bonding is typically larger than the one achieved with W2W boding, but smaller than the D2D case. Unlike W2W bonding, in D2W micro-bumps are needed to ensure sufficient bonding alignment between a die and the wafer. D2W bonding has a better throughput than D2D bonding, however, significantly lower compared to the W2W bonding solution. That is because dies have to be stacked sequentially over a wafer, one at a time [64]. To improve the production throughput of the D2W bonding, the application of liquid droplets on the wafer and hydrophilic adhesion layers on the dies have been proposed to allow the simultaneous placement of dies over a wafer [65].

2.4 Sequential 3D Integration

So far, it is clear that W2W bonding offers the highest density of 3D interconnects as compared to other alternatives. However, further improvements are inhibited by the relatively large depth (≈ 50 μm) of the TSVs, which in turn call for wide diameters. Indeed, the limited coverage of high-aspect ratio topologies by the barrier layer (i.e. Ta) limits the TSV aspect ratios to approximately 10 [66] and consequently their pitch to 10 μm (double their diameter). Even if the TSV pitch scaled bellow this value, the minimum 3D interconnect pitch would be limited by the alignment accuracy between two wafers during bonding. It is expected that utilizing the W2W bonding approach, the minimum TSV pitch that can be achieved is 800 μm [18]. For higher densities of 3D interconnect vias, a different integration scheme must be adopted in order to minimize the wafer alignment requirements. Towards this, sequential 3D integration has been proposed [25, 67, 68]. Instead of bonding two pre-patterned wafers, a semiconductor layer can be transferred over a patterned wafer and processed to form devices and interconnects over it. The procedure can be repeated with layers of devices being formed sequentially over already patterned ones, in a tier-by-tier manner. The sequential nature of tier-by-tier processing implies that the alignment accuracy between the device tiers is dictated only by the lithography stepper. Hence, in a sequential 3D processing scheme, the 3D interconnect vias, known as Monolithic Inter-Tier Vias (MIVs), can be processed like any other metal via, leading to tremendous improvements in the density of 3D interconnects.

However, this unprecedented increase in the density of 3D interconnects comes with a penalty on the maximum allowable thermal budget for the processing of the top-tier devices. The top-tier's thermal budget affects both the FEOL and the BEOL of the bottom tier. The impact of high temperatures on the bottom FEOL has been studied in [69]. In particular, bottom tier temperatures above 500^oC for a duration of more than two hours have been shown to cause a positive shift for the n-FET threshold voltage and a negative one for the p-FETs. This behavior is particularly pronounced for short-channel devices and is caused by the oxygen diffusion through the spacers, which in turn degrades the work-function of the metal gate. Bottom tier temperatures higher than 500^oC have also an adverse effect on the stability of the source and drain silicide regions, causing fluctuations in the transistor access resistances. On the other hand, no change has been noticed in dopant diffusion or de-activation, as it can be demonstrated by negligible changes in DIBL before and after thermal cycles. Furthermore, the thickness of the gate-stack, as well as the carrier mobility in the channel appear to be immune to temperatures around 500^oC. As for the bottom tier's BEOL, high temperatures can cause voids in the copper interconnects, as well as copper diffusion in the surrounding dielectric, leading to concerns about the interconnects reliability [70].

Typically, high bottom tier temperatures can be caused by a number of top tier processing steps, such as layer transfer, dopant activation, silicidation, etc. For conventional planar CMOS technologies, the step with the highest processing

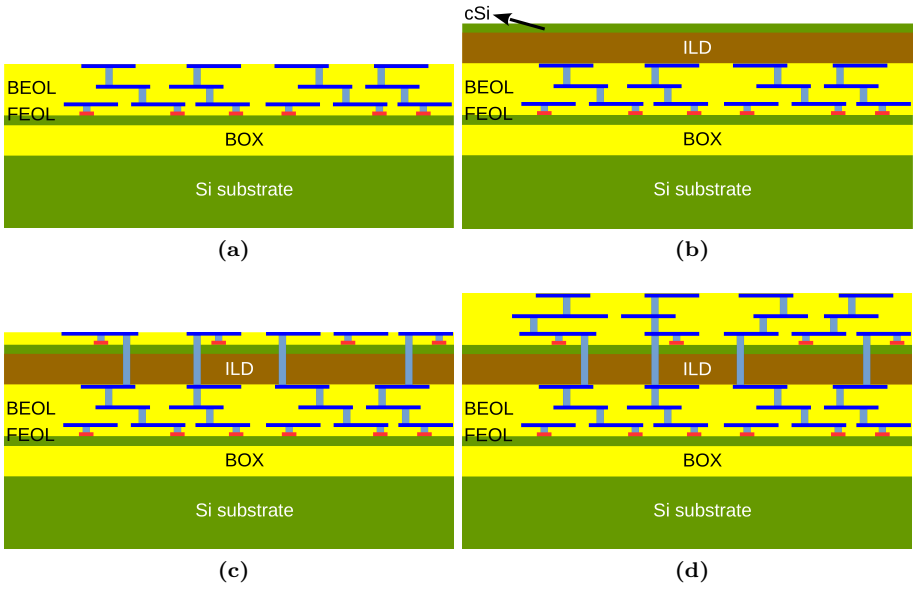


Figure 2.8: CMOS over CMOS S3D integration process steps: (a) Bottom tier with tungsten metal lines (b) Deposition of the ILD and transfer of a thin crystalline Si layer (c) Top tier FEOL and patterning of MIVs (d) Top tier BEOL

temperature is the dopant activation (>1000 ^{o}C) [24].To overcome these issues, various approaches have been proposed, which will be described thoroughly in the following subsections. These approaches can be categorized into (a) Si over Si and (b) heterogeneous integration.

2.4.1 Si over Si

The main challenge of the Si over Si S3D integration scheme is to achieve a low-temperature top tier FEOL process that yields devices with performance similar to their bottom tier counterparts. Based on the discussion above, care must be taken to ensure that the temperature in the bottom tier does not exceed 500^{o}C during the top tier processing. A processing scheme to allow Si-over-Si S3D integration is illustrated in Fig. 2.8 and it is as follows [71]:

- **Step 1.** An already patterned wafer can be used as the bottom tier. For the FEOL of the bottom tier, any device geometry can be used (FinFETs or planar), as well as any silicon substrate (SOI or bulk). For its BEOL, tungsten (W) is proposed to prevent any interconnect stability or contamination issues during the top tier processing [72].

- **Step 2.** A dielectric layer is deposited over the pre-patterned wafer. This layer serves as the Inter-Layer Dielectric (ILD) that separates the two tiers. To ensue low defect density during the subsequent layer transfer, the ILD must be first planarized through CMP. A thin silicon layer in crystalline form (cSi) is then transferred over the ILD and bonded through means of W2W bonding. For the layer transfer, SOI wafer bonding followed by etching to reveal the thin Si layer can be employed, as in [24]. This limits the top-tier devices to only SOIs.
- **Step 3.** The top tier devices can be now patterned on the transferred thin silicon layer. For the dopant activation, Solid-Phase Epitaxy (SPE) can be employed as in [71]. That way the maximum temperature required for the dopant activation in top-tier FETs has been reduced from over 1000^oC to 600^oC. Furthermore, the impact that SPE has on the performance of both top and bottom tier devices, appears to be negligible [24]. Following the patterning of top-tier devices, MIVs are formed to allow connection to the bottom tier's top metal layer. As already explained, the sequential processing enables ultra-tight MIV pitch, at the same scale with any other inter-metal via. Depending on the technology node, the MIV pitch could scale well below 100 nm [45].
- **Step 4.** Finally, the BEOL processing in the top-tier is carried out for the top-tier interconnects.

A laser pulse with a duration of a few hundred nanoseconds has been proposed as an alternative to SPE for the top-tier dopant activation [69, 73]. The main benefits of this approach is that since the laser beam has a low penetration depth, the produced thermal energy is confined in a small area close to the surface. By controlling the duration and energy of the beam, enough energy can be concentrated in the top tier to re-crystallize the silicon layer and to activate the dopants, while at the same time the temperature in the bottom tier can be kept below the safety barrier of 500^oC. Laser annealing offers similar benefits to Rapid Thermal Annealing (RTA) with the additional benefit of centralizing the energy concentration only where it is needed (top-tier). To better control the bottom tier temperature, some changes in the process flow of laser annealing have been proposed [69]:

- A silicon nitride capping layer (SiN) should cover the top tier surface before the application of the laser beam to ensure better uniformity of the laser energy across the top tier. Else, the distribution of the laser energy will depend heavily on the top tier topography, leading to potential damage of the exposed gate regions.
- Since silicon oxide is a thermal insulator, the ILD thickness should increase to reduce the thermal flow from the top to the bottom tier. An increase from 50 nm to 120 nm has proved sufficient in [69]. This step is vital for the stability of bottom tier metals [73].

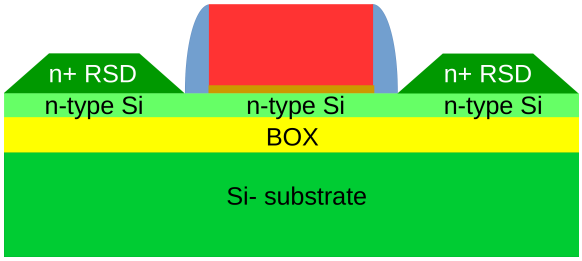


Figure 2.9: Structure of a n-type junctionless FET

- Similarly, if an SOI substrate is used in the bottom tier, the thickness of the Buried Oxide (BOX) should reduce (from 145 nm to 20 nm) to allow better dissipation of heat to the substrate and prevent it from building up in the bottom tier.

The effectiveness of the short duration laser annealing can be quantified by the sheet resistance of the source/drain regions. Typically, incomplete dopant activation leads to increased source/drain resistance. Experiments have shown that for most of the doping species, laser annealing yields sheet resistance values similar or even better than the ones obtained with RTA [69]. However, care must be taken to ensure that the energy dose of the laser beam does not exceed a specific value (≈ 0.9 J/cm^2), else the entire silicon layer liquidates. Once it solidifies, it becomes amorphous exhibiting very high sheet resistance [73].

2.4.1.1 S3D integration of junctionless transistors over Si-FETs

Another approach for CMOS over CMOS S3D integration is the use of junctionless transistors in the top tier [74]. Junctionless transistors, as the name implies avoid the use of p-n junctions in the FET structure, thus, the source and drain are of the same type as the channel region (see Fig.2.9). However, this feature limits the devices that can be integrated in the top tier to either n-type or p-type FETS, in other words no CMOS circuits can be designed in the top-tier. The threshold voltage in such topologies is set by calibrating the work-function of the gate material, which is achieved by selecting the appropriate material in the gate stack and its thickness. Threshold voltage depends also on the channel thickness and the doping concentration [74]. The main motivation for the use of junctionless transistors in the top tier is that the implantation and activation of dopants can occur using standard high temperature processing, before the W2W bonding. In [74], a thin silicon layer with uniform doping of 3e18 cm^{-3} has been used. Following the transfer of the doped and activated silicon layer over the bottom tier, the gate stack can be patterned. No silicidation of the source/drain regions has been carried out, instead

a low-temperature epitaxial growth has been considered for Raised Source/Drain (RSD) structures. This step has incurred also the highest temperature (525^oC).

Despite their obvious advantages for S3D integration, junctionless transistors suffer from a number of drawbacks, performance-wise:

- They suffer from relatively high source/drain resistance, which in turn degrades the transconductance g_m of the devices. One way to improve it, is to increase the doping concentration in the transferred layer. However, this leads to more pronounced short-channel effects and increased threshold voltage [74].
- Junctionless transistors suffer from random dopant activation, as opposed to Fully Depleted SOI (FDSOI) devices, due to the presence of dopants in the channel region. Thus, FDSOIs outperform junctionless transistors in terms of matching, limiting thus the potential of junctionless transistors in RF/AMS circuits and systems.

2.4.2 Heterogeneous S3D Integration

Another approach to achieve high performance top tier devices with a low thermal budget is the use of materials other than silicon for the top tier. Such materials must be characterized by a lower thermal budget for their dopant activation. They also need to have a bandgap value similar to silicon to allow compatibility with the state-of-the-art design methodologies. High bandgap values impede the efficient scaling of the threshold-voltage, leading to increased power supplies and consequently, power consumption. On the other hand, a low bandgap can lead to substantial off-current values, I_{off}. Furthermore, the material used for the top-tier channel region, should preferably exhibit carrier mobilities higher than the silicon ones to improve performance. Examples of such materials are:

(a) Carbon Nano-Tubes (CNTs)

(b) III-V materials, i.e. InGaAs

(c) germanium (Ge)

2.4.2.1 S3D integration of CNTs over Si-FETs

The use of CNTs for the top-tier devices has been motivated by the competitive energy-delay product of Carbon Nano-Tube FETs (CNTFETs), which exceeds that of silicon by an order of magnitude [75]. Improvements in the energy-delay product lead to energy-efficient designs with no performance trade-offs. The sequential 3D stacking of CNTFETs over planar silicon FETs is depicted in Fig. 2.10 [76]. In short, over a patterned Si wafer, a dielectric layer is deposited, followed by the opening of the MIVs. Then, the back-gates of the CNTFETs are patterned and a thin Al_2O_3 dielectric is formed, that serves as the back-gate dielectric. Over

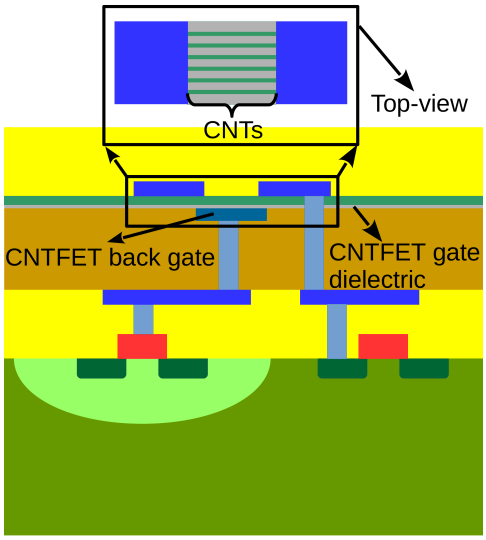


Figure 2.10: S3D integration of CNTFETs over Si FETs. The inset shows the top-view of a CNTFET

the back-gate dielectric, a layer of pre-processed CNTs is transferred, followed by the formation of source and drain terminals and the interconnect lines. That way, the high temperature required for the processing of CNTs (875^{o}C) is decoupled from the top tier's thermal budget. The temperature required for the CNT layer transfer is 130^{o}C, which makes the top-tier's BEOL the most critical step in terms of the top-tier's thermal budget (250^{o}C) [77]. The above S3D approach has been extended in [41], to also include tiers of Resistive RAMs (ReRAMs), which do not require high processing temperatures ($\approx$ 200^{o}C), opening thus the path for very tight co-integration of logic and memory.

One of the main drawbacks of this integration approach is the unconventional structure of the CNTFETs themselves. As shown in Fig. 2.10, CNTFETs are controlled by a back-gate which however can only be accessed from the bottom tier. Furthermore, CNTFETs suffer from two serious reliability issues [75]: (a) mispositioned CNTs and (b) metallic CNTs. Mispositioned CNTs can create unwanted connections between CNTFET circuits, whereas metallic CNTs inhibit the gate control over the nano-tubes leading to always "on" transistors. A layout-based technique to handle misaligned CNTs has been proposed in [78] and it involves "safety" regions between back-gates, where any CNTs are etched away. However, such an approach would lead to an area overhead that reduces the footprint gains of S3D integration. In general, the concern of misaligned CNTs limits significantly the potential applications of this S3D approach. As for the issue of metallic CNTs, their selective etching has been proposed in [77], which however requires additional

processing steps leading to increased costs and reduced wafer throughputs.

2.4.2.2 S3D integration of III-V FETs over silicon

III-V FETs could offer significantly higher electron mobilities compared to Si-FETs and they are considered as a possible candidate for future high-performance n-type FETs [79]. However, since III-V FETs suffer from very low hole mobilities (lower than in Si) [80], which limits the S3D integration of a III-V layer over silicon, to applications with only n-type FETs in the top tier. Such an integration scheme has been proposed in [81]. The basic processing steps are very similar to the ones employed for the Si over Si integration ([71]). The main difference occurs during the layer-transfer of an InGasAs layer over the ILD: an InP donor wafer is employed instead of a SOI one. Another difference is that no silicidation of the top tier's active regions occurs. Instead, a RSD approach is employed, which is formed by an InGaAs epitaxial growth with in-situ doping. The processing of RSDs is the most critical step in terms of thermal budget. However, the RSDs fail to suppress significantly the source/drain access resistance, as it can be demonstrated by the relatively high unit gain frequencies of the top tier transistors, f_T ($f_T \approx 16$ GHz for a InGaAs FinFET with L= 10 nm and W= 20 μm) [82]. There are two main concerns with this heterogeneous S3D approach:

(a) There is a size mismatch between the silicon wafer and the InP donor wafer; the Si wafer's diameter is 8 in versus the InP wafer's 2 in. This difference could limit the production throughput of S3D dies, which is defined by the size of the smaller donor wafers, leading also to increased costs per die.

(b) There are no interconnect lines in the bottom tier. Thus, the terminals of the bottom tier FETs are accessed only from the top tier, limiting the routing flexibility of this integration approach and leading to potential routing congestion issues (the amount of devices that need to be routed in a specific area doubles).

The performance metrics of the top tier InGasAs planar FETs are summarized in Table 2.1. It is clear that InGaAs planar FETs suffer from high Sub-threshold Slope (SS) values, i.e. 150 mV/dec for 50 nm long devices [81]. On the other hand, the situation is improved with InGaAs FinFETs built in the top tier, exhibiting SS $\approx$ 104 mV/dec.

A modification of this processing scheme has been proposed in [83]. There, III-V n-type FETs were integrated over p-type SiGe devices with no n+ implantation carried out in the bottom tier. Such an approach would limit the number of processing steps leading to potential cost gains. Nevertheless, the adoption of this scheme in large-scale applications would be inhibited by the consequent lack of ESD diodes, since diodes require both n+ and p+ implantations in the same tier.

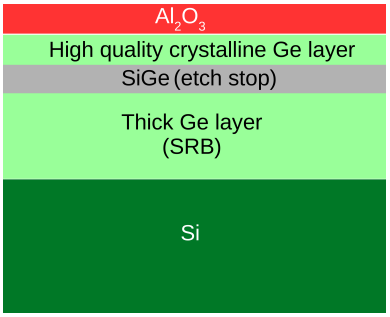


Figure 2.11: Formation of a high quality, thin and crystalline Ge layer over a Si wafer, as proposed in [2]

2.4.3 S3D integration of Ge FETs over Si-FETs

Contrary to III-V materials, Ge is characterized by hole mobility that is much larger than that of silicon's. Additionally, the processing of high performance n-type Ge FETs poses a lot of challenges [80]. Consequently, the heterogeneous S3D integration with Ge over Si, is limited to only p-type FETs in the top tier. An integration solution for Ge on top S3D ICs has been proposed in [43]. The main steps are similar to the ones used for the Si over Si integration in section 2.4.1. A Si donor wafer one is used, on which a high quality Ge layer is epitaxially grown. One of the main challenges of this approach is the lattice mismatch between silicon and germanium, which can lead to defects in the germanium layer and deformation [42]. Towards this, the use of a relatively thick Ge layer as a strain relaxed buffer (SRB) has been proposed. Then, on top of the SRB, the thin Ge layer that will serve as the top tier device layer is grown. The resulting stack of the donor wafer is shown in Fig. 2.11. For the W2W bonding, a layer of Al_2O_3 has been considered in the donor wafer, as it provides the best trade-off between surface roughness ($\approx$ 0.5 nm) and growth temperature (200^oC) [84]. Similarly, silicon oxide is used in the bottom tier wafer. Following the W2W bonding, the thin Ge layer needs to be revealed. This takes place in three steps. Firstly, the silicon substrate of the donor wafer needs to be etched away, followed by the SRB. During the SRB etching, a SiGe layer acts as an etch-stop to prevent damaging the thin Ge layer. Finally, the SiGe stop-etch is also removed.

As for the top-tier device fabrication, an Equivalent Oxide Thickness (EOT) of 4.3 nm was achieved through thermal oxidation (resulting in GeO_2) and deposition of 5 nm of Al_2O_3. The patterning of the gate-stack requires special attention since Ge's native oxide is soluble in water [80]. Furthermore, the GeO_2/Ge interface is prone to defects and large interface state densities, D_{it}, which can cause severe degradation of the Ge FET channel mobilities, and off-state leakage. The interface state density reported in [43], $D_{it} \approx$ 4.5e12 $eV^{-1}cm^{-2}$, is larger than the one ob-

Table 2.1: Heterogeneous S3D integration: Comparison between Ge and InGaAs top-tier FETs.

-	[81]	[43]
top tier material	InGasAs	Ge
Donor Wafer	InP	Si
Lg	50 nm	800 nm
$\mathbf{I}_{on}$	300 μA/μm	50 μA/μm
$\mathbf{I}_{on}/\mathbf{I}_{off}$	750	600
SS	150 mV/dec	170 mV/dec

tained for III-V materials in [81] (2e12 $eV^{-1}cm^{-2}$). The most critical step in terms of the top tier's processing temperature has been the dopant activation with 600^oC. No silicidation process followed the dopant activation, neither epitaxial growth for RSD, as it was the case in [81], leading thus to a high source/drain resistance.

The performance metrics of the top-tier Ge-FETs are also summarized in Table 2.1, together with the performance of III-V FETs. Due to the lack of a top-tier silicide or RSD structure, only long channel devices (800 nm) were reported. Ge-FETs exhibited really large SS values (170 mV/dec), larger than the one obtained for InGaAs FETs, which had a significantly shorter channel length (50 nm vs 800 nm). A better quality interface between the germanium layer and the gate-dielectric would significantly reduce the off-state leakage and improve the performance of Ge FETs. One possible solution for higher quality interfaces is the deposition of a thin silicon capping layer, above the channel region [80]. Furthermore, improved electrostatic control over the channel and improved SS can be achieved by thinning the germanium layer. Despite the relatively poor performance of germanium transistors, as exhibited by the current trends in their processing technologies, they offer a major advantage over III-V FETs: with the use of a SRB, any widely available Si wafer can be employed to form donor wafers, leading to increased production throughput (typically the size of Si wafer is larger than the InP ones) and reduced costs, as compared to the III-V solution.

2.5 Heat Dissipation in 3D ICs

3D integration opens up the path to higher integration densities, which consequently result in higher power densities. This, in turn, gives rise to concerns about the efficiency of heat dissipation in 3D ICs. Indeed, thermal simulations indicate that 3D ICs exhibit higher peak temperatures than conventional 2-D ICs, with the increase depending on the employed floorplan [85]. To better understand heat dissipation in 3D ICs, the thermal behavior of the materials used in an IC stack needs to be described. In general, silicon and metals act as heat conductors, whereas the

various dielectrics as insulators. Heat in a 3D IC flows in two directions: laterally, mainly across the silicon layers and vertically towards heat sinks where it is dissipated. The first mechanism helps reduce the intensity of hot spots, whereas the latter reduces the average temperature in the die. The main technology parameters that impact the heat dissipation in 3D ICs are [86]:

- The thickness of the stacked semiconductors layers. The thicker the semiconductor layer, the more efficient its lateral thermal conductivity is.
- The density of 3D vias. More 3D vias correspond to more vertical paths for the heat to flow, thus improving the vertical heat conductivity.
- The number of stacked tiers. Power density scales up with the number of stacked tiers, leading to higher peak and average temperature in the IC.

A comparison between the efficiency of heat dissipation among various 3D integration technologies (TSV-based D2D and D2W bonding, W2W bonding with TSVs, S3D integration) is carried out in [86] and a summary is provided here.

- **TSV-based D2D and D2W bonding.** Both of these two approaches require the use of micro-bumps. The thickness of the silicon layers is relatively large ($\approx$50 μm), which improves the lateral heat flow. However, the underfill material that is inserted between the stacked dies to strengthen their bonding is a bad heat conductor. The latter, coupled to the large pitch of 3D interconnects results in very low values for the vertical thermal conductivity, which in turn leads to the worst heat dissipation efficiency among other 3D technologies. Simulation results indicate that the peak temperature depends on the density of TSVs. For a 3D stack of 8 dies, with a hot-spot present in the fourth die, the peak temperature exceeds 120^{o}C for a low count of TSVs and drops to 100^{o}C as their count increases [86].
- **W2W bonding with TSVs.** Unlike the previous case, no underfill layer is used in the W2W bonding scheme. Thus, given the large silicon thickness ($\approx$50 μm) and the lack of micro-bumps that limit the density of TSVs, this integration scheme results in the best heat dissipation efficiency with the peak temperature of an 8-die 3D stack reaching 90^{o}C.
- **S3D integration.** The very thin silicon layer in the top tier of a S3D stack, which is not continuous either, results in very low lateral thermal conductivity in the top tier. On the other hand, the combination of a very thin ILD, and the ultra-small pitch of MIVs results in very effective vertical heat diffusion. Simulations indicate that for a S3D process stack consisting of 8 device tiers, the peak temperature is approximately 95^{o}C, slightly higher than the case of W2W bonding.

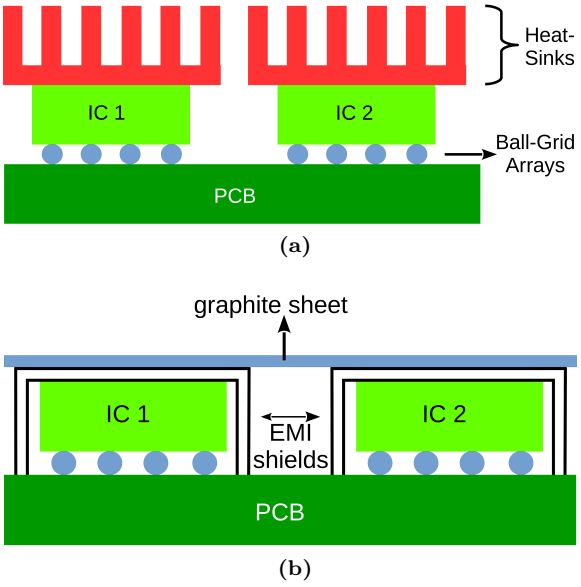


Figure 2.12: (a) Classic packaging solution for high power applications (b) New packaging approach for mobile applications

The peak temperature in an IC can be further reduced by adopting the packaging solutions used in today's mobile applications [87]. A comparison between such a package and a typical package for high-power applications is shown in Fig. 2.12. The main motivation behind the mobile-application package is area reduction with only minor penalties on heat dissipation efficiencies. This can be achieved through the use of a thin graphite layer and an Electro-Magnetic Interference (EMI) shield. They help to uniformly spread heat among packages (the graphite layer covers all the packages) and the PCB. As for their impact on 3D chips, they distribute heat inside the chips both downwards and upwards in a uniform manner, reducing the intensity of heat-spots and leading to lower peak and average temperatures [87]. All in all, heat dissipation does not appear to be a roadblock towards 3D integration with high densities of vertical interconnects in general and S3D integration in particular.

2.6 Summary

This chapter has provided a brief overview of the various processing technologies that have been developed for 3D ICs: TSVs, D2D, W2W and D2W bonding, S3D integration. Among them, S3D integration offers the lowest pitch of 3D interconnects, and consequently the largest improvements in circuit performance. Hence, they will be the focus of this thesis. The proposed solutions for S3D integration

can be categorized into homogeneous (Si over Si) and heterogeneous integration. As for heat dissipation, the high density of 3D interconnects ensures high heat conductivity in the vertical direction. This, in conjunction with appropriate packaging solutions, could lead to reduced peak and average temperatures, further solidifying the potential of S3D ICs.

Chapter 3

Sequential 3D Design Platform for Custom Integrated Circuits

In the previous chapter, various ways to overcome the S3D processing difficulties were described, highlighting the potential of S3D integration for future ICs. In the present chapter, the focus shifts to the development of S3D design platforms, which are required for the design of S3D circuits.

3.1 Sequential 3D Design Platforms

3.1.1 Overview

When it comes to new process integration technologies, the benefits of a design platform are twofold. Firstly, it enables an unimpeded collaboration between IC designers, process integration engineers and model developers. Second, it facilitates the study and exploration of design techniques for a new technology, without incurring high fabrication costs at an early stage of the process technology development.

Various design platforms for S3D integration have been proposed to explore circuit applications and in particular digital circuits. However, these design platforms have been developed for specific purposes only. For instance, design platforms for transistor-level S3D designs were developed in [34,44,88], whereas the ones in [89,90] aimed at the benchmarking of gate-level S3D circuits. Furthermore, some of the developed S3D design platforms were based on assumptions that have been proved wrong for S3D integration, such as the use of bulk transistor models for top-tier transistors [91]. In [92], a S3D 7 nm digital cell library was obtained from a 45 nm technology, simply by scaling the dimensions of the cells and the interconnects. Although such an approach was deemed sufficient for the scope of [92], it did not consider the complex Intermediate End Of Line (IEOL) that is used in advanced nodes and that could impact the parasitics inside the digital cells.

It is clear that for a reliable exploration of S3D circuit design, a S3D design

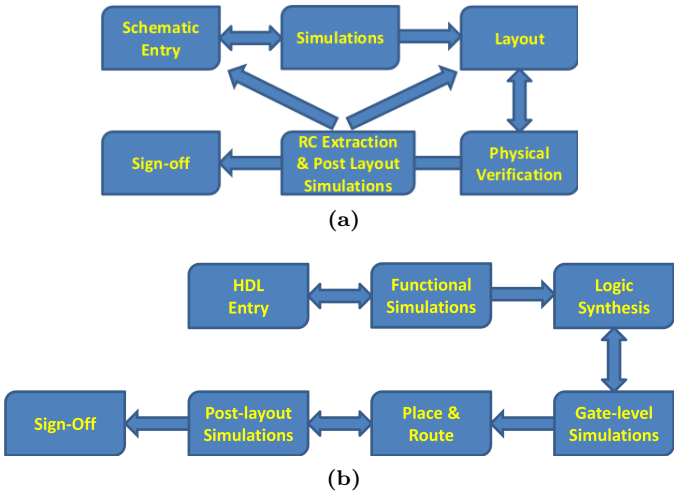


Figure 3.1: (a) Custom IC design flow (b) Digital IC design flow

platform needs to take into consideration the unique features of S3D integration, like the thin active layer in the top-tiers and the tungsten metal layers in the bottom. Furthermore, contrary to digital IC design platforms, a design platform for custom ICs could allow a complete exploration of S3D applications: analog, digital, mixed-signal, radio, etc. Therefore, S3D custom IC design platforms are the focus of this chapter and their development will be thoroughly described in the coming sections.

3.1.2 Custom IC Design Platforms

A design platform for custom ICs has a different structure than the one for digital ICs, reflecting the differences between the RF/AMS and digital IC flows. Custom IC design flows leave less room for automation in exchange for a larger design space exploration, with more physical implementation choices. These differences are illustrated in Fig. 3.1. A custom IC design flow starts with the schematic entry, in which the sizing and biasing of each device are defined. After validating the results with schematic-level simulations, the designer proceeds to the physical implementation of the design (layout). Care must be taken to ensure that the drawn layout corresponds to the schematic entry and that it conforms to the design rules set by the foundry. Finally, to account for layout parasitics, post-layout simulations that take into consideration the extracted parasitics are carried out. To accommodate these features, a custom IC design platform must exhibit a module-based structure, with each module responsible for specific design steps. This collection of modules, forms a Process Design Kit (PDK). The main modules of a PDK are:

- **Technology Description Files**. They provide a description of the process stack (i.e. metal layers, vias, some basic design rules) which is necessary for drawing layouts.
- **Parameterized cells (PCells)**. PCells facilitate both the schematic and layout entries, as well as schematic-level simulations. Their purpose is twofold: (a) They update the device models according to schematic parameters. For instance, a NFET PCell updates the corresponding model every time its dimensions are changed in the schematic. (b) They link the layout dimensions of a device to its schematic parameters, which enables the automatic generation of device layouts.
- **Device Models**. A trade-off exists between the accuracy of the device models and the simulations run-time. For the case of FETs, the best compromise is typically achieved with spice models (BSIM3, BSIM4, UTSOI, etc).
- **Physical Verification Scripts**. There are two main categories of physical-verification scripts: DRC and LVS. DRC scripts include the foundry design rules that dictate the minimum size, enclosure and spacing for various layout objects. On the other hand, LVS scripts define the connectivity between various layers (i.e. metal 1 connects to metal 2 through via 1) as well as the rules for identifying the various device types found in a layout. They also include extraction functions for the calculation of device dimensions, such as the width and length for FETs or the area and perimeter for capacitors. The calculated dimensions are needed for post-layout simulations.
- **Parasitic extraction flow**. A parasitic extraction flow requires typically a look-up table with the results of electro-magnetic (EM) simulations. These EM simulations are run only once for the whole process stack and calculate the coupling capacitance between various layout topologies, as well as the resistance of such structures. Then, during parasitic extraction, layout patterns are identified and their capacitance and resistance are estimated by extrapolating the corresponding look-up table values.

In this thesis, two S3D PDKs have been developed, one to support the development of the KTH in-house S3D process (KTH-S3D PDK) and a S3D predictive PDK (S3D PDK) that is used to explore the potential of S3D integration for custom ICs. The S3D PPDK is built upon a commercial CMOS process. Both will be described in detail in the following sections.

3.2 KTH-S3D PDK

The development of the KTH-S3D PDK has been carried out in close-collaboration with the process integration group at KTH and it reflects the current state of the in-house S3D process. Future changes in the process-stack can be easily annotated

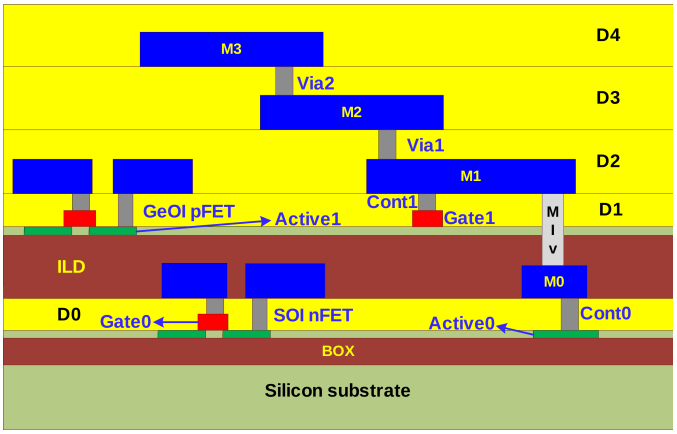


Figure 3.2: Process stack of the in-house KTH-S3D process

in the design platform by updating the necessary files. The KTH-S3D PDK is structured in such a way that it can be applied for the design of both S3D and conventional 2-D circuits. Indeed, 2-D circuits can be designed by utilizing only the top-tier layers. In such case, the parasitic extraction look-up tables need to be re-generated to account for the different process stack. However, all other PDK modules (i.e. physical verification scripts, device models, PCELLs, technology description files) can be re-used.

3.2.1 KTH-S3D process stack

Solid knowledge of the process stack is essential for the development of both the technology description files and the parasitic extraction flow. The current version of the KTH-S3D process stack is illustrated in Fig. 3.2. A SOI wafer is used for the FEOL of the bottom tier. The thickness of the bottom tier's active area (layer "Active0") is t_{Si} = 25 nm and that of the Buried Oxide (BOX) t_{BOX} = 145 nm. Both p-type and n-type FDSOI FETs can be processed in the bottom tier. Thermal oxidation generates a gate dielectric (SiO_2) with an EOT = 4.5 nm. To set the transistor's threshold voltage, a thin tinitride layer is deposited over the gate dielectric. The bottom tier interconnects consist of one tungsten (W) layer (M0). It is important to note that the dielectric between the bottom tier's active region and M0 is not planarized, leading to variations in the M0-Active0 distance, according to the underlying topology. More specifically, the distance of a M0 line to the silicon substrate is larger if it crosses over a poly-silicon pattern.

A 1000 nm thick SiO_2 layer that stretches from the bottom of M0 serves as the ILD, upon which a thin ($t_{Ge} \approx$ 25 nm) Ge layer is transferred through wafer bonding and a SRB [43]. Given the 500 nm thickness of M0, the distance between

Table 3.1: Characteristics of the in-house KTH-S3D process stack

Routing Layer	**Material**	**Thickness [nm]**	**Sheet resistance [Ω/□]**
Gate_0	Sd Poly-Si [1]	100	20
Gate_0_SBlk	Non-Sd Poly-Si [2]	100	120
M0	W	500	135m
Gate_1	Sd Poly-Si [1]	100	20
Gate_1_SBlk	Non-Sd Poly-Si [2]	100	120
M1-M3	Al	500	50m
Dielectric Layers	**Material**	**Thickness [nm]**	**Dielectric Constant ϵ_r**
D0 [3]	SiO_2 (CVD)	500	3.1
ILD	SiO_2 (CVD)	1000	3.1
D1 [3]	SiO_2 (CVD)	1000	3.1
D2-D4	SiO_2 (CVD)	500	3.1

[1] silicided poly-silicon
[2] non-silicided poly-silicon
[3] not planarized

the Ge-layer and M0 is 500 nm. The use of Ge as channel material for the top-tier FETs limits them to p-type only. Three aluminum metal layers comprise the top-tier's BEOL (M1-M3). Apart from the M1-Active1 dielectric (D1), all other top-tier inter-metal dielectrics are planarized. The dimensions, materials and properties of the used metal and dielectric layers are summarized in Table 3.1. The dielectric constant of D0-D3 is lower than the expected value for thermally gown oxides (≈ 3.9), since the employed deposition process, Chemical Vapor Deposition (CVD), results in less dense, porous-like structures and consequently lower ϵ_r.

3.2.2 KTH-S3D PCELLs

The in-house KTH-S3D process contains three different types of devices, with PCELLS available for each one of them:

- **FETs**. Both p-type and n-type FETs are available in the bottom tier. On the other hand, the use of Ge as the top active layer limits the type of top-tier transistors to pFETs only. The developed transistor PCells enable the automatic generation of transistor layouts, based on the schematic entries of the width, length and number of fingers. The contact arrays to the active or gate regions are also automatically generated. The minimum transistor length is

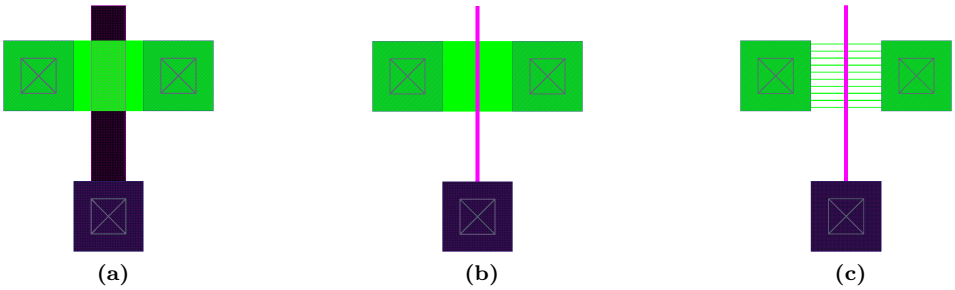


Figure 3.3: (a) Automatically generated layout for a pFET with L= 1 μm and W= 2 μm (b) Manually generated layout of a pFET with L= 100 nm and W= 2 μm (c) Manually generated layout of a nano-wire pFET with L= 100 nm. The device consists of 10 nano-wires, each 20 nm wide.

0.6 μm, limited by the resolution of the lithography stepper. Shorter channel devices are possible with the use of Electron Beam Lithography (EBL) instead of optical lithography. To enable designs with short-channel devices, two new layers have been added in the process-description files: EGate0 and EGate1. These layers are used to generate the masksets for the EBL patterning of transistor gates. Similarly, EBL can be also applied to pattern ultra-narrow active regions (W $\approx$ 20 nm), which can find use in nano-wire transistors. Therefore, two more layers have been added in the technology description files: EActive0 and EActive1. However, PCells are not available for short-channel FETs, neither for nanowire transistors, so their layouts need to be drawn manually. No models are available for these devices either. Fig. 3.3 demonstrates the automatic layout generation of a FET, as well as the manually created layouts of a short-channel transistor and a nano-wire FET.

- **Resistors**. Currently, four different resistor types are available in the in-house process, all in the bottom tier: silicided and non-silicided heavily doped active regions and silicided and non-silicided poly-silicon regions. The sheet resistance obtained with each of these options is shown in Table 3.2. In the future, top-tier silicided/non-silicided active regions (with Ge) are expected to be used as resistors, once efficient processes for low-temperature dopant activation and germanide reactions are established. The resistor PCells enable the automatic generation of their layouts based on the schematic-entries for width and length.

- **Diodes**. Since diodes require two regions with different dopant species (one with donors, the other with acceptors), they cannot be processed in the top-tier, where only p-type dopants are available. Bottom tier diodes are processed as PiN diodes (p+/intrinsic Si/n+), with the length of the intrinsic region set to 2 μm. The presence of the intrinsic region between the two

Table 3.2: Sheet resistance for the KTH-S3D resistors

Resistor Layer	**Sheet Resistance** [Ω/□]
Gate_0	120
Silicided Gate_0	20
n+ or p+ Si	600
silicided n+ or p+ Si	8

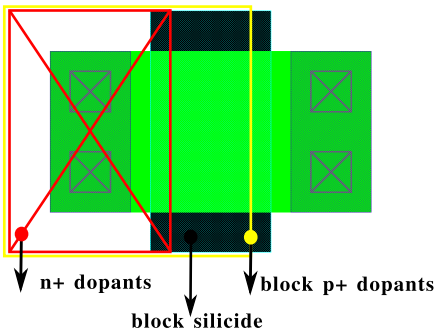


Figure 3.4: Layout of a KTH-S3D diode with W= 2μm.

heavily doped ones aims to reduce the potentially high electric field across the junction. Care must be also taken to ensure that silicidation occurs only along the contacts to the p+ and n+ regions, to avoid any risk of short-circuiting the two diode terminals. An example of a diode layout is shown in Fig. 3.4.

- **Capacitors**. No Metal-Insulator-Metal (MIM) or Metal-Oxide-Metal (MOM) capacitors are available in the KTH-S3D PDK. Structures with high capacitance densities can be obtained through Metal-Oxide-Semiconductor Capacitors (MOSCAPs), by short-circuiting the drain and source terminals of a FET. However, the total capacitance of such topologies is bias-dependent (it depends on the V_{GS}, V_{GD} values), which limits their use in ICs to few applications, such as decoupling capacitors.

3.2.3 KTH-S3D Device Models

Accurate device models are of paramount importance for the reliable validation of circuit performance and the development of new design methodologies. However, process-induced variations relax the need for highly accurate device models.

The two main sources of performance variations are geometry mismatches (spacer width variations [93], Line Edge Roughness-LER [94], etc), and Random Dopant Fluctuation (RDF) [95]. When compared to bulk transistors, the lack of dopants in the channel region translates to a reduced performance variation for SOI FETs. For instance, the threshold voltage variation, $\sigma_{V_{TH}}$, reaches 44 mV for bulk FETs, substantially higher than 17 mV for SOI transistors [96]. However, SOIs are prone to variations in the source/drain access resistance, R_{SD}, [97]. All in all, variations in I_{on} can exceed 10% for short-channel devices [95].

The S3D-PDK contains device models for the top and bottom tier transistors, as well as for the bottom tier diodes.

Transistor models. Spice-based models for the KTH-S3D transistors were preferred over Verilog-A models, as the former provide better integration with industry standard simulators and analysis types (i.e. transient, noise, harmonic balance, stability, etc). Two models have been developed for lowly doped or fully depleted thin channels: (a) the BSIM-IMG (Independent Multi-Gate) model by University of California Berkley [98] and (b) the UTSOI2.1 (Ultra-Thin SOI) model by CEA-LETI [99, 100]. The BSIM-IMG model assumes that the surface controlled by the back-gate is not inverted. On the other hand, the UTSOI2.1 model makes no such assumptions and is valid across a wide range of back-gate biases. Thanks to their versatility, UTSOI2.1 models will be used for the the KTH-S3D transistors. The model parameters have been calibrated to match measurement results. However, due to the lack of an effective silicidation-like process in both the top (Ge pFETs) and bottom tiers (Si n- and pFETs) only long channel devices were used for the model calibration, in particular devices with L = 2 μm. The transfer (I_D versus V_{GS}) and output (I_D versus V_{DS}) characteristics in Fig. 3.5 highlight the good agreement between the UTSOI2.1 model and measurement results for both bottom tier (Si) and top-tier (Ge) transistors. The calibration of the transistor model parameters against actual measurement results makes sure that the impact of the Ge layer on the performance of top-tier transistors is accounted for in the KTH-S3D PDK. This is in contrast to [34, 88], where the same models have been considered for the top and bottom tier FETs.

Diode model. To model the behavior of the in-house diodes, a simple DIODE level-1 model has been employed. It accounts for the I-V response in the forward and reverse bias regions, the breakdown voltage, parasitic resistances, as well as diffusion, depletion and overlap capacitances. Measurement results of the in-house diodes' I-V response are plotted in Fig. 3.6 together with the corresponding results obtained with the calibrated level-1 diode model.

3.2.4 KTH-S3D Physical Verification

Cadence® Physical Verification System (PVS) has been employed for the LVS and DRC runs in the KTH-S3D PDK. The main motivation behind this choice has been to allow the flawless integration of the developed PDK with industry standard CAD-EDA tools. Setting-up the physical verification scripts, and in particular the DRC

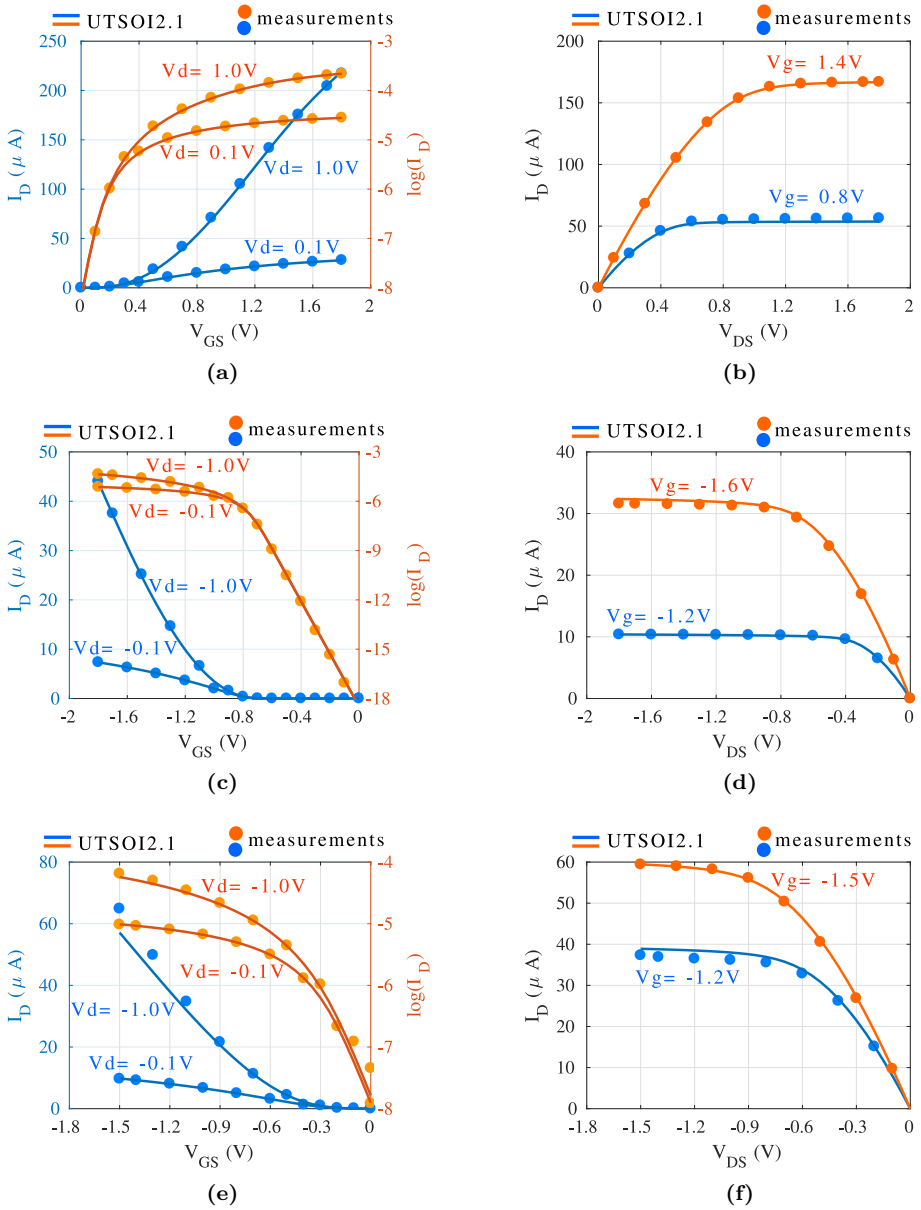


Figure 3.5: UTSOI2.1 model and measurement results for (a)-(b) a Si nFET in the bottom-tier, (c)-(d) a Si pFET in the bottom tier and (e)-(f) a Ge pFET in the top tier. The plots in the left show the transfer characteristics and the ones in the right the output characteristics. All devices have equal dimensions, W=L= 2 μm.

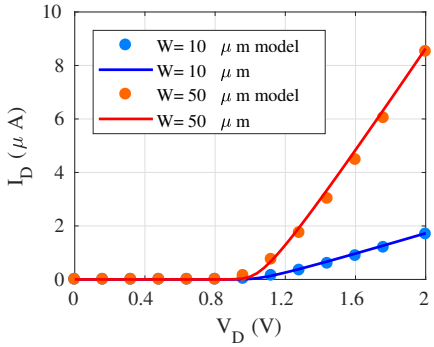


Figure 3.6: Level-1 diode models versus measurement results for the I-V charcteristics of two diodes fabricated by the in-house process.

Table 3.3: Minimum width and spacing for the layers in the KTH-S3D PDK

Layer	**Min. Width μm**	**Min. Spacing μm**
Active0	2	2
Gate0	0.6	1
Contact0	1	1
M0	2	2
Active1	2	2
Gate1	0.6	1
Contact1	1	1
MIV	1	1
M1-M3	2	2
V1-V2	1	1

ones requires the design rules definition of the in-house process. The resolution of the photolithography stepper at the KTH cleanroom allows the scaling of critical dimensions down to 500 nm. However, processing at the tool limits can affect adversely the yield. The minimum widths and spacings shown in Table 3.3 were set as the best compromise between "ease" to design (designability) and high yield.

Special care must be also taken for handling the MIVs relative to Contact1, Active1 and Gate1 layers, given that they run parallel to the latter, as shown in Fig. 3.7. This is a unique feature of S3D integration. The minimum distance between MIVs and the aforementioned layers was set equal to any other via-via / contact-contact minimum distance, i.e. 1 μm.

A characteristic of the KTH in-house S3D process is the inability to form via-

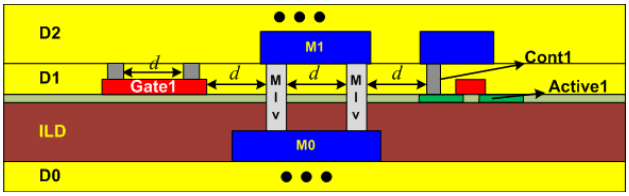
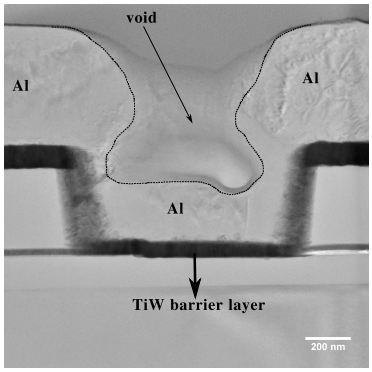
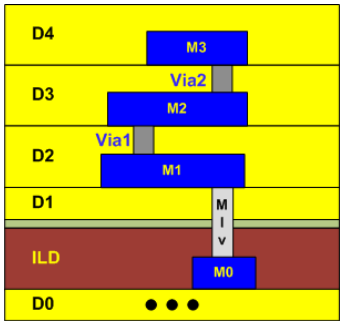


Figure 3.7: MIVs parallel to top-tier contacts (Cont1), poly-silicon (Gate1) and active regions (Active1). The minimum distance between these layers is d= 1 μm.



(a)



(b)

Figure 3.8: (a) TEM image of a via processed with the in-house process (b) Allowed stacking of non-successive vias (Via2 over MIV in this case).

stacks, owing to the incomplete via-filling shown in Fig. 3.8(a). Indeed, the incomplete filling of a via-hole can cause reliability issues with the etching and filling of the next via above it. Thus, an additional design rule has been added that prevents the stacking of directly successive vias. However, there is no restriction in placing non-successive vias (i.e. Via2 over MIV, etc) on top of each other, as shown in Fig. 3.8(b).

3.2.5 KTH-S3D parasitic extraction flow

A parasitic extraction flow is essential for the study of any potential coupling between the stacked tiers. A very thorough parasitic extraction flow has been proposed in [34, 44]. It incurs two steps: (a) to extract the parasitics of each tier separately, using commercial tools and (b) extract the coupling between stacked circuits with Synopsys® Sprocess and Sinterconnect flows. The main drawback of this solution is that it only suits parasitic extraction inside digital cells. Furthermore, it cannot be adopted by the established design flows, as it requires the combination of the

generated parasitics from each step in a single netlist.

In this work, a S3D parasitic extraction flow has been proposed and developed, based on Quantus from Cadence® , to ensure not only compatibility with established design flows but with industry-standard tools as well. Cadence® Quantus consists of two separate sub-tools: (a) QRC Techgen [101] and (b) QRC Extraction [102]. QRC Techgen is run only once to generate the electromagnetic (EM) simulation data for the whole process stack, as it is described in a process description file (.ict). The EM simulation data is then stored in look-up tables. QRC Extraction reads the output of the LVS run and calculates the layout parasitics by extrapolating data from the look-up tables. For higher accuracy in the calculation of parasitic capacitances, QRC Extraction can utilize a built-in Field-Solver (FS) that runs directly on the layout topology without incurring any extrapolation-related errors. However, this improvement in accuracy comes with a sever degradation in the extraction run-time and thus, field-solver assisted parasitic extraction is encouraged only for small regions of a layout [102].

The main effort in developing a parasitic extraction flow for the KTH-S3D PDK has been directed towards setting-up the process description file required by QRC Techgen. The main challenge has been Techgen's inability to handle more than one active-layer [101]. An interesting method to overcome this impediment has been introduced in [103]: the top-tier active regions have been handled first as conductors and then as dielectrics. The parasitics associated with each case are then combined to get the final value. However, there is little physical meaning in treating the top-tier active regions as dielectrics. More specifically, the source/drain regions are heavily doped, exhibiting low sheet resistances. Hence, in this work, top-active layers have been defined as conductors [104, 105]. As for the channel regions, typically, they are excluded from parasitic extraction, since the parasitics associated to the channel are accounted for in the device models. Additionally, the channel regions are shielded by the gate stacks, which prevent the electric field, generated by neighboring structures in the top tier, from reaching them. The relative large thickness of the ILD, minimizes also the coupling between channel regions and bottom tier structures. QRC Techgen has been also set to include the sidewall capacitance between the MOS gate and the contacts to the active regions in its calculations.

Not all of the PDK layers are considered during parasitic extraction. For instance, as mentioned before, a clear distinction is made between parts of the active regions that are placed under gate layers (channel regions) and those that are not. Active layers used in resistors have been also excluded from parasitic extraction to avoid counting the resistance between the device terminals twice, in the device model and again during parasitic extraction. The active layers used in PiN diodes are not considered for parasitic extraction either, since the series resistance in diodes is already considered in the diode model.

To ensure the validity of the proposed parasitic extraction flow and consequently its suitability to analyze and explore S3D circuits, its extracted parasitics have been compared against Electromagnetic (EM) simulations. To do so, the values of the

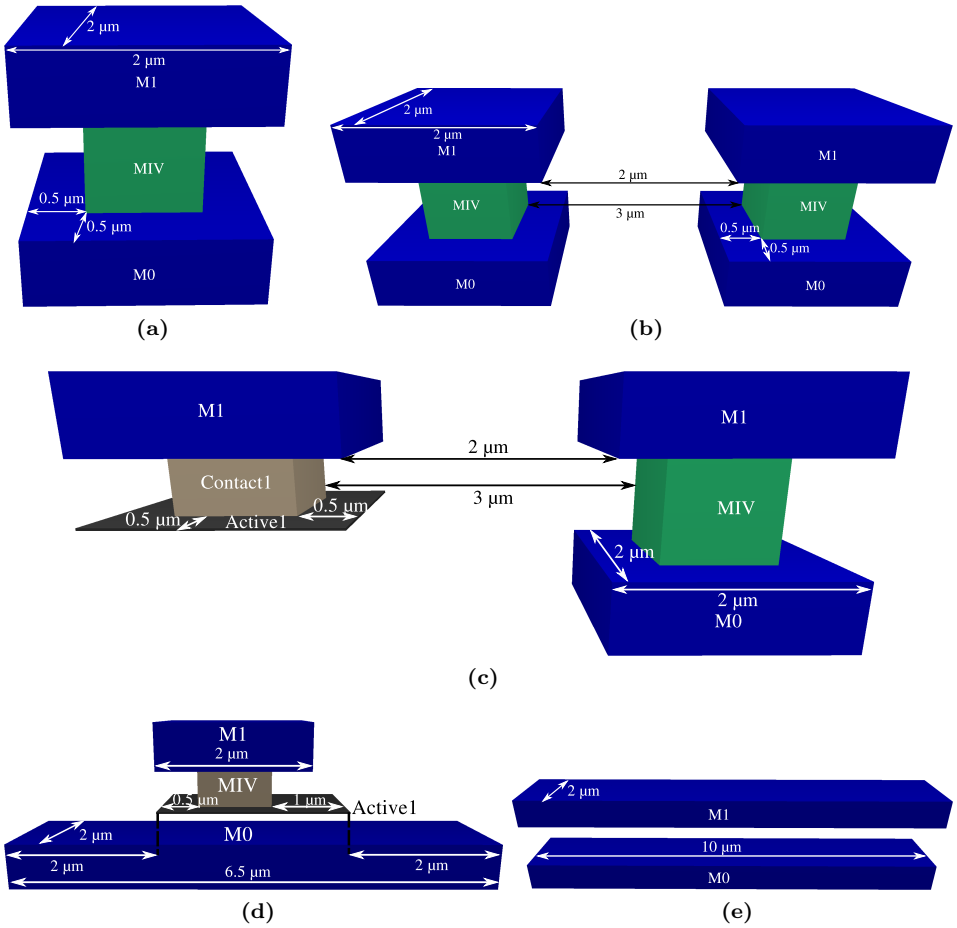


Figure 3.9: Layout structures to extract the parasitics of: (a) a single MIV (b) the coupling between two neighboring MIVs (c) the coupling between a MIV and a top-tier contact (Contact1) (d) the coupling between a M0 line and a top tier active (Active1) region (e) the capacitance between a M0 and a M1 line.

Table 3.4: Comparison between the parasitic capacitance obtained with the KTH-S3D parasitic extraction flow and with Momentum EM simulations (by Keysight®)

Layout Structure	**EM Simulations**	**QRC Extraction with FS**	**QRC Extraction without FS**
standalone MIV (Fig: 3.9(a))	0.63 fF	0.94 fF	1.2 fF
MIV-to-MIV (Fig: 3.9(b))	0.45 fF	0.54 fF	0.65 fF
MIV-to-Contact1 (Fig: 3.9(c))	0.38 fF	0.43 fF	0.58 fF
M0-to-Active1 (Fig: 3.9(d))	0.98 fF	1.19 fF	1.32 fF
M1-to-M0 (Fig: 3.9(e))	0.15 fF/μm	0.17 fF/μm	0.18 fF/μm

QRC extracted parasitics for each of the layout structures in Fig. 3.9 are compared against the corresponding values obtained with Keysight® ADS electromagnetic (EM) simulations (using the Momentum simulator). These test structures allow the extraction of the following parasitics:

- the parasitics of a single (standalone) MIV.
- the coupling between two MIVs. The distance between the two MIVs is set to 3 μm, owing to the KTH-S3D design rules that set the minimum enclosure of MIV by M1 and M0 shapes to 0.5 μm in all sides and the minimum distance between neighboring M1/M0 structures to 2 μm. Alternatively, the minimum distance between two MIVs could be employed (1 μm), assuming that both MIVs are enclosed by the same M0 and M1 structures. However, the latter case would result in iso-potential MIVs and consequently inability to calculate their capacitive coupling.
- the coupling between a MIV and a top-tier contact (Cont1). The same motivations with the previous case led to setting the MIV-Cont1 distance to 3 μm.
- the coupling between a M0 line and a top-tier active region (Active1) that is located directly above the M0 line.
- the coupling between two parallel lines, one in M1 and another in M0.

For the parasitics extracted by Quantus, two cases have been considered: (a) with and (b) without enabling the FS. The results are summarized in Table 3.4.

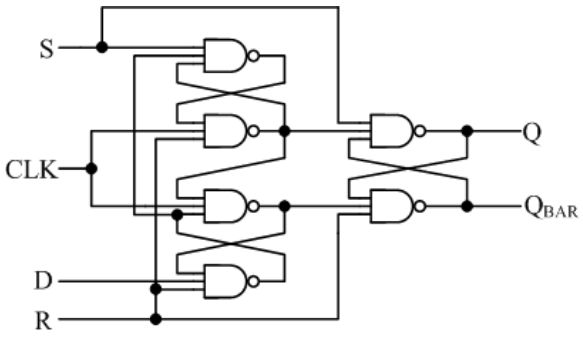


Figure 3.10: Schematic of a 3-input NAND-based D flip-flop.

Given the relatively small aspect ratio of the MIVs (1 μm width and 0.925 μm depth) the parasitic resistance associated with MIVs is negligible and has been omitted from the comparison. The results indicate that QRC Extraction tends to overestimate layout parasitics, especially for the case of a single MIV. Furthermore, FS assisted QRC Extraction yields more accurate results than the case without the FS, as it was expected. A final note is that for the in-house KTH-S3D process, the parasitics induced by the MIVs are rather low, particularly in comparison with TSV-based 3D integration (250 fF in [20]).

3.2.6 Digital Cells Library

A digital library consisting of 7 cells, both combinatorial and sequential has been designed for the S3D PDK. These cells are:

- inverters: INVX1, INVX3, INVX9
- buffer: BUFF
- NAND gate: NAND_I2 (2-input), NAND_I3 (3-input)
- D-flip/flops with set and reset functions: DFFSR

The cells count may appear small, however the logic functions represented by them enable the efficient generation of all other logic functions. The S3D cells were based on 9-tracks planar cells that were designed and fabricated at KTH using a SOI 2D in-house process. A transistor-level partitioning scheme has been adopted with Ge-pFETs in the top-tier and Si-nFETs in the bottom. The power lines run parallel to each other: the supply line runs in the top tier, connected to the pFETs and the ground line in the bottom tier, connected to the nFETs. The height of the KTH-S3D cells is set by the minimum height required to layout the largest cell in the library, i.e. the DFFSR, and it equals 6 tracks. The lack of switches

Table 3.5: Truth Table for the KTH-S3D DFFSR cell

Set	**Reset**	**Q**	$\mathbf{Q}_{bar}$
0	0	1	1
0	1	1	0
1	0	0	1
1	1	D	**NOT** D

in the in-house process limits DFFs to NAND-based implementations only. The schematic of the DFFSR is shown in Fig. 3.10. This implementation enables low-active, asynchronous Set and Reset functions and triggering at the rising edge of the clock signal, CLK. The truth table of the DFFSR cell for the rising edges of the clock is given in Table 3.5.

For the layout of the DFFSR, a module-based design methodology has been adopted, devised specifically for the in-house process. Naturally, the NAND cells that are used in the DFF could serve as modules. An optimized module layout, could improve the area efficiency of the DFF, particularly in terms of pin-access and intra-cell connectivity. The layout of the S3D DFFSR is shown in Fig. 3.11(a-c). The NAND-based module of the DFFSR is also illustrated in the inset of Fig. 3.11(c). The intra-cell routing is limited to bottom tier interconnects along with Gate1 and M1 layers from the top-tier, leaving thus M2 and M3 open for inter-cell routing, with no corresponding blockages inside the cells. The bottom tier interconnects are not considered for inter-cell routing due to tungsten's high resistivity.

The size of the cells is limited by the KTH-S3D design rule that does not allow the stacking of successive vias. Indeed, additional area overhead is required, close to the pin locations, to ensure their access from M2/M3 lines through Via1 and Via2 (Via1 and Via2 cannot be stacked). Had the stacking of successive vias been possible, the height of the S3D DFFSR would be reduced to 5 tracks as shown in Fig. 3.12, leading to further area gains. The potential of five track cells has been also demonstrated in [106] for a planar technology with FinFETs. However, two intermediate metal layers (Intermediate End Of Line - EOL) were added between the active region and M1, which were closed to the router. Furthermore, the active regions of both the pull-up and pull-down networks were limited to 1 fin for NMOS, severely limiting the driving power of these cells. To compensate the reduction of the cell's driving strength, the height of the FinFET fins could increase, resulting however in an increase of the intra-cell parasitics [106, 107]. On the other hand, the KTH-S3D cells do not trade-off area reduction with driving strength. On the contrary, thanks to the use of Ge in the top tier, the driving strength of the pull-up network would increase compared to the 2D 9 tracks implementation.

The KTH-S3D digital cells have been simulated using the models described in Section 3.2.3. Post-layout parasitics have been also included in the simulations.

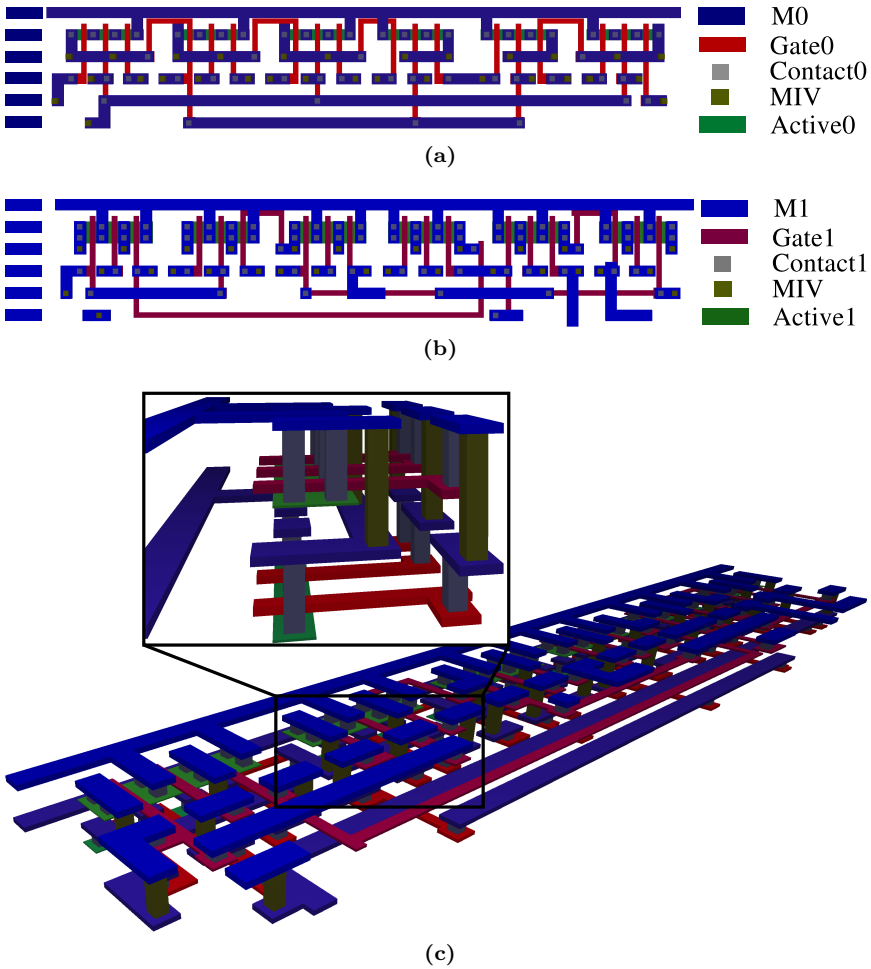


Figure 3.11: Layout view of the KTH-S3D DFFSR cell in (a) the bottom tier and (b) the top tier. (c) 3D view of the KTH-S3D DFFSR. The NAND based module is shown in the inset.

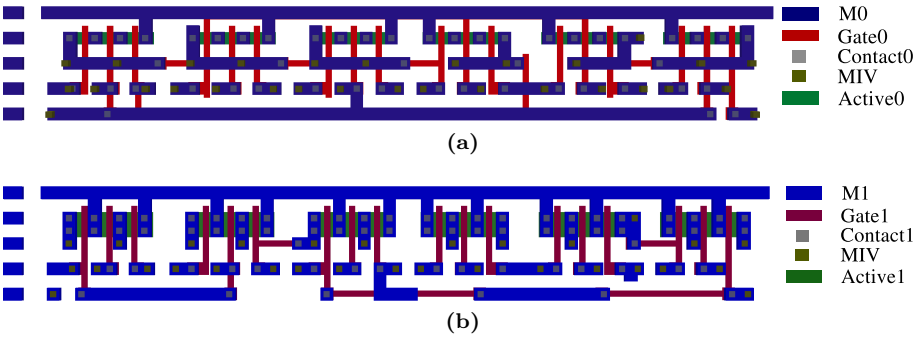


Figure 3.12: Layout view of the ideal KTH-S3D DFFSR cell (with allowed stacking of successive vias) in (a) the bottom tier and (b) the top tier.

Table 3.6: Performance comparison between the KTH-S3D inverter cell and its 2-D in-house implementation

Performance Metric	KTH-S3D INV	2-D INV
rise time†	5.3 ns	5.1 ns
fall time†	4.9 ns	4.2 ns
cell height	6 tracks	9 tracks

† measured for a 10 ns input ramp and a 500 fF load

The time-domain operation of the DFFSR cell is shown in Fig. 3.13. A comparison between the delay of the KTH-S3D inverter cell and its 2D in-house implementation has been also carried out. The results are shown in Table 3.6 and indicate that the 2D implementation outperforms the S3D cell, due to the latter's larger interconnect parasitics. Similar observations for the time-domain behavior of S3D cells were also made in [35]. However, the higher driving strength of the Ge pFETs as compared to Si ones (see Fig. 3.5) minimizes the impact of the increased parasitics in the rise-time, resulting in nearly equal values between the two cells.

3.3 S3D Predictive PDK

The main purpose of the S3D-KTH PDK is to support the development of the in-house S3D process. However, to explore applications that could benefit from S3D integration, as well as to identify and overcome potential challenges, the study needs to be decoupled from the capabilities and restrictions of the in-house process. Towards this, a S3D predictive PDK (PPDK) has been also developed. Based on [23], in which S3D integration can occur on existing process technologies, the S3D PPDK is built upon a commercial 150 nm bulk process with one poly-silicon

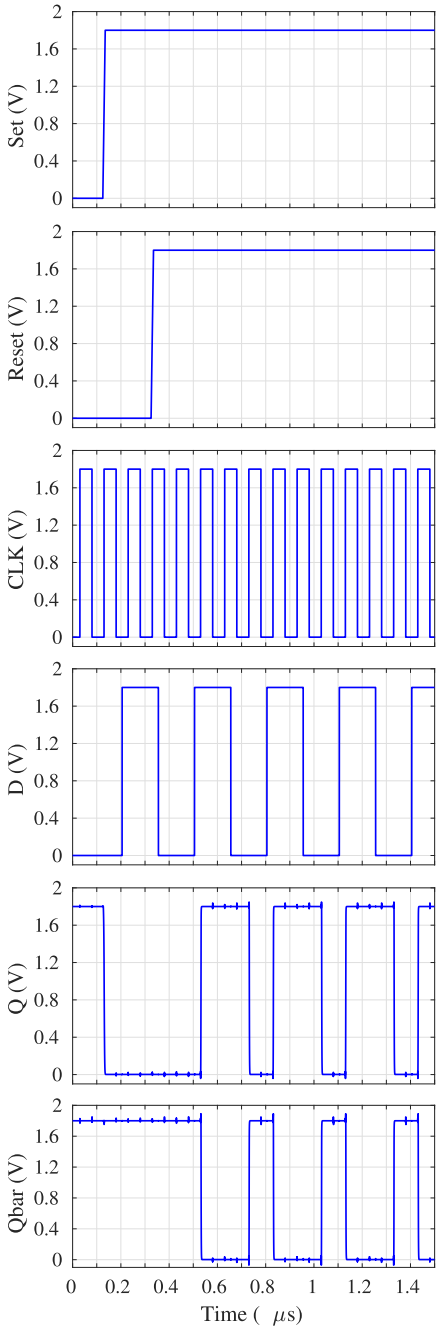


Figure 3.13: Time-domain operation of a KTH-S3D DFFSR cell

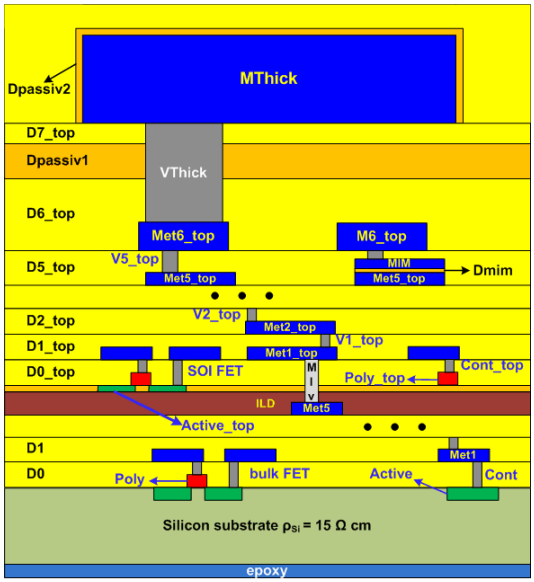


Figure 3.14: Process stack of the S3D PPDK.

and 6 metal layers (1P6M). This bulk process serves as the bottom tier and is referred to as the "base process". A detailed description of the considered S3D process stack is provided in subsection 3.3.1. Extending an existing process to a S3D one enables the use of many of its devices and its design rules. Circuits that have been previously designed in the "base process" could be also reused, at a schematic level of-course. Although the considered base process is not a state-of-the-art CMOS node, it is ideal for AMS applications, thanks to the high intrinsic gain A_v of its transistors and their negated Short-Channel Effects (SCEs), typical of long channel devices. Furthermore, it comes with a wide range of passive devices (resistors, capacitors, diodes and inductors). All these features, along with its low processing cost (compared to more advanced nodes) makes it suitable for exploring the potential of S3D RF/AMS circuits and systems. In addition, any research finding that is based on the S3D PPDK could be projected to more advanced S3D processes. In the following sub-sections, the main modules of the predictive S3D-150nm PPDK will be described: the process stack, the available devices and their models, the design rules and the parasitic extraction flow of the S3D-150nm PPDK.

3.3.1 S3D PPDK Process Stack

The considered process stack for the S3D PPDK is shown in Fig. 3.14. An epoxy layer is considered in the bottom of the stack, as an adhesive layer to the package.

The base process' FEOL is considered for the bottom tier transistors and five tungsten metal layers (Met1-Met5) are assumed for the bottom tier interconnects. This is in contrast to the in-house S3D process, in which only one metal layer is used. The higher number of bottom tier interconnects results in improved connectivity and consequently, higher flexibility in the type of circuits that can be implemented there. Over Met5, a 500 nm thick SiO_2 layer is deposited, which serves as the ILD. A 25 nm active layer is transferred and bonded on top of the ILD, to be used for the processing of top-tier FETs. The same gate-stack is assumed for both the top and the bottom tier transistors. The base process's BEOL (6 Al lines) is used for the top tier interconnects. A MIM layer is sandwiched between Met5_top and Met6_top for the formation of plate capacitors with the underlying Met5_top. The top tier metalization consists also of a thick metal layer (MThick) that is reserved only for the patterning of inductors. No routing is allowed on this layer, in accordance with the design rules of the base process. To suppress electro-migration in MThick, aluminum enriched with copper has been considered for this layer. A special passivation scheme is used above and below MThick, as illustrated in Fig. 3.14 (Dpassiv1, Dpassiv2). All vias and contacts in both tiers are filled with tungsten. The main features (i.e. thickness, material, etc) of the S3D PPDK's interconnects and dielectrics are summarized in Table 3.7. An important difference between the KTH-S3D process and the S3D PPDK is that all the dielectric layers in the latter are considered planarized.

3.3.2 S3D-150 nm PPDK Devices and PCells

In addition to the base-process devices, the S3D PPDK includes also top-tier FETs. More specifically, the devices available in the S3D PPDK are:

- **Bottom tier FETs**. The PCells of the base process FETs are re-used for the S3D PPDK. The bottom tier FETs are bulk transistors, both n- and p-type with minimum dimensions of L= 150 nm and W= 350 nm and EOT= 3.3 nm. Two different types for the bottom-tier transistors are available: (a) with standard threshold voltage, V_T, for typical applications and (b) with high V_T for low-leakage applications.

- **Top tier FETs**. New PCells had to be created for the top tier FETs. Their minimum dimensions have been set equal to their bottom-tier counterparts. The thickness of the Si-active region is T_{Si}= 25 nm and the same EOT as for the bottom tier transistors has been considered. The PCells for the top tier FETs enable the automatic generation of contact arrays, as well as the layout generation of multiple finger transistors with automated generation of connections between the multiple drains/sources/gates.

- **Diodes**. A bottom tier implementation of the base process diodes has been adopted in the S3D PPDK. The minimum length and width of these diodes is L = W = 350 nm.

Table 3.7: Characteristics of the S3D PPDK considered process stack

Routing Layer	**Material**	**Thickness [nm]**	**Sheet resistance [Ω/□]**
Poly	Sd Poly-Si [1]	250	10
Poly_SBlk	N-Sd Poly-Si [2]	250	400
Met1 - Met5	W	400	170m
Poly_top	Sd Poly-Si [1]	250	10
Poly_SBlk_top	N-Sd Poly-Si [2]	250	400
Met1_top - Met5_top	Al	400	70m
MIM	Ti-Nitride	100	8
Met6_top	Al	900	30m
MThick	Al [3]	6100	4m
Dielectric Layers	**Material**	**Thickness [nm]**	**Dielectric Constant ϵ_r**
D0 [3]	-	800	4.2
D1 - D4	-	900	4.2
ILD	-	900	4.2
D0_top	-	800	4.2
D1_top - D5_top	-	900	4.2
D6_top	-	1800	4.2
Dmim	-	70	8.0
Dpassiv1	-	500	6.5
D7_top	-	300	4.2
Dpassiv2	-	120	6.5

[1] silicided poly-silicon
[2] non-silicided poly-silicon
[3] planarized

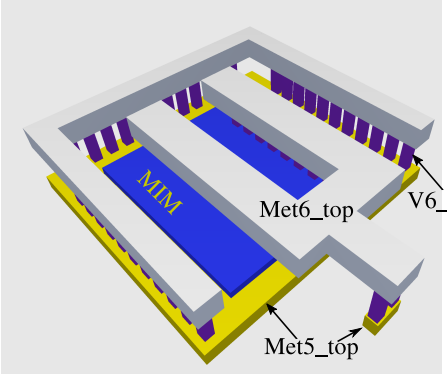


Figure 3.15: 3-D layout view of a S3D PPDK MIM capacitor.

- **Resistors**. Four different resistor types are available in the S3D PPDK, two in the top tier and two in the bottom tier. For high sheet resistance values (400 Ω/□), the S3D PPDK resistors are implemented with non-silicided poly-silicon layers, whereas silicided poly-silicon layers are preferred for low sheet resistance values (10 Ω/□).

- **Capacitors**. Apart from MOS capacitors, which can be readily obtained by short-circuiting the source and drain transistor terminals, base-process MIM capacitors are also included in the S3D PPDK. They are implemented in the top-tier between Met5_top and a thin (100 nm) Ti-nitride layer (MIM layer). A 70 nm insulator layer (Dmim) with a dielectric constant ϵ_{Dmim}= 8 yields a unit area capacitance of 1 fF/μm^2. The minimum dimensions for the MIM capacitors are L=W= 5μm, so the minimum obtained capacitance is ≈ 25 fF. The MIM capacitors' PCells enable the automatic generation of vias from the MIM layer to Met6_top. A 3D view of a MIM cpacitor in the S3D PPDK is shown in Fig. 3.15.

- **Inductors**. The S3D PPDK contains also a number of inductors, with the same topology as the base process inductors. These cells are implemented in MThick and cover the frequency range 1-10 GHz. However, the S3D PPDK inductors have a pre-defined topology and so, they cannot be parameterized. As a trade-off, they can be modeled by accurate spice models with no need for carrying out electromagnetic simulations.

3.3.3 S3D-150nm PPDK Device Models

Spice models are included in the S3D PPDK for the following cells:

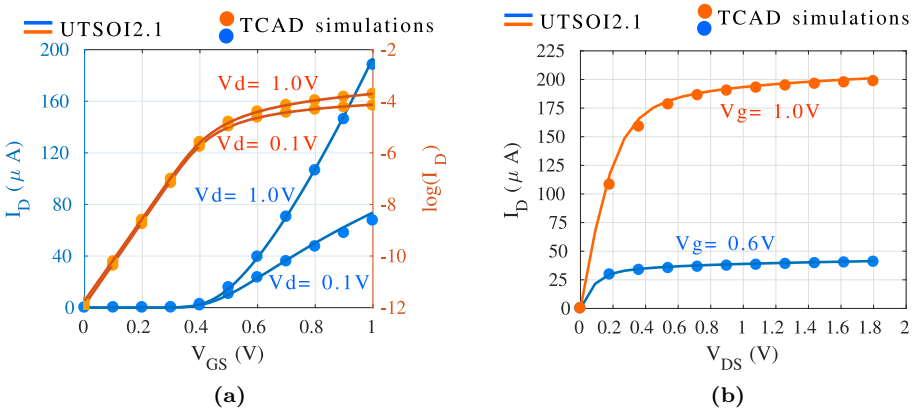


Figure 3.16: Calibrated UTSOI2.1 model and TCAD simulations for the transistor (a) transfer and (b) output characteristics. The plots refer to a n-type FET with L= 250 nm and W= 1 μm

Bottom-tier FETs. In [69] it has been shown that, as long as the processing temperature of the top tier does not exceed 500°C, the impact on the performance of bottom-tier transistors can be considered insignificant. Therefore, in this work, the base-process BSIM3v3 models have been employed for the bottom tier transistors with no alterations.

Top-tier FETs. The UTSOI2.1 model [99, 100] has been used for the top-tier FETs, given their SOI topology, as it was the case for the KTH-S3D PDK. The lack of fabricated short-channel transistors led to the calibration of the model's parameter deck based on TCAD simulations of SOI FETs with T_{Si}= 25 nm and EOT= 3.3 nm. The transfer and output characteristics obtained by the UTSOI2.1 model demonstrate good agreement with the TCAD results, as shown in Fig. 3.16. Since the main purpose of the S3D PPDK is the exploration of applications of S3D integration, other than digital, the devices' noise models have been also included in the PDK. In general, the noise behavior of SOI devices is complicated, with both the front- and back-gate interfaces contributing to 1/f noise [108]. As a first approximation, the noise parameters of the top-tier FETs have been leveraged to match the input noise of bottom tier FETs with the same biasing, as shown in Fig. 3.17.

Inductors. Migrating the base-process inductors to the S3D-150nm PPDK results in an increased distance between the inductor and the silicon substrate, requiring updates in the inductor spice models. As a first approximation, the spice model parameters related to the parasitic capacitance between the inductors' coils and the silicon surface were scaled proportionally to the increase in the inductor-silicon distance. However, full electromagnetic simulations need to be carried out, if higher accuracy is desired.

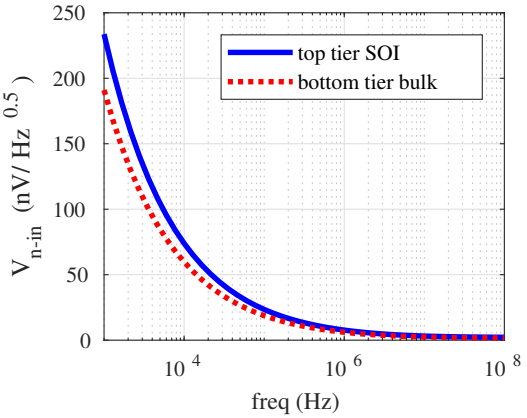


Figure 3.17: Input referred noise voltage, V_{n-in} of a nFET in the top and bottom tier. Both devices have the same dimensions and biasing: L= 150 nm, W= 10 μm, I_D=200 μA and V_{DS}= 0.5 V

Table 3.8: Basic design rules for S3D-150nm PPDK layers

	Layer	Min. Width μm	Min. Spacing μm
Bottom Tier	Active	0.4	0.4
	Poly	0.15	0.3
	Cont	0.2	0.3
	Met1 - Met5	0.3	0.3
	Via1-Via4	0.25	0.25
Top Tier	Active_top	0.4	0.4
	MIV	0.3	0.3
	Poly_top	0.15	0.3
	Cont_top	0.18	0.3
	Met1_top - Met5_top	0.3	0.3
	Via1_top - Via5_top	0.25	0.25
	MIM	1.5	5
	Met6_top	0.6	0.6
	MThick	6	6
	VThick	3	3

Table 3.9: Comparison between the parasitic capacitance obtained with the S3D-150 nm PPDK's parasitic extraction flow and Momentum EM simulations (by Keysight®)

Layout Structure	EM Simulations	QRC Results with FS	QRC Results without FS
single Met5 - Met1_top via stack [Fig. 3.18(a)]	0.48 fF	0.5 fF	0.12 fF
Met5 - Met1_top via stacks [Fig. 3.18(b)]	0.67 fF	0.6 fF	0.24 fF
Met5 - Met1_top Fig. 3.18(c)	0.1 fF/μm	0.1 fF/μm	0.09 fF/μm
MIV - Cont_top Fig. 3.18(d)	0.18 fF	0.12 fF	0.06 fF
M5 - Active_top Fig. 3.18(e)	0.10 fF	0.10 fF	0.08 fF

3.3.4 S3D PPDK Physical Verification

Cadence® PVS has been employed for the S3D PPDK DRC and LVS checks, as it was the case with the KTH-S3D PDK. The PPDK's design rules have been based on the corresponding ones from the base process. Since the same photolithography equipment is assumed for both the bottom and top tiers, the minimum dimensions and spacing rules do not alter between the two tiers. The clearance between MIVs and Poly_top/Active_top/Cont_top was set equal to the base process minimum Cont-Cont spacing (0.3 μm). An overview of the most important S3D-PPDK's design rules is given in Table 3.8.

3.3.5 S3D-150 nm PPDK parasitic extraction flow

To set up the S3D PPDK's parasitic extraction flow, the same methodology as in the KTH-S3D PDK has been adopted. The MIM, VThick and MThick layers have been excluded from parasitic extraction, since the layout parasitics associated with these layers are already included in the corresponding device models (MIM capacitors and inductors). EM simulations (with Keysight® 's Momentum simulator) have been employed to validate the accuracy of the QRC extraction flow for the structures shown in Fig. 3.18. The aspect ratio (AR) of the MIVs in the S3D PPDK is larger than in the in-house process (AR $\approx$ 5), which implies that the the parasitic resistance induced by a MIV is not negligilbe. In fact, EM simulations indicate that the resistance of a via stack from Met1 to Met1_top is 5 Ω. The results are summarized in Table 3.9. It appears that the FS assisted QRC extraction yields higher accuracy for the S3D PPDK than the KTH-S3D PDK. Furthermore, disabling the FS limits the credibility of the extraction flow significantly and hence,

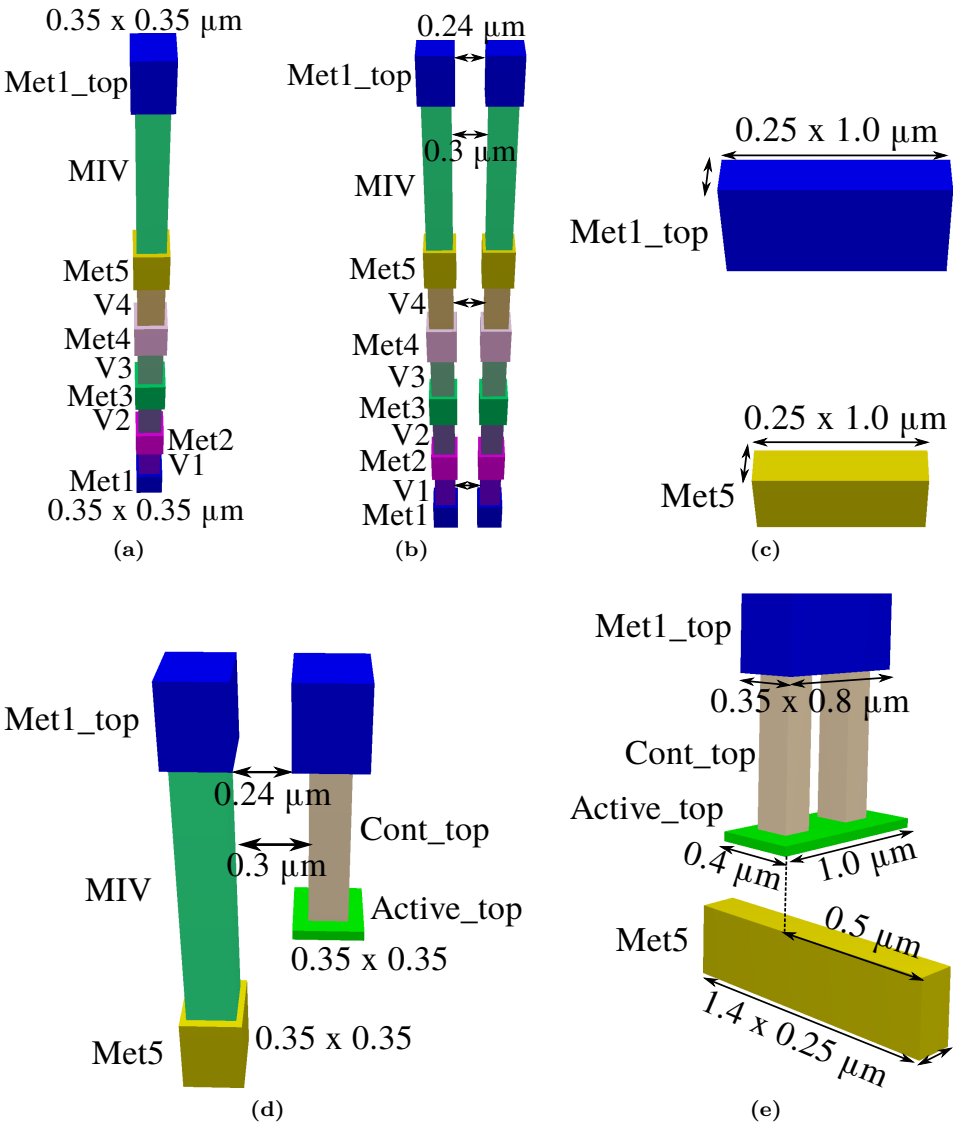


Figure 3.18: Layout structures to extract the parasitics of: (a) a single Met5 -> Met1_top via stack (b) the coupling between two neighboring Met5 -> Met1_top via stacks (c) the coupling between a Met1_top and a Met5 line (d) the coupling between a MIV and Cont_top (e) the coupling between a Met5 line and an Active_top region.

it should be avoided as the layout dimensions shrink.

3.4 Summary

Chapter 3 provided an overview of the developed S3D PDKs for custom ICs: (a) one PDK to support the development of the in-house S3D process and enable circuit-design in it (KTH-S3D PDK) and (b) one PDK to support the exploration of S3D custom circuits (S3D PPDK). These two S3D design platforms cover a void in state-of-the-art research trends that are totally skewed towards digital applications of S3D integration. A novel parasitic extraction flow for S3D ICs has been developed and validated against EM simulations, highlighting the importance of the FS assisted parasitic extraction for small layout features. Special emphasis has been given to the calibration of the device models against both measurement results and TCAD simulations. All in all, both S3D PDKs enable a thorough exploration of S3D custom ICs.

Chapter 4

Sequential 3D Receiver Front-End

This chapter focuses on applications that could benefit from the S3D integration technology. The prospect of area-efficient RF/AMS circuits is identified and as a proof of concept, a S3D receiver front-end is designed in the S3D-PPDK, which was introduced in Chapter 3.

4.1 Applications of S3D integration

4.1.1 State-Of-The-Art

Digital applications have been the main driving force behind S3D integration. Consequently, significant research endeavors have been directed towards novel design methods for S3D digital applications. These methods concern mainly the circuit/system partitioning into two or more device tiers and they can be grouped into three broader categories: (a) transistor-level, (b) gate-level and (c) block-level. Transistor-level S3D designs are based on S3D digital cells, in which the pFETs and nFETs are implemented in separate tiers. A major benefit of this approach is that it incurs an one-time engineering effort to create the library of the S3D digital cells. All the available place-and-route tools and methodologies can be practically reused. The height of the S3D cells cannot be halved with respect to the 2-D ones, due to the area overhead introduced by the MIVs [91]. Thus, this design approach cannot offer 50% area gains compared to the original 2-D implementations. It is important to note that apart from the obvious performance and miniaturization benefits, larger area reduction could improve the throughput of a S3D process, which in turn can compensate the high S3D fabrication costs at wafer level. Another drawback of transistor-level designs is that they are prone to routing congestion [109]. More specifically, the footprint reduction achieved through sequential 3D integration, coupled to the same interconnect dimensions as the 2-D case, result in reduced space for routing. Two solutions have been proposed to overcome routing congestion: (a) adding more metal layers in the bottom-tier [110] and (b) trading-off the reduction in the S3D cells' height with more space for the interconnects [35]. However, the

former solution requires additional processing steps, leading to elevated production costs, whereas the second one tends to cancel out the area reduction benefits offered by S3D integration.

On the other hand, in gate-level S3D designs, both nFETs and pFETs are available in each tier and the stacking occurs at the cell-level, i.e. a NAND gate over a NOR gate. Unlike the case of transistor-level S3D designs, gate-level ones require novel place-and-route tools with the capability of allocating each cell to a specific device tier. Various such place-and-route algorithms have been proposed [30–32]. Assuming balanced area utilization among the device tiers, the resulting area reduction reaches 50% [111] , contrary to transistor-level designs, the area gains of which are limited by the MIV overhead [35,91]. However, none of these works considers the performance mismatch between the top and bottom tier devices, caused by the different fabrication process utilized in each: a high-thermal budget process for the bottom-most tier and a low-thermal budget one for the rest. To accommodate this performance skew between the device tiers, additional engineering effort at the CAD/EDA level is necessary. For instance, two cell libraries need to be developed, for the top and bottom tiers respectively. The number of necessary cell libraries increases with the number of device tiers. Furthermore, using these tier-related cell libraries could reduce the complexity of the place and route tools, by handling the tier-allocation for each cell during logic synthesis. For example, the synthesis tool can select the optimal tier implementation for each cell, based on timing/power constraints. The performance skew between the stacked tiers was taken into consideration in [112], in which a 10-20% drop in the on-current of the top-tier devices, as compared to the bottom tier ones was assumed. However, the impact of low-temperature processing on top-tier devices appear to be more complicated than just a simple scaling of the on-current [113]. For instance, the lack of a "reliability anneal" can cause up to 60 mV shift in the threshold voltage of top-tier devices, $V_{T_{top}}$, if the same gate-stack as in the bottom tier is used.

S3D partitioning at block-level can better handle the performance variation between the stacked tiers. This can be achieved in a twofold way by (a) allocating blocks that can better handle performance degradation in the slowest tier and (b) modifying the size of blocks allocated in the slowest tier to counteract their performance drop [29]. Although block-level S3D designs do not exploit the ultra-low MIV pitch, they can achieve an area reduction up tp 50%, assuming balanced area occupation in the stacked tiers. Furthermore, additional cost benefits can be expected, when a low-cost, specialized process is employed for the top tier [114].

4.1.2 Motivation for S3D RF/AMS circuits

Block-level partitioning can potentially offer higher area gains than transistor-level S3D designs, while at the same time handling efficiently the performance skew between stacked tiers. Hence, they offer the most appealing option for reliable S3D ICs. Furthermore, the use of specialized processes in the top tier could also minimize the performance skew between top and bottom tier devices, or even revert

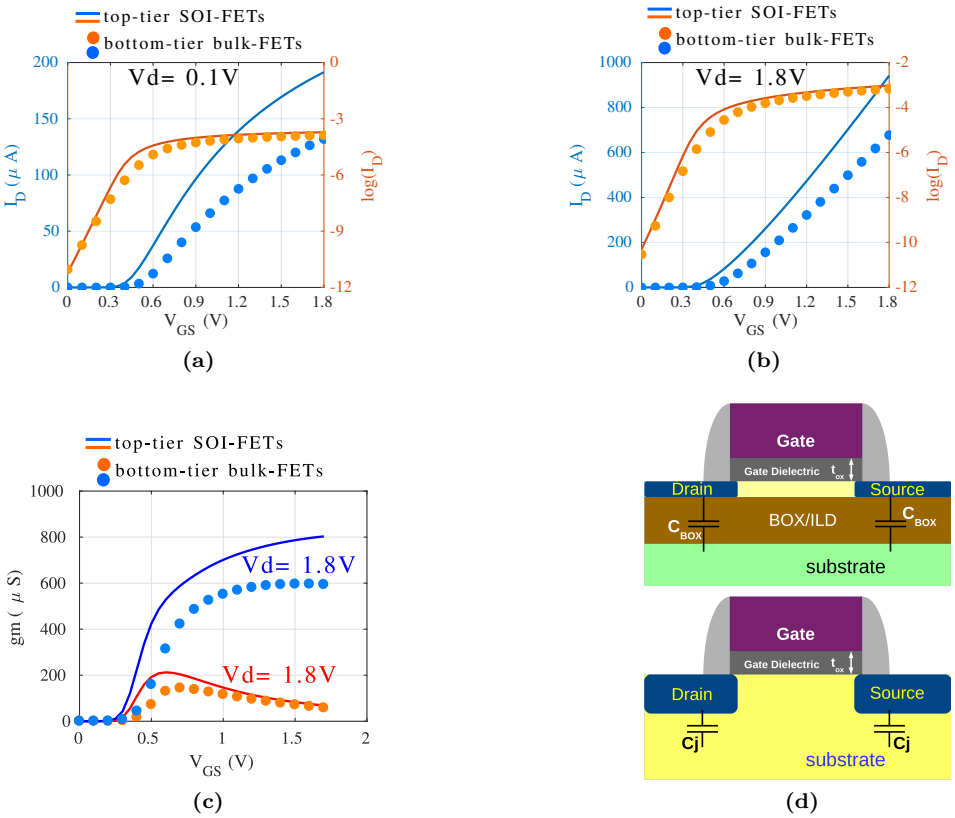


Figure 4.1: Id versus Vgs for (a) V_{DS} = 0.1 V and (b) V_{DS} = 1.8 V. (c) Transconductance-gm versus Vgs for V_{DS} = 0.1 V and 1.8 V. (d) Structures of the top-tier FDSOI (top structure) and bottom tier bulk (bottom structure) FETs. Note that for visual purposes, the dimensions of the various device regions are not scaled proportionally

it for specific applications. Identifying such applications is the scope of this section. The coming analysis will be based on the S3D PPDK.

The SOI nature of the top-tier devices offers a number of performance benefits over the bulk transistors in the bottom tier. For instance, the top-tier transistors of the S3D PPDK exhibit 40% higher on-current than the base-process transistors, as shown in Fig. 4.1(a-b) (945 μA vs 645 μA). The higher on-current of the top-tier FETs is linked to their higher transconductance, gm, (see Fig. 4.1(c)), which in turn translates to a better noise performance compared to bottom tier transistors:

$$\bar{V_{n,in}}^2 = \frac{4kT\gamma}{gm} \tag{4.1}$$

with ${V_{n,in}}^2$ referring to the input-referred thermal noise of the transistors.

Furthermore, SOI transistors exhibit very low source/drain junction capacitances, C_j, thanks to the presence of the BOX (ILD for top-tier FETs) beneath the active regions (Fig. 4.1(d)). The capacitance across the BOX of the top-tier transistors, C_{BOX}, is negligible, thanks to the relatively thick ILD (500 nm to the top-surface of Met5). The combination of higher on-current and lower device parasitics for the top-tier FETs yields better performance at high frequencies over their bottom tier counterparts. This is clearly illustrated by a 16% increase in the transit frequency, f_T, of top tier FETs compared to the bottom-tier ones (86 GHz vs 74 GHz). As a result, top-tier transistors require less current than bottom tier FETs to achieve a specific bandwidth.

For all these reasons, the top-tier is an ideal option for the implementation of high-frequency blocks in the S3D PPDK [105]. On the other hand, the bottom tier offers more benefits for low-frequency circuits, thanks to the availability of both standard and high V_T transistors. These findings highlight the clear potential for S3D RF/AMS circuits, when such a frequency-based partition scheme is employed. Contrary to the case of digital applications, the benefits that S3D integration offers to RF/AMS circuits are not limited to just area reduction. Instead, performance benefits can be also expected, thanks to the improved performance of the top-tier FDSOI transistors at high frequencies (improved speed, bandwidth and noise performance). However, the previous observations hold only for bulk bottom-tier processes and do not extend to the case with SOI transistors in both tiers. An additional benefit of S3D integration to the performance of RF/AMS circuits, irrespective of the type of the bottom tier process, originates from the increased distance between the top-most thick metal and the silicon substrate. This in turn leads to reduced inductor losses to the substrate, as well as stronger attenuation of the noise coupled from the substrate to the inductors.

Among the various radio applications, low-power/short-range standards (bluetooth, Wireless Personal Area Networks - WPAN) are of particular interest for S3D integration. The main reasons stem from the advent of applications with many interconnected devices, such as Internet of Things (IoTs) and Internet of Medical Things (IoMTs) [115]. These applications can clearly benefit from the system miniaturization achieved through S3D integration. Furthermore, the low-power and short-range specifications of these standards minimize the risk of potential interference between the stacked device tiers, which could adversely affect the operation of the whole system.

A receiver front-end for the ZigBee standard [3] has been selected to serve as a proof of concept for S3D RF/AMS circuits. The circuit has been designed in the S3D-PPDK employing the proposed frequency based partition scheme, in which high frequency blocks were placed in the top-tier and low frequency ones in the bottom. Based on the standard specifications [116], ZigBee receivers should operate either in the 868/915 MHz or in the 2.4 GHz band. The considered receiver is designed for operation at 2.4 GHz, with 16 communication channels, each with a 5 MHz bandwidth.

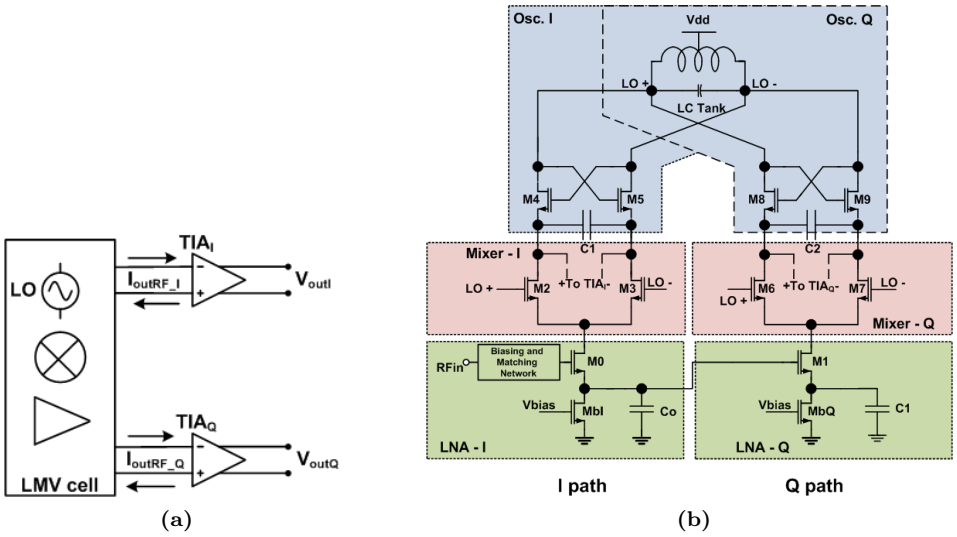


Figure 4.2: (a) Block-level representation of the considered receiver front-end, based on the LMV cell. (b) Schematic of the LMV cell, adopted from [3].

4.2 Design of the S3D receiver front-end

The block-level diagram of the considered front-end is shown in Fig. 4.2(a). The receiver makes use of the LMV cell that has been introduced in [117] and consists of a Low-Noise Amplifier (LNA), a mixer and a Voltage-controlled local oscillator (LO). The LMV cell, which is depicted in Fig. 4.2(b), is based on the current re-use principle to achieve low-power: the Low-Noise Amplifier (LNA), mixer and local oscillator are cascoded so that they all share the same biasing current. Further improvements in area efficiency are possible with S3D integration. A low intermediate frequency (low-IF) architecture has been chosen, thanks to its better handling of the 1/f noise, as compared to a zero-IF architecture. In this design, the RF signal is down-converted to the 1 MHz - 6 MHz frequency band. The operation principles behind the LMV cell, as well as the design choices for each of its consisting blocks are analyzed in sub-section 4.2.1. In short, the LNA converts the antenna signal to an RF current, which is then fed to the mixer and down-converted to the IF. The down-converted current flows to the virtual ground created by the Trans-Impedance Amplifier (TIA), where it is converted to voltage and amplified.

4.2.1 High frequency blocks in the top tier

LNA. A simple common-source (CS) topology with a resistive termination at the input has been employed for the LNA (see Fig. 4.3(a)). This choice has been

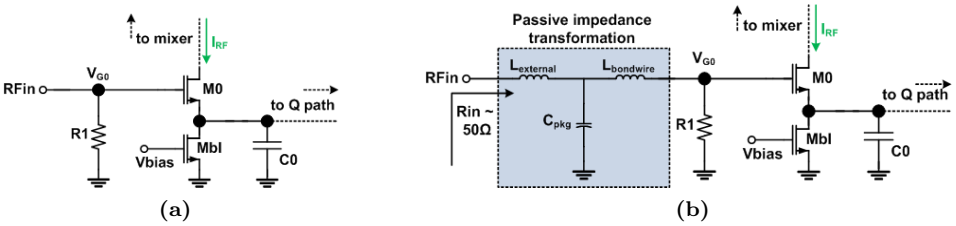


Figure 4.3: (a) CS LNA with simple resistive input termination. (b) CS LNA with a passive impedance transformation network.

motivated by the topology's robust input matching properties and the standard's relaxed specifications for the noise figure, NF <15 dB. An inductive degenerated CS stage could be an alternative implementation for the LNA [117]. However, the required additional inductor would cancel the area-saving motivations in [3].

The CS stage produces a RF current, $I_{RF} = G_{LNA}V_{G0}$, that flows through the mixer of the LMV cell. Due to the capacitive degeneration of C_0, the transconductance of the LNA, G_{LNA} is:

$$G_{LNA} = \frac{g_{m0} j\omega C_0}{g_{m0} + j\omega C_0} \tag{4.2}$$

where g_{m0} is the transconductance of M0. Thus,

$$I_{RF} = \frac{g_{m0} j\omega C_0}{g_{m0} + j\omega C_0} V_{G0} \tag{4.3}$$

with V_{G0} representing the RF signal at the gate of M0.

The main purpose of the capacitor C_0 is twofold: (a) it is used for the quadrature generation, as it will be explained later and (b) it sinks part of the noise generated by the biasing transistor MbI. It should be noted that the topology's biasing current could be also produced by M0, avoiding totally the need for MbI. However, by doing so, the size of M0 should increase, leading to excessive RF losses that could corrupt the performance of the LMV cell [117]. To ensure good matching and low reflection losses, the real part of the topology's input impedance, $Re\{Z_{in}\}$ needs to be matched to the antenna resistance, R_{ant}, i.e:

$$Re\{Z_{in}\} = R1 = R_{ant} \tag{4.4}$$

To simplify the noise analysis of the considered LNA, the noise contribution of the transistor MbI is assumed insignificant. The motivation behind this assumption is twofold: (a) the transistor MbI acts as the biasing of the LMV cell, hence its transconductance is typically low and (b) part of the noise generated by MbI is sinked by C_0. Under this assumption,

$$NF = \frac{SNR_{in}}{SNR_{out}} \tag{4.5}$$

$$NF = (1 + \frac{R_{ant}}{R1})(1 + \frac{\gamma}{g_{m0}R_{ant}} + \frac{\gamma}{g_{m0}R1}) \tag{4.6}$$

where R_{ant} represents the antenna impedance. For perfect matching ($R1 = R_{ant}$),

$$NF = 2 + \frac{4\gamma}{g_{m0}R_{ant}} \tag{4.7}$$

which implies that $NF_{min} \approx 3$ dB. A drawback of the resistive termination at the input of the LNA is that for the case of perfect matching, the voltage at the gate of M0, V_{G0}, is attenuated by a factor of 2, as compared to the value received at the antenna. To cancel, or even reverse this attenuation, a passive impedance transformation network (PITN), such as the one shown in Fig. 4.3(b), is used. This PITN consists of an external inductor, $L_{external}$, along with the inductance of the bondwires ($L_{bondwire} \approx$1.7 nH) and the capacitance associated with the package connections ($C_{pkg} \approx$ 16fF). By using a PITN, R1 can be set independently of R_{ant}. The matching specifications can be then met by calibrating the value of $L_{external}$, so as to guarantee that the impedance seen from the antenna, R_{in}, is matched to R_{ant}. The use of the PITN results in voltage amplification, A_{PITN}, from the antenna input to the gate of M0. Assuming perfect matching ($R_{in} = R_{ant}$), the obtained voltage gain and noise figure are [118]:

$${A_{PITN}}^2 = \frac{R1}{4R_{ant}} \tag{4.8}$$

$$NF = \frac{16}{g_{m0}R1} + 2 \tag{4.9}$$

It is clear that higher values of R1 result in lower NF and improved gain. On the other hand, the increased gain limits the linearity of the receiver. Larger values of R1 are impractical as they result in insufficient impedance matching.

Quadrature Generation. In the considered topology, quadrature generation is obtained at the RF signal path, through the capacitor C_0 and an identical LMV cell [3], as shown in Fig. 4.4. Indeed, based on (4.3), the voltage sensed at the gate of the Q path's LNA, V_{G1}, is

$$V_{G1} = \frac{I_{RF}}{j\omega C_0} \tag{4.10}$$

Assuming that C_1 is so large that it practically grounds the source of M1 at the input frequency, the current through M1 is:

$$I_{RF\_Q} = g_{m1}\frac{I_{RF}}{j\omega C_0} \tag{4.11}$$

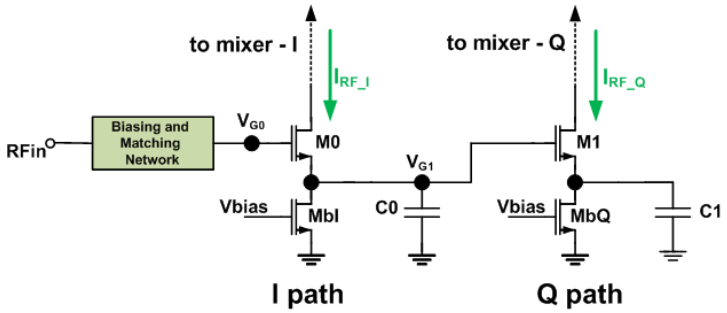


Figure 4.4: Quadrature generation through C0 and M1.

Combining (4.11) and (4.3),

$$I_{RF\_Q} = g_{m1}\frac{g_{m0}}{g_{m0} + j\omega C_0}V_{G0} \tag{4.12}$$

Since the devices in the I and Q paths are identical, $g_{m0} = g_{m1} = g_m$. A comparison between (4.3) and (4.12) indicates that the phase difference between I_{RF} and I_{RF_Q} is always 90^o. For an efficient image rejection however, the amplitude of I_{RF} and I_{RF_Q} should be also equal, which occurs when

$$g_m = \omega C_0 \tag{4.13}$$

In other words, quadrature generation is efficient only for a narrow signal bandwidth. However, the relaxed image rejection specifications of the ZigBee standard make such an implementation advantageous in terms of both area and power [3].

Mixer. An active mixer topology consisting of the transistors M2-M3 (M6-M7 for the Q-path) is cascoded over the LNA, as shown in Fig. 4.5. A decoupling capacitor is placed between the LO signal and the switching transistors M2-M3, else the large DC level of the LO (equal to the supply V_{DD}= 1.8 V) could drive M2-M3 to the triode region. The capacitor C1 is sized so that it presents a high impedance at the IF. Thus, the down-converted current I_{IF} flows into the virtual ground formed by the TIA. Assuming no parasitics losses and ideal abrupt switching, the amplitude of I_{IF} is

$$I_{IF} = \frac{2}{\pi}I_{RF} \tag{4.14}$$

The noise contribution of the mixer transistors M2 and M3 to the output current can be approximated by [118]:

$$I_{n,M2-3,out} = \frac{I_{BIAS}}{\pi V_{LO,p}}V_{n,M2-3} \tag{4.15}$$

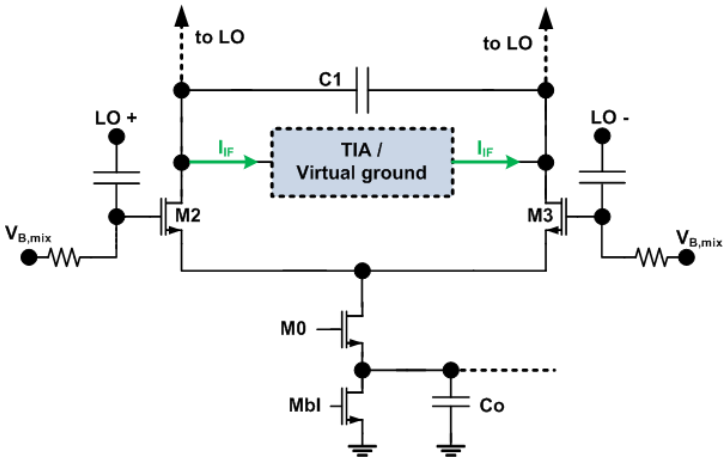


Figure 4.5: Mixer sub-block in the LMV cell.

where, I_{BIAS} is the biasing of the LMV cell, set by the transistor MbI, $V_{LO,p}$ is the peak of the LO's swing and $V_{n,M2-3}$ is the input referred noise of the mixer transistors (both thermal and 1/f). Equation (4.15) suggests that, because of the cascoding of the LNA, mixer and VCO in the LMV cell, there is little room in leveraging the performance of the mixer without degrading the performance of the LNA and/or mixer. Indeed, the noise contribution of M2-3 can be reduced by lowering the biasing current of the cell, which in turn decreases the LNA gain. Thus, the efforts to improve the mixer's noise performance were restricted to reducing the flicker noise of M2-3. After all, because of the stacking of the LO over the mixer, the mixer's flicker noise can be up-converted to higher frequencies, leading to excessive phase noise for the LO. To lower the 1/f noise of the mixer, wider transistors have been employed.

LO. The LO has been implemented as a cross-coupled pair consisting of transistors M4-5 for the I-path and M8-9 for the Q, as shown in Fig. 4.2. This is achieved by sizing the capacitors C1 and C2 so that they present a very low impedance at f_{LO}, effectively shorting together the sources of M4-5. The resulting oscillation frequency, f_{LO} equals 2.4 GHz. A center-tap inductor from the base process, with a total inductance of 1.8 nH and a quality factor of Q=17 at 2.4 GHz has been employed for the LC tank, thanks to its area efficiency compared with two separate inductors. The absence of a varactor in the S3D-PPDK has led to the use of a simple MIM-capacitor in the LC-tank, sacrificing thus any tuning capability. This, however, has been considered sufficient for the scope of this design, i.e. to serve as proof of concept for S3D RF/AMS circuits.

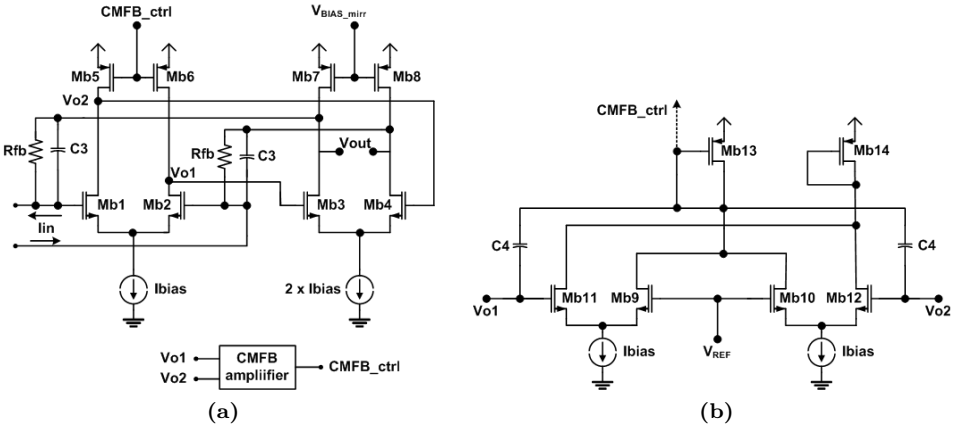
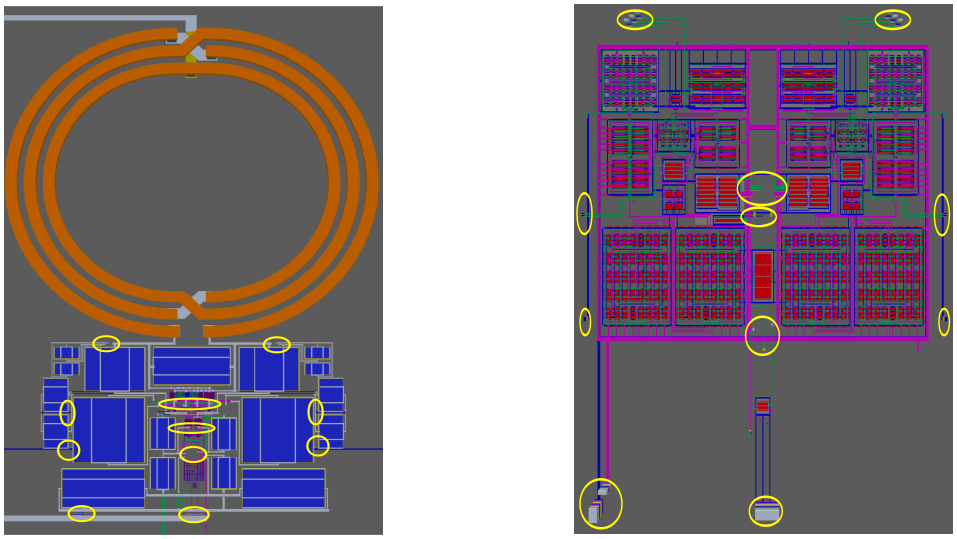


Figure 4.6: (a) Schematic of the TIA. (b) Schematic of the CMFB circuit.

4.2.2 Baseband blocks in the bottom tier

The main role of the TIAs is to convert the IF current generated by the mixer to voltage and amplify it. Two identical TIAs are designed, one connected to the output of the I-path (TIA_I) and the other to the Q path (TIA_Q). Their outputs can be further processed (i.e. additional amplification, image rejection, analog-to-digital conversion, etc). However, unlike the case in [3] where the complete receiver chain has been implemented, in this work, only the receiver front-end is designed, as shown in Fig. 4.2(a). In other words, the S3D receiver does not perform any further amplification or image rejection, since it does not include variable gain amplifiers (VGAs) or complex filters respectively. Both the TIA_I and TIA_Q are placed in the bottom tier, except for their capacitors (in the S3D PPDK, the MIM layer is available only in the top tier).

The schematic for the designed TIAs is shown in Fig. 4.6(a). They consist of two differential pair stages. For the frequency compensation of the amplifier, the phantom-zero method is employed by connecting the capacitor C_3 in parallel to the feedback resistor Rfb. The differential configuration of the TIAs (see Fig. 4.6(a)) requires a common-mode feedback (CMFB) circuit to control the common-mode voltage at the output of each differential stage. The schematic of the CMFB circuit is shown in Fig. 4.6(b). It senses the terminals Vo1 and Vo2 from the first stage and compares them against a reference voltage (V_{REF}). The generated output controls the biasing of the first stage's active load (transistors Mb5-Mb6). Simulations have shown that setting the common mode voltage at the output of the first stage is enough to control the common mode at the output of the second stage as well.



(a) (b)

Figure 4.7: (a) Layout of the designed S3D receiver in (a) the top and (b) the bottom tier. The location of MIVs are highlighted with yellow circles

4.3 S3D layout of the receiver front-end

This section describes the layout of the S3D receiver. The layout of each tier is shown in Fig. 4.7, while a 3-D view of the S3D receiver is shown in Fig. 4.8. As already mentioned, high-frequency blocks were placed in the top-tier, along with the capacitors from the low-frequency blocks. The rest of the low-frequency blocks were placed in the bottom tier. The resistors of both the high- and low- frequency blocks have been also placed in the bottom tier, no matter the block they belong to. This choice has been motivated by the need to reduce the size-skew between the top and bottom tier layouts, towards improved area efficiency. This size skew is owed to the relatively large area of the inductors and capacitors. Indeed, the area occupied in the top-tier equals 0.12 mm^2, with 0.09 mm^2 occupied just by the inductor. On the other hand the size of the bottom tier blocks equals 0.02 nm^2. Therefore, the total area of the S3D receiver is dictated by the the top-tier (0.12 mm^2). No bottom-tier blocks were placed under the inductors, to avoid any coupling between them.

All of the MIV locations are highlighted with yellow circles. For the four interfaces between the high- and low-frequency blocks (two at the output of the I mixer and two at the output of the Q mixer), a cluster of 36 MIVs has been employed, with the goal to reduce the parasitic resistance between the two blocks. Despite the large count of MIVs, their impact on the parasitic capacitance in the mixer

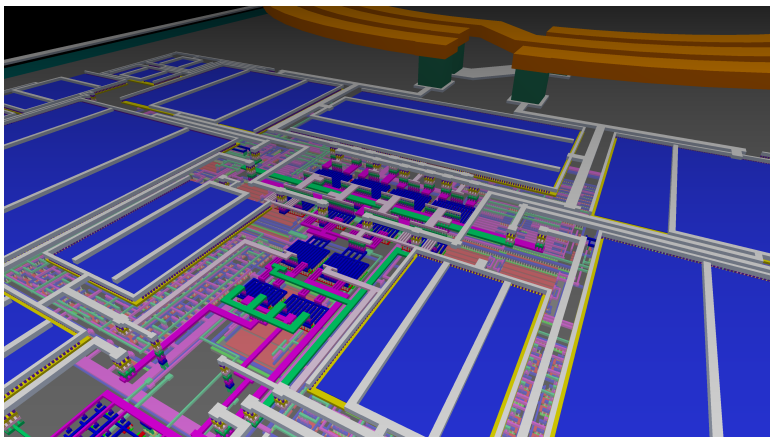


Figure 4.8: 3-D view of the S3D receiver front-end. Top tier blocks are shown with opaque colors and bottom tier with transparent.

output is rather small, approximately 100 fF. This parasitic capacitance needs to be low, to avoid any significant degradation of the LMV's gain [117]. For the rest of the connections between the top and bottom-tier, i.e. the connection to the compensation capacitors in the TIAs, the MIV count is smaller.

4.4 Performance of the S3D receiver front-end

Post-layout simulations have been used to evaluate the performance of the designed S3D receiver. A field-solver assisted parasitic extraction was carried out because of the higher accuracy it offers. For a more reliable simulation setup, the I/O pads of the base process were included at the input, supply and ground connections. Furthermore, as mentioned before, the package capacitance and bond-wire inductance have been also included in the simulation setup as part of the input matching network. The most important simulation results are plotted in Fig. 4.9(a)-(d).

The reflection losses at the input are quantified with the S_{11} parameter, which is plotted in Fig. 4.9(a). Overall, at the frequency of interest (2.4 GHz) $S_{11} \approx$ -17 dB. The gain at the output of the TIA is plotted in Fig. 4.9(b). The S3D receiver chain achieves a gain of 34 dB and a bandwidth of 19 MHz. The simulated bandwidth is larger than the bandwidth of a communication channel in the ZigBee standard (5 MHz). However, it can be corrected with the addition of complex filters in the S3D receiver. Fig 4.9(c) plots the phase noise of the LO. At a frequency offset of 3.5 MHz, the resulting phase noise equals 120 dBc/Hz. The double-sideband NF of the receiver is shown in Fig. 4.9(d), which, when integrated over the IF band results in 11 dB. The addition of the VGA and the complex filter in the receiver

Table 4.1: Simulated performance of the S3D receiver front-end

	S3D receiver	**[3]**	**Specifications** [a]
Technology	S3D 150 nm	90 nm	-
Area	0.120 mm^2	0.186 mm^2	-
Supply	1.8 V	1.2 V	-
Power	3.6 mW	3.6 mW [b]	-
Gain	34 dB [c]	75 dB	65 dB
NF	11 dB	12 dB	$<$ 15.5 dB
IIP3	-17.5 dBm	-12.5 dBm	$>$ -32 dBm
Phase Noise @ 3.5 MHz offset	-120 dBc/Hz	-108 dBc/Hz	$<$ -102 dBc/Hz
LO leakage at the input	-75 dBm	-60 dBm	-
S_{11}	-17 dB	-13 dB	$<$ -10 dBm

[a] As extracted from [3].
[b] It refers to the complete receiver chain with the VGAs and the complex filters.
[c] The S3D receiver lacks the VGAs and complex filters.

chain is not expected to affect the NF significantly, thanks to the gain of the LMV cell and TIAs.

A summary of the S3D receiver's performance is provided in Table 4.1, along with the corresponding performance of the original implementation [3] and the standard's specifications. The most important benefit of the S3D receiver is, undoubtedly, the 35.5% area reduction, despite the fact that [3] was implemented in a 90 nm process node. It should be noted that for a fair comparison, the VGAs and complex filters of the original implementation were not included in the area calculations. Given the large area consumed by the MIM capacitors, additional area gains could be possible, had the S3D PPDK been developed over a newer process node with a larger MIM capacitance density. The significant improvement of the phase noise over the 2-D implementation is not related to S3D integration. Instead, it should be attributed to the use of a MIM capacitor in place of a varactor in the LC-tank, since the latter tends to drop the tank's quality factor. As for the S3D receiver's gain it could easily rise up to the standard specifications with the insertion of a VGA after the TIAs.

The performance metric for which the original implementation outperforms the S3D design quite significantly is the power consumption: with the full chain implemented, the power of the 2-D receiver is exactly the same with the S3D one, which, however, includes only the LMV cell and the TIAs. The main reason for this is

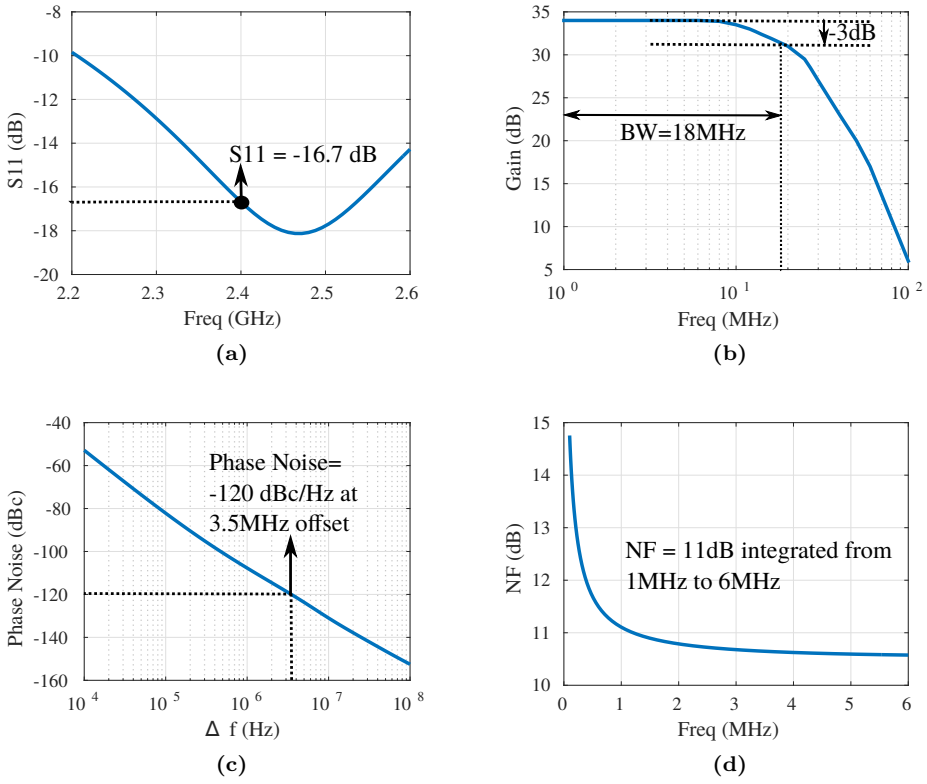


Figure 4.9: (a) S_{11} , (b) Gain of the receiver chain, (c) Phase Noise and (d) double-sideband NF versus frequency.

the lower f_T of the top-tier 150 nm transistors, compared to the 90 nm devices. As already explained, transistors with higher f_Ts require less current to achieve a specific performance. For the S3D receiver, the current drawn from the 1.8 V supply is 2 mA, of which 1.6 mA are allocated to the LMV cell and 400 μA to the two TIAs. Hence, the adoption of a more advanced node could lead not only to further area reduction (assuming increased MIM capacitance densities), but to power gains as well. All in all, the S3D receiver performance is on-par with [3] for most of the performance metrics. This, together with the 35.5% area gains proves the potential of S3D RF/AMS circuits.

4.5 Summary

In this chapter, the benefits that S3D integration can offer to applications other than digital circuits have been studied and its potential for RF/AMS circuits has been identified. Towards this, a frequency-based partition scheme has been devised with high frequency blocks implemented in the top-tier and low-frequency ones in the bottom. As a proof of concept, a receiver front end has been designed in the S3D PPDK and compared against the original 2-D implementation. The results indicate a 35.5% area reduction with no performance degradation.

Chapter 5

Inductors in a S3D process

In Chapter 4, the potential of S3D RF/AMS circuits has been demonstrated through the design of a S3D receiver front-end. The designed circuit achieved a 35.5% area reduction compared to the original 2-D implementation, with no performance degradation. It has been also observed that the total area of the S3D receiver is dictated by the area of the high-frequency blocks, due to the large dimensions of the required integrated inductor. The latter consumed 75% of the total area in the top tier. In general, inductors find wide use in RF circuits and systems, such as input matching networks, LNAs, VCOs and power amplifiers. Therefore, a study of integrated inductors in a S3D process is crucial for the complete analysis of S3D RF/AMS circuits, to highlight their benefits and potential shortcomings. This study will be the focus of this chapter.

5.1 Inductor Topologies for S3D Integration

The process stack of the S3D PPDK will be the basis of the coming analysis. This process stack has been depicted in Fig. 3.14 and its main features are summarized in Table 3.7. The total distance between MThick and the silicon substrate equals 13225 nm. On the other hand, the same distance in the base process equals 7900 nm. The thickness of MThick (the top-most thick metal layer) is 6.1 μm, while the thickness of the Si substrate, T_{Si}, and its resistivity, ρ_{Si}, are assumed equal to the base process's corresponding values, 200 μm and 15 Ωcm, respectively. The resistivity of the Si substrate plays a crucial role in the quality factor of integrated inductors. Higher resistivity values result in higher quality factors, thanks to the suppression of substrate losses [119, 120]. The benefits of high resistivity substrates extend also to the suppression of substrate noise. However, to achieve sufficiently high resistivity ($\rho_{Si} > 1$ kΩ cm), complex processing steps are needed. For example, in [121], the resistivity of the silicon substrate under the inductors was increased through its bombardment with helium ions. Naturally, such processing could increase the fabrication cost of S3D ICs, so they have not been considered in

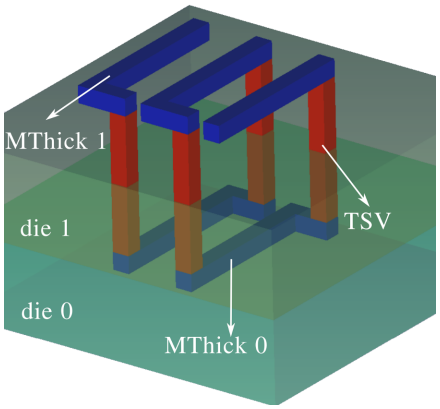


Figure 5.1: 3D solenoid implemented with TSVs

this study. Nevertheless, the value of ρ_{Si} that is considered in this work (15 Ωcm) is sufficient to suppress any current induced in the substrate by the inductor's magnetic field (eddy currents) [122]. These currents tend to reduce the inductance and increase the resistive losses of inductors, causing thus a significant drop in the inductors' quality factors. A grounded epoxy layer is considered underneath the S3D process stack, which serves as both the adhesive layer to the package and the current return path for the following EM simulations.

As for the optimal configuration for inductors in a S3D process, 3D solenoids, like the one in Fig. 5.1, have been already studied for TSV-based 3D ICs [123–125]. The main benefit of 3D solenoids lies with their high inductance densities, thanks to the relatively large thickness of the stacked dies (in the order of 50 μm). More specifically, a 2-4 times larger inductance density, compared to planar inductors, was reported in [125] for TSV-based 3D solenoids. Nevertheless, despite their high inductance densities, TSV-based 3D solenoids are susceptible to noise coupling between the TSVs and neighboring active devices [123]. Moreover, the Si substrate that surrounds the TSVs increases the solenoids' losses quite significantly. Placing cooling micro-channels around the TSVs [124] or surrounding them with grounded TSVs [125] have been shown to improve the quality factors of 3D solenoids with maximum values in the range 13-20.

However, 3D solenoids are not suited for S3D integration. The distance between MThick (top-most metal layer) and Met1 (bottom-most metal layer), in other words the maximum distance (D) between the top and bottom segments of a S3D solenoid, equals approximately 12 μm, which is significantly lower than 60 μm in [124] or 200 μm in [125]. This small distance results in substantial negative inductive coupling between the top and bottom segments of the solenoids, which, in turn, impacts adversely the obtained inductance and quality factors. Another drawback of S3D solenoids stems from the use of tungsten in the bottom tier in-

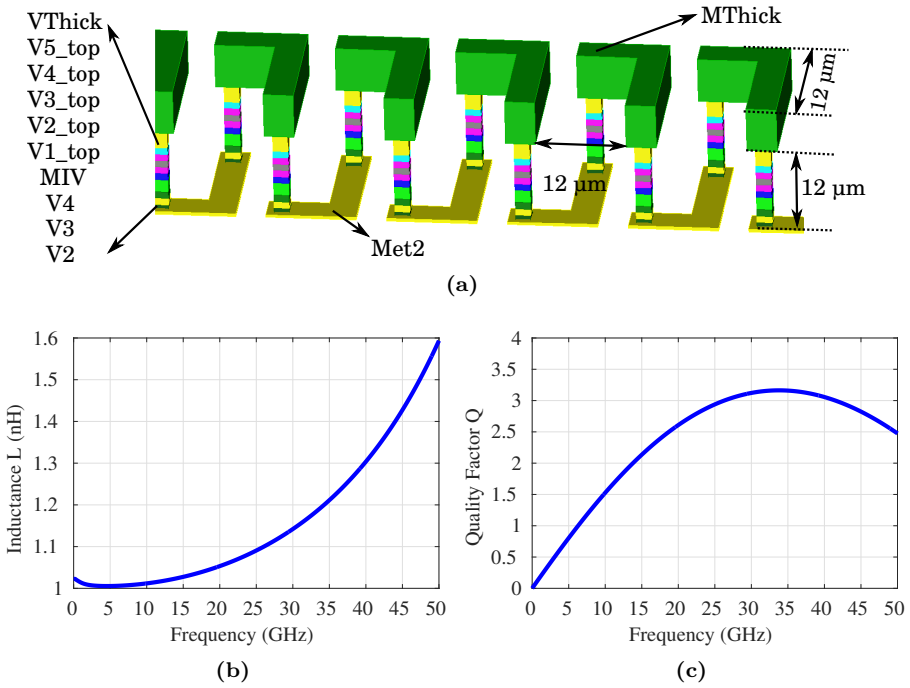


Figure 5.2: (a) Structure of the S3D solenoid. (b) Inductance and (c) Quality factor of the S3D solenoid.

terconnects. The higher resistivity of tungsten, along with the smaller thickness of the bottom tier metals (with regard to MThick) increase the resistance of the bottom segments, further reducing the S3D solenoids' quality factors. To verify these observations, a 3D solenoid has been designed in the S3D PPDK, and its snapshot is shown in Fig. 5.2(a). Its geometrical features were set in accordance to a TSV-based solenoid from [124] that has achieved 1 nH inductance and a maximum quality factor, $Q_{max} = 8.6$. To maximize the solenoid's quality factor, the length of the segments, l, needs to be equal to the vertical distance between the top and bottom parts of the solenoid, i.e. $l = D$. Since the value of D in S3D solenoids is much smaller than in TSV-based implementations, the number of turns has to be up-scaled to compensate for the drop in the overall length of the S3D topology. Eventually, 35 turns were used to yield an inductance value of 1 nH. All EM simulations were carried out with the Momentum simulator by Keysight® . A differential excitation between the solenoid ports has been assumed, so the following formulas were used to extract the differential inductance L_{diff} and quality factor Q_{diff}.

$$L_{diff} = \frac{1}{\omega} Im(Z_{diff}) \tag{5.1}$$

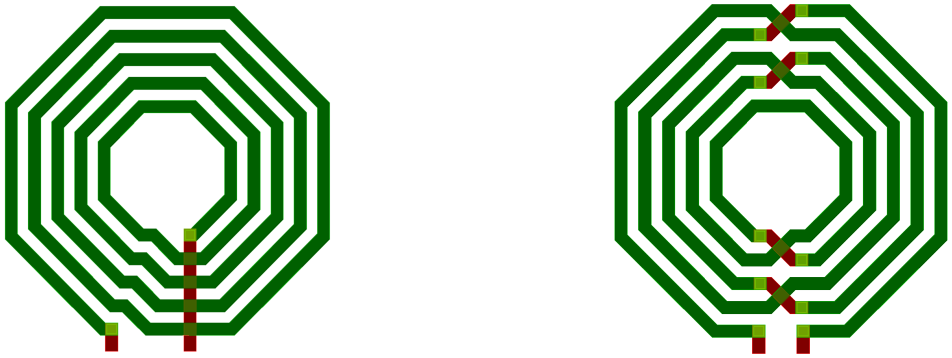


Figure 5.3: (a) Asymmetric and (b) Symmetric inductor topology.

$$Q_{diff} = \frac{Im(Z_{diff})}{Re(_{diff})} \tag{5.2}$$

$$Z_{diff} = Z(1,1) + Z(2,2) - Z(1,2) - Z(2,1) \tag{5.3}$$

The results are plotted in Fig. 5.2(b)-(c). As mentioned before, the topology achieves an inductance value of L $\approx$ 1 nH. However, the resulting quality factor is prohibitively low (< 5) rendering the topology impractical for RF/AMS circuits.

Therefore, planar inductors emerge as the most optimal configuration for S3D ICs. They need to be implemented in a thick metal layer in the top tier, i.e. MThick in the S3D PPDK. Bottom tier inductors would suffer from very low quality factors caused by the thin tungsten metal layers (0.4 μm) and the close distance to the substrate. This also inhibits the design of transformers in S3D technologies with one coil in the top tier and another in the bottom. A thick metal in the bottom tier to enable bottom tier inductors would be impractical, as it would result in excessive parasitics for the inter-tier vias.

Planar inductors can be implemented as either asymmetric (Fig. 5.3(a)) or symmetric topologies (Fig. 5.3(b)). As explained in [126], symmetric inductors can find use in both single-ended and differential topologies, unlike asymmetric ones that can be used only in single-ended applications. Hence, symmetric inductors will be the focus of this chapter. However, the results can be safely extended to asymmetric inductors as well.

5.2 Impact of S3D Integration on Planar Inductors

S3D integration can be based on any existing fab-process (base process) and can extend its capabilities by transferring an additional device layer on top of it [22,23].

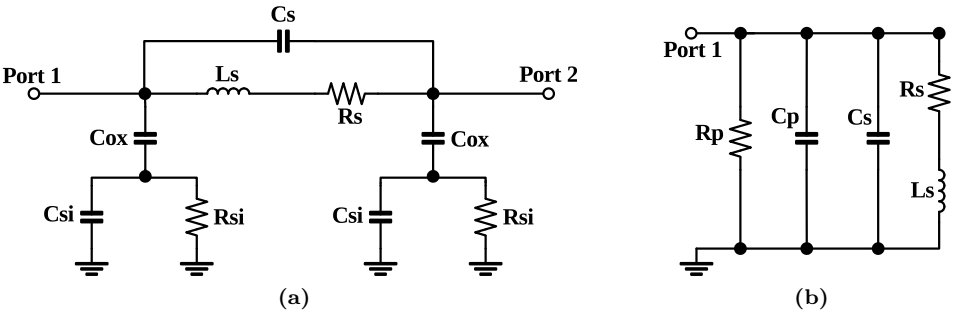


Figure 5.4: (a) Lumped network employed for inductors. (b) Simplification of the inductor lumped network, assuming Port 2 is grounded.

That was clearly the case with the S3D PPDK, which was built upon a 1P6M CMOS process. Therefore, it would be interesting to compare the inductance value and quality factor between a base-process inductor and a same topology inductor, implemented in the S3D PPDK.

This comparative study has been based on the lumped network shown in Fig. 5.4(a). The values of the network's components depend on design and process parameters, allowing thus an easy conceptualization of how these parameters impact the inductor performance. In particular:

- $\mathbf{L}_s$ models the topology's inductance at low frequencies. It depends mainly on the design parameters like the number of inductor turns, n, the topology's outer diameter, D_{out}, the width of the metal lines, w, and the spacing between the inductor coils, s.

- $\mathbf{R}_s$ models the series resistance of the topology. At low frequencies, it depends on the total length of the metal lines, their width and thickness. However, at higher frequencies, the advent of skin effects leads to a frequency dependence for R_s. At high frequencies, the current that flows in neighboring metal lines also affects R_s (current crowding and proximity effects).

- $\mathbf{C}_s$ models the capacitive coupling between the two ports of an inductor. It depends on the spacing of the metal lines, s, and the distance between Port 1 and Port 2. Similar to the case of R_s, C_s also exhibits a frequency dependence at higher frequencies.

- $\mathbf{R}_{Si}$ and $\mathbf{C}_{Si}$ account for substrate losses and depend on the inductor geometry, as well as the resistivity of the Si substrate, ρ_{Si}.

- $\mathbf{C}_{ox}$ accounts for the capacitance between the inductor and the substrate. It depends on the inductor geometry, its distance to the substrate, T_{ox}, and the dielectric constant of the oxide.

If one of the inductor ports is grounded, the lumped network of Fig. 5.4(a) is converted to the simplified network of Fig. 5.4(b), in which R_P and C_P are given by the following expressions [127]:

$$R_P = \frac{1}{\omega^2 C_{ox}^2 R_{Si}} + \frac{R_{Si}(C_{ox}+C_{Si})^2}{C_{ox}^2} \tag{5.4}$$

$$C_P = C_{ox}\ \frac{1+\omega^2(C_{ox}+C_{Si})C_{Si}R_{Si}^2}{1+\omega^2(C_{ox}+C_{Si})^2 R_{Si}^2} \tag{5.5}$$

The inductor's quality factor can be expressed as [127]:

$$Q = Q_{ideal} \cdot P_{subst} \cdot P_{SR} \tag{5.6}$$

with

$$Q_{ideal} = \frac{\omega L_s}{R_s} \tag{5.7}$$

$$P_{subst} = \frac{R_P}{R_P + R_s[1+(\omega L_s/R_s)^2]} \tag{5.8}$$

$$P_{SR} = 1 - \frac{R_s^2(C_s+C_P)}{L_s} - \omega^2 L_s(C_s+C_P) \tag{5.9}$$

Q_{ideal} represents the quality factor of an ideal inductor with its series resistance being the only source of losses. P_{subst} represents the inductor losses caused by the substrate and P_{SR} the losses due to the inductor's self-resonance.

Transferring a base process inductor to the derived S3D process, impacts mainly the oxide capacitance, C_{ox}, since the only process parameter that changes is the distance to the substrate T_{ox} (it increases, so C_{ox} drops). The rest of the process parameters remain practically constant. The impact of the increase in T_{ox} on the inductor's quality factor can be approximated by [126]:

$$\Delta Q = Q_{ideal} P_{subst}\left(\frac{R_s^2}{L_s} + \omega^2 L_s\right)\frac{C_{ox,2-D}}{T_{ox,2-D}}\Delta T_{ox} \tag{5.10}$$

For typical values of the parameters in (5.10), ΔQ is proved insignificant [126]. To verify this observation, an inductor with n= 3, w= 8 μm, s= 5 μm and D_{out}= 260 μm has been designed in the S3D PPDK and compared against its base process implementation. The results for the inductance and quality factor are plotted in Fig. 5.5(a-b), demonstrating indeed the negligible impact of S3D integration on the quality factor. More specifically, the S3D inductor exhibits only 1.2% increase in Q_{max}. Similarly, a small change of 2.1% is observed for the self resonance frequency, f_{SR}, i.e. the frequency at which the effective inductance of the topology drops to zero. As expected, the impact on the low-frequency inductance is negligible (3.3 nH for both the base process and the S3D PPDK inductors), as it does not depend on

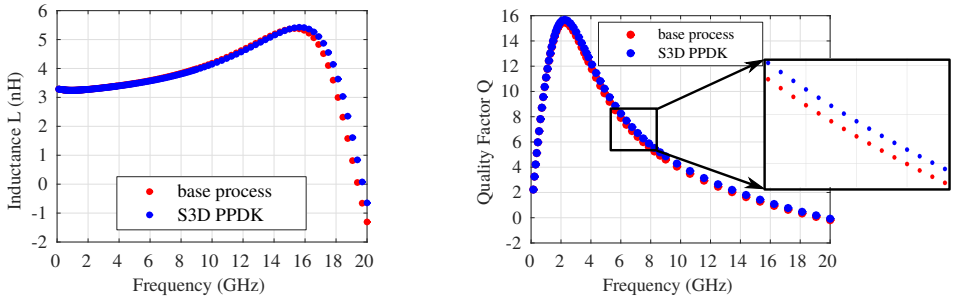


Figure 5.5: (a) Inductance and (b) quality factor for an inductor built in the base process and the same topology transferred to the S3D PPDK.

the inductor-substrate distance. This small improvement in the inductor's quality factor tends to become insignificant as the number of the inductor turns increases [126]. The reason is that the losses related to the series resistance R_s and the coupling capacitance C_s increase with the number of turns, suppressing any gains offered by the increase in T_{ox}. Hence, when building a S3D process over an existing 2-D one, the inductor models included in the base-process PDK can be re-used, as long as the thickness of the thick metal and the dielectric constants are kept the same.

5.3 Impact of shields

One popular approach to improve the quality factor of an integrated inductor lies on minimizing its substrate losses, in other words maximizing P_{subst} in (5.6). To do so, the inductor's electric field that reaches the substrate needs to be minimized. Towards this, various techniques have been proposed [127–129]. They all employ conductive structures (shields) underneath the inductor that terminate its electric field and prevent it from penetrating into the substrate. The conductive structures can be either grounded (patterned ground shields - PGS, [127]) or floating (like the horse-shoe shield in [128, 129]), with the latter finding use in differential applications only. The shields need to contain slots that inhibit the flow of loop currents induced by the inductor's magnetic field. This is easily achieved, by laying out the slots perpendicularly to the inductor's current flow. However, the presence of the shield between the inductor and the substrate increases the capacitance C_{ox}, and consequently reduces the term P_{SR} in (5.6) and the self-resonance frequency, f_{SR}.

To study the impact of shields on S3D inductors, the structure shown in Fig. 5.6 has been simulated with the Momentum simulator. The inductor topology has been selected to ensure that its losses are dominated by the substrate parasitics and

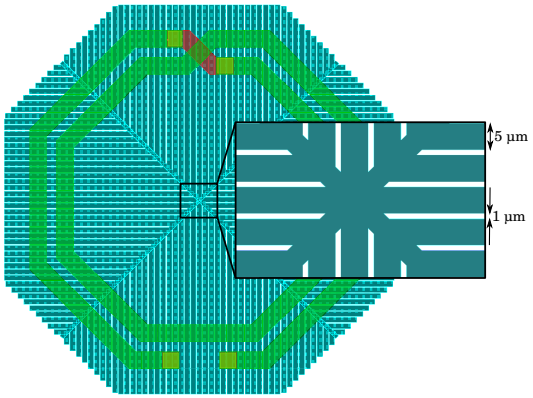


Figure 5.6: Inductor with a PGS. The shield contains metal stripes, each 5 μm wide. The spacing between adjacent stripes is 1 μm.

Table 5.1: Impact of the PGS on a S3D inductor

PGS layer	$f_{Q_{max}}$ [1] [GHz]	Q_{max}	L [2] [nH]	f_{SR} [GHz]
unshielded (M3D)	2.9	22.7	1.8	30.6
Slcd Active_Top[3]	2.9	22.3	1.8	27.3
Met5	2.8	21.7	1.8	27.2
Met4	2.9	21.6	1.8	26.2
Met3	2.9	21.8	1.8	28.1
Met2	2.9	22	1.8	28.5
Met1	2.9	22	1.8	28.8
Slcd Poly[4]	2.9	22.8	1.8	29.3

[1] frequency at which Q_{max} occurs
[2] inductance at $f_{Q_{max}}$
[3] silicided Active_top layer
[4] silicided ply-silicon

not by the series resistance or coupling capacitance between the coils. The reason is that the latter are not affected by the presence of a shield. The studied inductor contains n= 2 turns, with the spacing between each turn set to $s = 10$ μm. To further minimize the series resistance losses, the width of the turns was set to $w = 17$ μm. The inductor's output diameter is $D_{out} = 338$ μm, so that an inductance

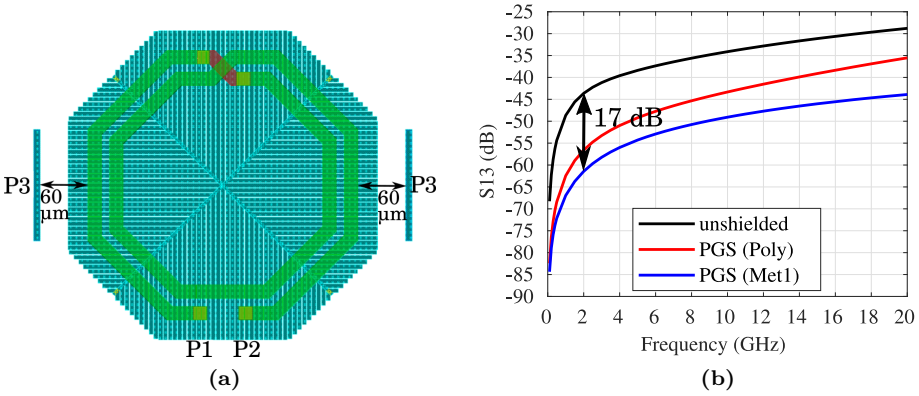


Figure 5.7: (a) Considered structure to validate the effectiveness of the PGS in suppressing substrate noise (b) Isolation between the inductor and the substrate, quantified by the S_{13} parameter.

value of 1.8 nH can be obtained. The considered PGS consists of metal fingers that are shorted along two diagonal lines. Each finger is 5 μ wide, while the spacing between adjacent fingers is 1 μm. For the shield implementation, each one of the bottom tier routing layers is considered (both metals and poly-silicon), in addition to the silicided active region of the top tier. A differential excitation is assumed for the inductor. The results for the inductance, L, the self-resonance frequency, f_{SR}, the maximum value of the quality factor, Q_{max}, and the frequency that it occurs, $f_{Q_{max}}$, are compared against the unshielded case in Table 5.1

It appears that the insertion of a PGS has only a minor impact on Q_{max}, which is opposite from what was expected: the inductor's Q_{max} drops with the insertion of a PGS. With the exception of the Active_top case, the further the shield is placed from the inductor, the smaller the drop in Q_{max} gets. Unlike the case of Q_{max}, the impact of the PGS on f_{SR} matches the theoretical analysis. As far as the inductance L and $f_{Q_{max}}$ are concerned, the presence of a PGS has no effect on them. The reason why the PGS fails to improve Q_{max} in the examined topology is related to the direction of the electric field in the substrate. Shields help to terminate only the electric field that is tangential to the substrate; they have no impact on the electric field that is vertical to the substrate surface [130]. Thus, in the studied topology the reduction of substrate-losses (increase in P_{subst}) fails to counter the increase in the self resonance losses, (reduction of P_{SR}).

Another important application of PGS is to suppress the coupling of substrate noise to the inductors. This is particularly useful for RF/AMS systems in which many blocks (i.e. digital, other inductors, etc) could inject noise to the substrate. Assuming the receiver front-end of Chapter 4, the substrate noise coupled to the inductor could increase its phase noise, corrupting thus the down-converted signal.

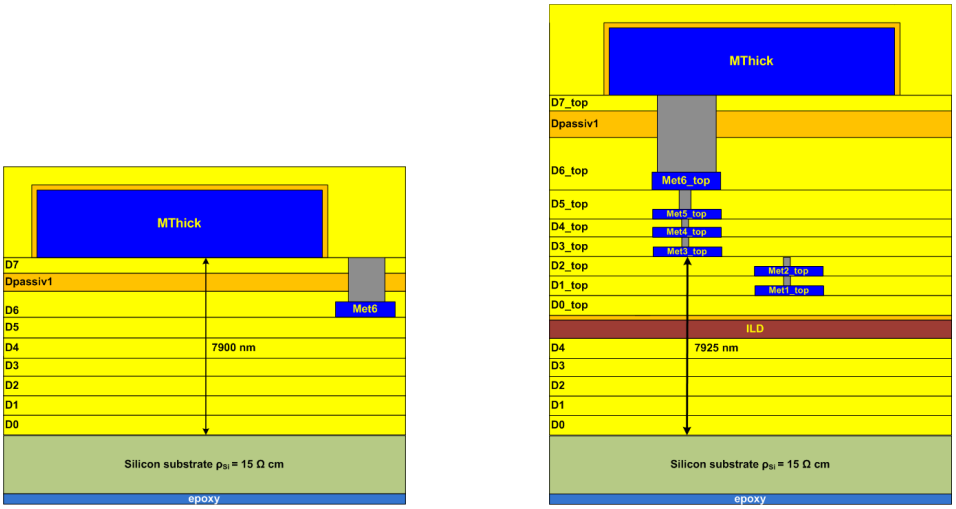
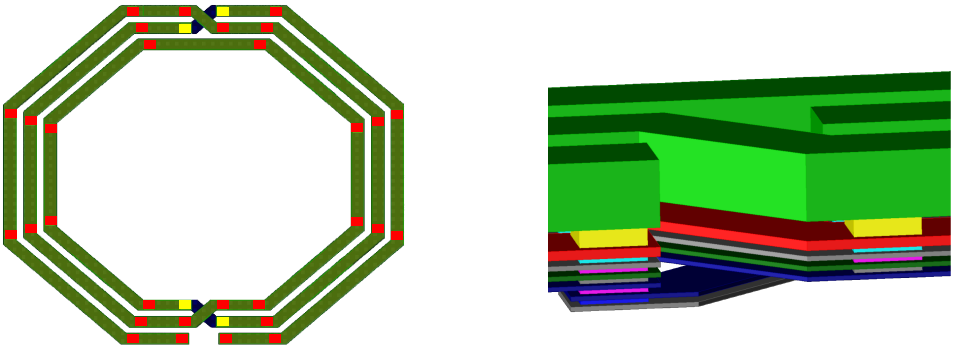


Figure 5.8: (a) Base process inductor (b) S3D inductor formed by shunting together the top tier metals MThick ->M3_top.

To evaluate the effectiveness of PGSs in suppressing substrate noise, the topology shown in Fig. 5.7(a) is examined. Noise is injected to the substrate from the M1 lines that are located left and right from the inductor, through an array of substrate contacts. Both lines are located 60 μm from the inductor edges and are both driven by port P3. The inductor-substrate isolation is quantified through the S_{13} parameter, which is plotted in Fig. 5.7(b). The results show that the presence of a Met1 PGS improves the inductor-substrate isolation by 17 dB at 2 GHz, compared to the unshielded case. Furthermore, it is evident that the metallic PGS provides better isolation from the substrate than the poly-silicon PGS, as it provides a less resistive path to ground.

5.4 Multiple-Metal S3D Inductors

Apart from reducing substrate losses, another approach to improve an inductor's quality factor is to minimize its series resistance, thus improving the term Q_{ideal} in (5.6). S3D integration offers an appealing way to achieve this, by shunting together multiple metal layers from the top tier, as conceptually shown in Fig. 5.8. In this way the low-frequency series resistance R_s is reduced significantly, while the distance to the substrate remains approximately the same with the base process case. Thus, an improvement in the inductor's quality factor should be expected. Furthermore, a reduction in f_{SR} is also expected due to the increase in the capac-



(a) (b)

Figure 5.9: (a) Multi-metal S3D inductor. The red rectangles are the vias that shunt together the multiple metal layers (b) 3D view of the multi-metal S3D inductor

itive coupling between the inductor-turns, as well as the increase in the fringing capacitance of the multi-metal implementation. To validate these observations, a multi-metal inductor topology with n= 3 turns, w= 8 μm, s= 5 μm and D_{out}= 250 μm is considered. It is formed by shunting together the top tier metals from MThick to M3_top, reserving M1_top and M2_top for the underpasses. The via stacks that shunt the metal layers are placed at the edges of the orthogonal segments of the inductor, as shown in Fig. 5.9(a). No via stacks are placed in the diagonal inductor segments, due to their relatively narrow width and the base-process design rule that prohibits VThick vias narrower than 6 μm.

The performance of this multi-metal inductor (S3D_multiMet inductor) is compared against its corresponding base process implementation on MThick with no metal shunting (2-D inductor). An additional inductor topology is also used as reference (eq. multiMet inductor). Its distance to the substrate is equal to the S3D_multiMet case, and it is implemented on a new metal layer, MThick_eq with a thickness of t_{eq}. The resistivity of MThick_eq is equal to MThick and its thickness is set so that it yields the same sheet resistance with the shunted metals. In other words:

$$\frac{t_{eq}}{\rho_{MThick}} = \frac{1}{r_{MThick}} + \frac{1}{r_{Met6\_top}} + ... + \frac{1}{r_{Met3\_top}} \tag{5.11}$$

where t_x is the thickness of the x-layer, r_x its sheet resistance listed in Table 3.7 and ρ_{MThick} the resistivity of MThick. A differential excitation is assumed and the results for the inductance and the quality factor are plotted in Fig. 5.10(a-b). Contrary to the theoretical expectations, shunting multiple metal layers results in a significant drop in Q_{max}, from 15.7 to 12.5. This unexpected drop is proved to be caused by a combination of strong proximity effects and negative magnetic coupling between the stacked metal layers [126]. Negative magnetic coupling between the

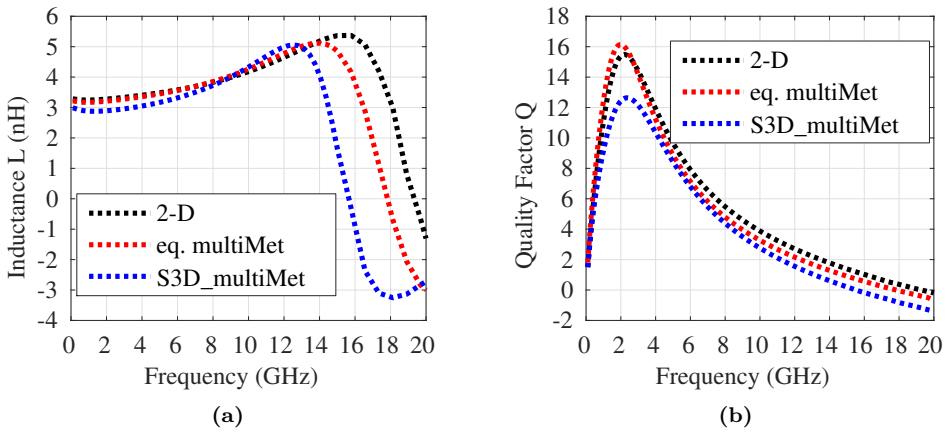


Figure 5.10: (a) Inductance and (b) quality factor of the multi-metal S3D inductor.

stacked lines, leads to a drop in the obtained inductance. On the other hand, proximity effects tend to increase the resistance of metal lines at high frequencies, as shown in [131]. In [131], the dependence of proximity effects on the distance between neighboring metal lines was demonstrated. Since the distance between the stacked lines is typically smaller than between adjacent lines, proximity effects are more pronounced in the former case. Indeed, for the S3D PPDK, the vertical distance between MThick and Met6_top is 1.7 μm. For lower metal layers, the vertical distance reduces further, reaching values as low as 0.5 μm between Met4_top and Met3_top. These values are significantly lower than the minimum distance between two MThick lines (6 μm as shown in Table 3.8).

To visualize the impact of the strong proximity effects and negative magnetic coupling between the stacked metal layers of the S3D_multiMet inductor, its resistance and inductance have been plotted in Fig. 5.11(a) and (b) respectively. The case of a multi-metal inductor formed by shunting only MThick and Met6_top is also included for comparison. Multi-metal inductors suffer indeed from higher resistance values at high frequencies, compared to single metal implementations. Furthermore, counter-intuitively, shunting more metals results in higher resistance values at high frequencies. However, the increase in the topology's resistance occurs mainly for frequencies higher than 5 GHz, which explains the reduced quality factor values for $f > f_{Q_{max}}$. The drop in the quality factor of the S3D_multiMet inductor at lower frequencies should be attributed to the inductance reduction, shown in Fig. 5.11(b). These findings suggest that S3D inductors formed by shunting multiple top-tier metals should be avoided.

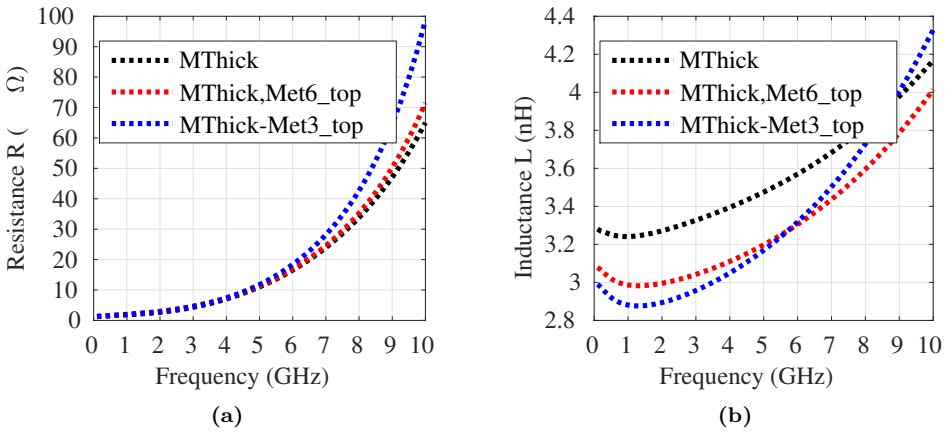


Figure 5.11: (a) Resistance and (b) low-frequency inductance of the multi-metal S3D inductor versus frequency.

5.5 Placement of bottom-tier blocks under top-tier inductors

The placement of bottom tier blocks under top-tier inductors could significantly improve the area-efficiency of S3D RF/AMS circuits and systems. To enable this placement, the main issue that needs to be addressed is minimizing the electromagnetic coupling between the inductors and the bottom tier blocks placed under them. Equivalently, the EM isolation between the inductor and the bottom tier blocks (inter-tier EM isolation) needs to be maximized. At the same time, care must be taken not to decrease the inductor's quality factor and inductance. The next sections consider both analog and digital blocks as potential candidates for placement under top-tier inductors and propose design guidelines for each case.

5.5.1 Analog Blocks

The position of bottom tier blocks relative to the inductor plays a pivotal role in the inter-tier EM isolation. To identify its impact, an inductor topology with n= 2 turns, w= 17 μm, s= 10 μm, D_{out}= 338 μm and a bottom-tier resistive network that is shown in the inset of Fig. 5.12(a) serve as a test-case. The resistive network consists of non-silicided poly-silicon resistors, each with a length of 8.5 μm and a width of 3.5 μm. The choice of a resistive network as a test vehicle has been motivated by the difficulty in combining EM simulations with transistor-level simulations. The main obstacle lies with the bias dependence of the transistors' characteristics, like the transconductance g_m and the output resistance r_{out}. Nevertheless, any findings from the conducted analysis can be safely extended to blocks

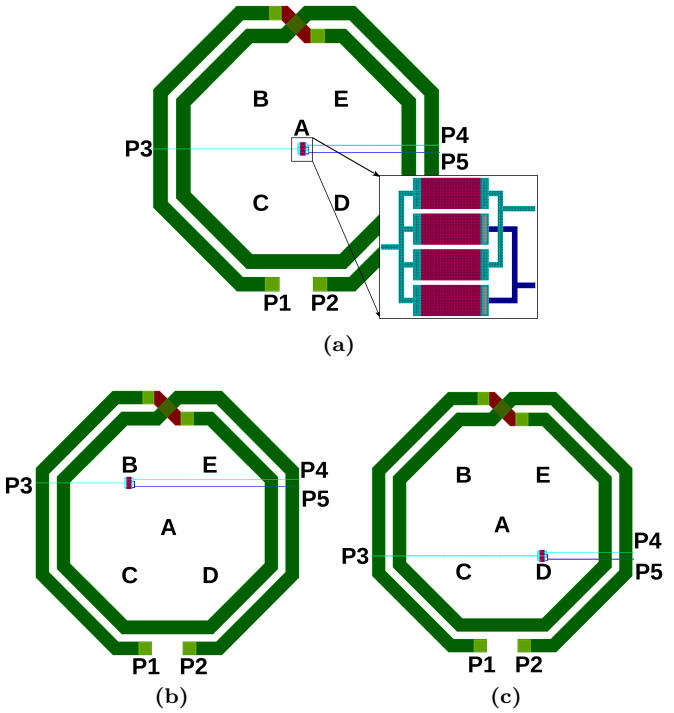


Figure 5.12: Considered structure to study the impact of the location of bottom tier blocks under the inductor on the inter-tier EM isolation.

containing transistors, assuming small electromagnetic coupling between the inductor and the channel area of the transistors. The latter holds true, especially for newer nodes. The position of the resistive network is swept between locations A, B and D shown in Fig. 5.12(a)-(c). Location A corresponds to the inductor center, whereas locations B and D are both symmetric to the inductor center. The inductor is driven by ports P1-P2, whereas the resistive network by P3-P5. The inter-tier EM isolation is quantified by the S_{13} parameter, which is defined as:

$$S_{13} = \frac{V_1^-}{V_3^+}|_{V_1^+=0} \tag{5.12}$$

Its attractiveness lies in the fact that it can be also used to quantify the cross-talk and voltage degradation between ports P1 and P3. Indeed, if port P3 is properly terminated (no reflections at P3) and given that port P1 is not excited ($V_1^+ = 0$), then S_{13} yields the voltage ratio V_1/V_3. Assuming S_{13} = -40 dB, 1 V swing at port P3 would result in 10 mV swing at the inductor port P1. The results for the inter-tier EM isolation is plotted in Fig. 5.13 for two scenarios: (a) with and (b) without a PGS between the inductor and the resistive network. To free-up bottom

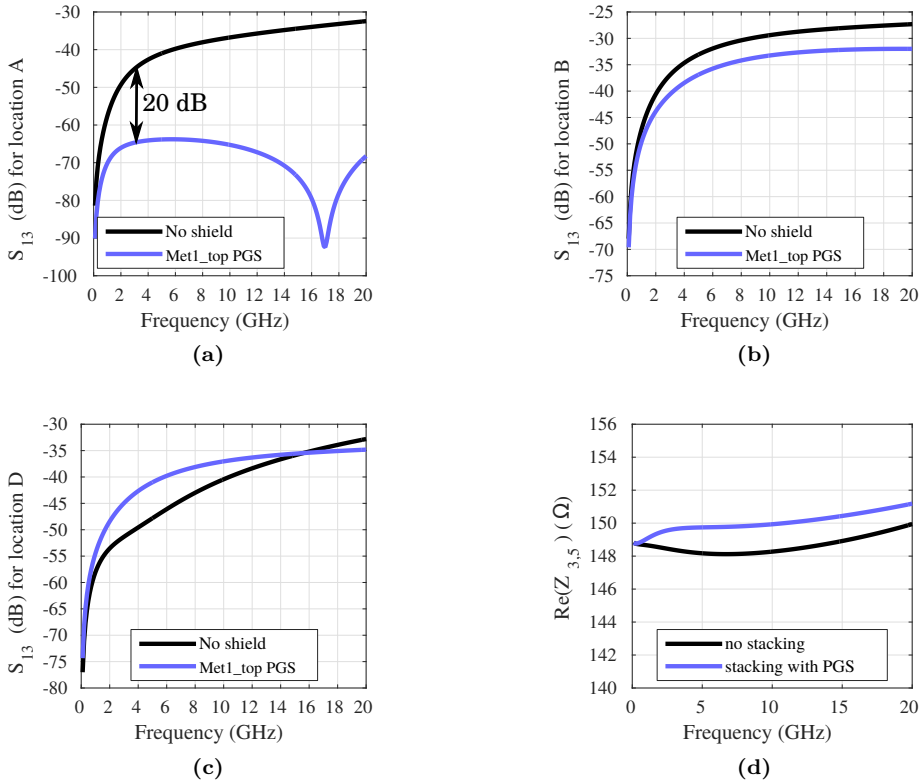


Figure 5.13: Inter-tier EM isolation when the resistive network is placed under the inductor at position (a) A (b) B and (c) D. (d) Resistance between ports P3 and P5 before and after placing the resistive network under the inductor center (location A).

tier interconnects, no bottom tier implementation for the PGS has been considered. However, the closer the shield is placed to the inductor, the larger the drop in both its Q_{max} and f_{SR} would get. Thus, a Met1_top implementation for the PGS was selected.

Fig. 5.13(a)-(c) indicate that at low frequencies, well below 1 GHz, the inter-tier EM isolation is sufficiently high, irrespective of the location of the resistive network or the presence of a PGS. This indicates that at low frequencies, the inductor's electric field does not penetrate deep into the studied structure and hence it does not reach the resistive network. For higher frequencies, the best location for the resistive network in terms of inter-tier EM isolation is at the inductor center (location A). For this case, the PGS improves the inter-tier EM isolation by 20 dB at 3 GHz. A detailed analysis on the behavior of S_{13} with frequency is provided in [126]. The

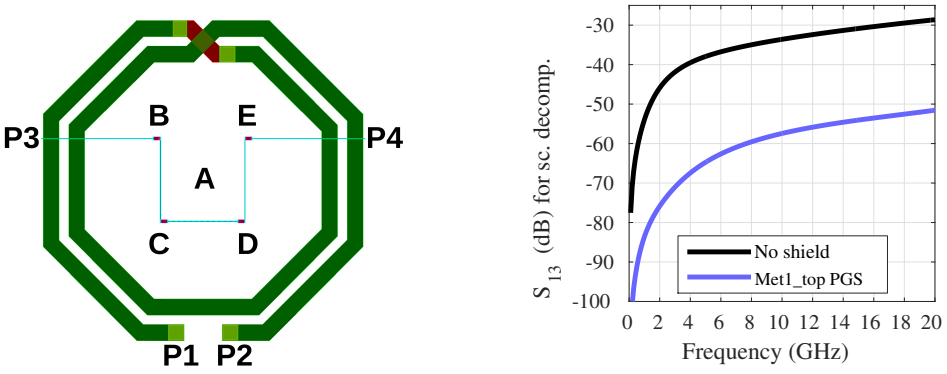


Figure 5.14: (a) Studied topology to emulate larger blocks and (b) inter-tier EM isolation.

placement of the resistive network under the inductor has practically no impact on the network's resistive behavior, causing a change of less than 2 Ω in the resistance between ports P3 and P5, at 3 GHz (see Fig. 5.13(d)).

To emulate blocks that are placed under the inductor and occupy a larger area, the considered resistive network is de-composed into its four resistors. The devices are then connected in series and laid out around the inductor center, as shown in Fig. 5.14(a). Interestingly, despite the low inter-tier EM isolation at positions B and D, the resulting S_{13} is sufficiently low, as shown in Fig. 5.14(b). This observation highlights the importance of laying-out bottom tier blocks symmetrically with respect to the inductor center.

Based on these findings a set of design guidelines have been drawn to enable an efficient placement of bottom tier blocks under top-tier inductors. In short:

- Blocks operating at low frequencies, well below 1 GHz, can be placed at any location underneath top-tier inductors. For such blocks, the inter-tier EM isolation is sufficiently high (S_{13} < -65 dB) even without the use of a PGS. Still, despite the high inter-tier EM isolation, blocks with high sensitivity requirements, like high-resolution data converters should be avoided.

- Blocks operating at higher frequencies should be placed at the inductor center. The use of a PGS is also imperative. Special care must be taken to ensure that the intra-block routing is symmetrical with respect to the inductor center. Furthermore, the placement of high frequency blocks that process signals with low power-levels, such as LNAs, under top-tier inductors with large voltage swings (i.e. LOs) should be avoided.

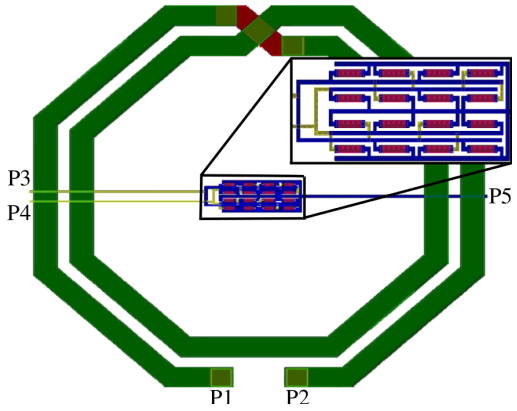


Figure 5.15: 4x4 resistor array under a top-tier inductor

Table 5.2: Inter-tier EM isolation

		Inter-tier EM isolation
16 res. array	S_{13} at 3 GHz	-59.7 dB
	S_{14} at 3 GHz	-74.8 dB
	S_{25} at 3 GHz	-60.1 dB
32 res. array	S_{13} at 3 GHz	-60.5 dB
	S_{14} at 3 GHz	-71.4 dB
	S_{25} at 3 GHz	-59.7 dB

The proposed guidelines, refer only to S3D ICs. For conventional 2-D processes, the relatively small distance between the inductors and the blocks placed under them would exacerbate the EM coupling between the two. Furthermore, in 2-D ICs, the layers available for the PGS would be rather close to the inductor, significantly lowering their quality factors.

As a proof of concept, an array of 16 bottom tier resistors, shown in the inset of Fig. 5.15, has been placed under a top-tier inductor that has the same topology as in Fig. 5.12. Each of the resistors was 10 μm long and 2 μ wide. The proposed guidelines were followed and a Met1_top PGS was included between the inductor and the resistor array. To evaluate the inter-tier EM isolation for larger bottom tier blocks, an array with double the resistor count (i.e. 32) has been also considered. The results for both cases are presented in Table 5.2. The inter-tier EM isolation is quantified through the parameters S_{13}, S_{14} and S_{25}, as the ports P3-P4 are on the same side with P1, whereas P5 is on the same side with P2. The results indicate that all three parameters remain sufficiently low ($\leqslant$-60 dB for frequencies up to

Table 5.3: Characteristics of the passive devices before and after stacking

		Before Placement		After Placement
		w/o PGS	with PGS	
16 resistors array	$f_{Q_{max}}$	2.9 GHz	2.6 GHz	2.6GHz
	Q_{max}	22.6	20.3	20.2
	$L(f_{Q_{max}})$	1.8 nH	1.8 nH	1.8 nH
	f_{SR}	30.6 GHz	26.3 GHz	26.3 GHz
	$Re(Z_{45})$	84 Ω	-	84.4 Ω
32 resistors array	$f_{Q_{max}}$	2.9 GHz	2.6 GHz	2.6GHz
	Q_{max}	22.6	20.3	20.1
	$L(f_{Q_{max}})$	1.8 nH	1.8 nH	1.8 nH
	f_{SR}	30.6 GHz	26.3 GHz	26.3 GHz
	$Re(Z_{45})$	65.2 Ω	-	65.8 Ω

3 GHz). It appears that the inter-tier EM isolation is independent of the size of the bottom-tier block. The impact of this placement on the characteristics of both the inductor and the resistors array is shown in Table 5.3. It is proved that the inductor's Q_{max}, f_{SR} and inductance at $f_{Q_{max}}$ do not depend on the presence of bottom-tier analog blocks under it. In fact, the inductor's behavior depends only on the presence or absence of a PGS. As for the resistor-array, its placement under the top-tier inductor has an insignificant impact on its resistive behavior, as illustrated by the less than 1% change in the real part of the impedance Z_{45}, i.e. the impedance between ports P4 and P5.

5.5.2 Digital Blocks

Digital blocks can find wide use in RF/AMS circuits and systems. Examples of their applications include controlling the gain of VGAs [132], and calibration of mixers to compensate for process, temperature and supply variations [133]. The design guidelines proposed in section 5.5.1 need to be extended to be able to cover the placement of digital blocks under top-tier inductors as well. The main features of digital circuits that need to be considered in order to guarantee their efficient placement under top-tier inductors are:

- The rail-to-rail voltage swings of digital signals call for sufficiently high inter-tier EM isolation (in excess of 60 dB), to prevent their coupling to the inductor.
- The closed-loop that is formed by the block's power ring fails to suppress magnetically induced eddy currents, with detrimental effects on the inter-tier EM isolation, the inductor's quality factor and its inductance.

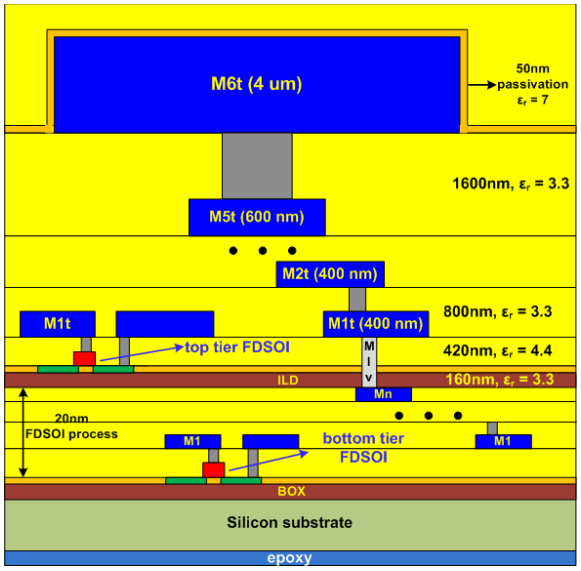


Figure 5.16: S3D process stack used to study the EM coupling between bottom-tier digital blocks and top-tier inductors. The process data of the 20 nm FDSOI process are provided on an "as is" basis through ASCENT, funded by the EU H2020 Infrastructure Programme (H2020-INFRAIA-2014-2015) under Grant Agreement 654384

- Care must be taken to ensure a minimum penalty on the loading of digital wires, and consequently prevent re-running the synthesis, as well as place and route flows.

Process stack. For the analysis of the EM coupling between digital blocks and a top-tier inductor, a new process stack needs to be considered, due to the absence of a digital cell library in the S3D PPDK. This process stack is depicted in Fig. 5.16. A commercial 20 nm FDSOI process is considered for the bottom tier, after changing the material of its metal lines to tungsten. However, this process stack is optimized for digital applications and high integration densities. It features thin metal layers and a small distance between them, with both features being unsuitable for RF/AMS applications. Consequently, a custom BEOL has been considered for the top-tier. It consists of five Al metal layers (M1t-M5t) and a thick one, M6t, reserved only for inductors. It is clear that the characteristics of top-tier's BEOL resemble the corresponding values of the S3D PPDK. The relatively large distance between M6t and M5t aims to increase the distance between the inductor and the Si substrate and so, to reduce the losses to the substrate.

PGS implementation. For the implementation of PGSs, a trade-off exists

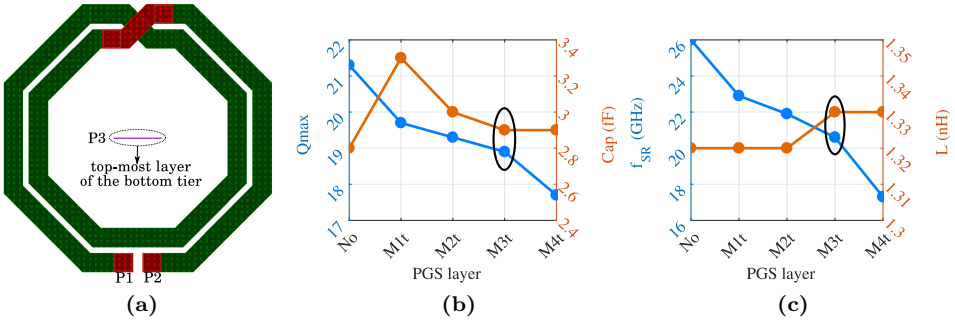


Figure 5.17: (a) Studied topology to identify the most optimal PGS layer (b) Impact of the PGS on inductor's Q_{max} and wire's *Cap*. (c) Impact of PGS on f_{SR} and L_{diff}.

between the resulting loading of the digital signals and the drop in the inductor's Q_{max}. In particular, the closer the shield is placed to the inductor, the smaller the capacitive loading of digital wires gets. However, the penalty on the inductor's Q_{max} becomes more severe. Thus, to determine the most optimal implementation for the PGS, the setup of Fig. 5.17(a) is analyzed. The inductor design parameters are: n= 2, w= 20 μm, s= 6 μm, D_{out}= 280 μm. A 50 μm long wire, on the top-most metal layer of the bottom tier is placed at the center of the inductor and driven by port P3. A PGS with the same topology as in Fig. 5.6 is inserted between the inductor and the wire. Four different scenarios are considered for the PGS layer: (a) M1t PGS, (b) M2t PGS, (b) M3t PGS and (b) M4t PGS. An unshielded case is also considered as a reference. For each scenario, the inductor's behavior (Q_{max}, L, f_{SR}), as well as the capacitance of the wire, *Cap*, are simulated. The latter is calculated by:

$$Cap = \frac{1}{\omega} Im[Y_{33}] \tag{5.13}$$

The results in Fig. 5.17(b)-(c) demonstrate that the M3t PGS is the most optimal option for the shield. The capacitance of the wire, *Cap*, increases by only 3.6%, from 2.8 fF to 2.9 fF, compared to the unshielded case. At the same time the inductor's Q_{max} drops by 10%, and its self resonance frequency, f_{SR}, by 21%. A M4t implementation would offer no further reduction in *Cap* but would cause a further reduction in Q_{max} by 6% and in f_{SR} by 17%, compared to the M3t case. Naturally, the inductance is practically independent of the PGS layer and equal to 1.3 nH.

Power Ring. To study the impact of eddy currents on the power ring (PR) of the digital cells, the topology of Fig. 5.18(a) is considered, with the same inductor topology as in Fig. 5.17. A M3t PGS is also used, which, however, it is not shown in Fig. 5.18(a). As already discussed, eddy currents result in reduced inductance

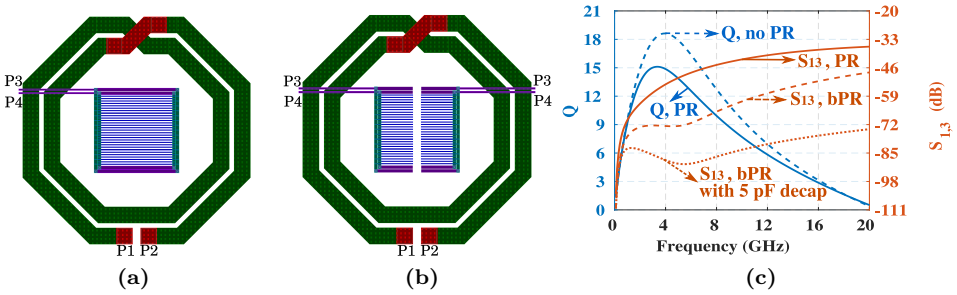


Figure 5.18: (a) Top tier inductor and bottom tier power ring.(b) Proposed broken power ring to suppress eddy currents. (c) Inter-tier EM isolation between the inductor and the power ring, and quality factor for the inductor.

and increased series resistance for the inductor, causing a significant drop in the inductor's Q. This is clearly illustrated in Fig. 5.18(c), where the inductor's Q_{max} drops by 18%. To tackle this issue, the power ring can be broken in two parts, as shown in Fig. 5.18(b) [134]. The two halves can be reconnected outside the inductor area, where the influence of its magnetic field is insignificant. The breaking of the power ring cancels out the drop in the inductor's quality factor. It also improves the EM isolation between the inductor and the PR by 15 dB at 3 GHz, as it is shown in Fig. 5.18(c). The EM isolation between the inductor and the PR can be further improved through the connection of a decoupling capacitor between the power rails. For the studied topology, a 5 pF decoupling capacitor provides very high EM isolation (86 dB at 3 GHz).

Proof of Concept. To demonstrate the effectiveness of the proposed design guidelines, a 3-bit adder has been synthesized in the 20 nm FDSOI process. A clock frequency of 1 GHz has been targeted during its synthesis and place-and-route. Despite the small complexity of the design, all of the bottom tier's metals are used for the adder's implementation. The synthesized adder is placed under the top-tier inductor of Fig. 5.17(a) at its center. Its PR is then split into two halves, which are reconnected outside the inductor, and a 5 pF decoupling capacitor is connected between the power rails. A M3t PGS is inserted between the inductor and the adder. The studied topology is shown in Fig. 5.19, which, however, does not include the PGS. For the EM simulations, the active regions, both of the source/drain and channel areas are considered as conductors with a sheet resistance of 12 Ω/□. This assumption tends to overestimate the EM coupling between the inductor and the digital block, a trend that is considered more preferable that underestimating it. To account for the worst case scenario, in which all of the digital wires switch simultaneously and towards the same direction, all of the block's I/O pins are connected to port P4. The worst case inter-tier EM isolation is then quantified by

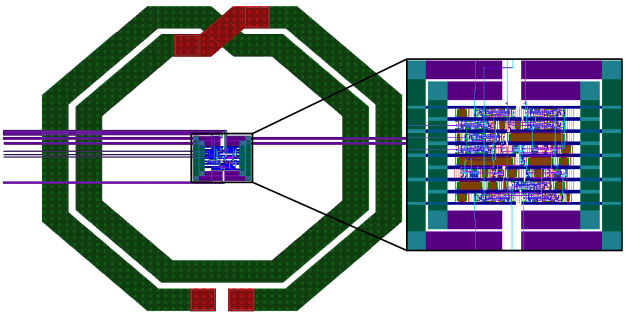


Figure 5.19: A 3-bit adder is placed in the bottom tier under a top-tier inductor. Details of the adder implementation are shown in the inset.

Table 5.4: Analysis of the placement of a 3-bit adder under a top-tier inductor

	Before Placement		**After Placement**
	w/o PGS	with PGS	
L	1.3 nH	1.3 nH	1.3 nH
Q_{max}	18.9	21	18
C_{clock}	20 fF	20 fF	20.3 fF
S_{wrst}	-	-	-75 dB
$S_{V_{DD}}$	-	-	-86 dB

S_{14} (= S_{wrst}). Port P3 drives the supply line of the power rails (V_{DD}). Hence, the S_{13} parameter is used to quantify the isolation between the inductor and the power ring.

The results are shown in Table 5.4. The inductance is not affected at all by the placement of the 3-bit adder under it (L = 1.3 nH). As for the quality factor, the 14% drop in the value of Q_{max} is mainly caused by the insertion of the PGS. The additional decrease in Q_{max} from 18.9 to 18 could be caused by the presence of closed-loops in the bottom tier, formed by the intra-block routing. Nonetheless, the obtained value of Q_{max} after the placement of the digital block is sufficiently high for most applications. The penalty on the capacitive loading of the clock-signal, ΔC_{clock}, after the placement of the adder under the inductor is small, 0.3 fF or 1.5% of its value before the placement. Here, it should be noted that the penalty on both Q_{max} and C_{clock} can be further minimized by calibrating the top-tier's BEOL, for instance by increasing the distance between M6t and M5t, or by employing insulators with a lower dielectric constant between the metal layers. As for the inter-tier EM isolation, breaking the PR and utilizing a decoupling capacitor results in ultra high EM isolation between the inductor and the PR. For the worst

case scenario, when all the signals are switching simultaneously and to the same direction, the inter-tier EM isolation reaches 75 dB. For instance, 1 V swing at the digital signals would cause 180 μV swing at he inductor's port P1. Such a scenario however is extreme.

5.6 Summary

To further highlight the potential of S3D RF/AMS circuits, an analysis of S3D inductors has been undertaken. Planar coils in the top-most thick metal have been found as the most optimal configuration for inductors in a S3D technology. The considered shield structures were proved successful in suppressing substrate noise. However, this may be accompanied by a decrease in the inductors' quality factors. Shunting multiple metal layers in the top-tier of a S3D process should be avoided, as it can give rise to proximity effects and reduced inductances and quality factors. Finally, a set of design guidelines have been drawn to enable the placement of bottom tier blocks, both analog and digital, under top-tier inductors. Such an approach could further improve the area efficiency of S3D RF/AMS circuits.

Chapter 6

Heterogeneous S3D Integration: Ge over Si

This chapter describes the prospects of heterogeneous S3D integration and focuses on its impact on circuit performance. The study is based on the in-house S3D process with Ge in the top-tier and Si in the bottom. However, since the latter is limited to only long-channel transistors, the performance of short-channel Ge FETs is predicted from measurement results of in-house devices and Technology Computer Aided Design (TCAD) tools.

6.1 Current Status of Heterogeneous S3D Integration

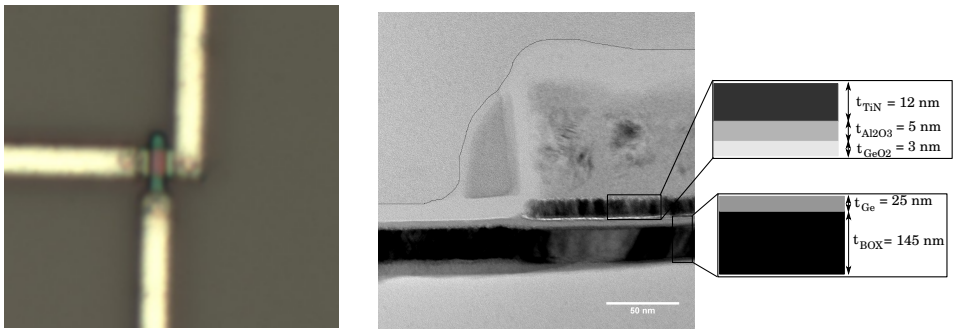
To overcome the barrier of low-temperature processing in the top-tier, the use of active layers other than Si, for instance III-V materials and Ge, has been investigated. Apart from the low processing temperatures that they enable, III-V materials and Ge exhibit also high bulk mobilities and band-gap values similar to Si. III-V materials are favorable for the implementation of nFETs, so they enable a S3D integration scheme with n-type transistors in the top-tier. Similarly, the use of Ge in the top-tier favors a S3D integration scheme with pFETs in the top. In contrast, a homogeneous Si-over-Si S3D scheme enables both n- and p-type transistors in any tier.

A comparison between the performance of III-V and Ge FETs has already been made in Chapter 2, based on data extracted from state-of-the-art works in the field [43, 81] . Both cases exhibit high SS, 150 mV/dec and 170 mV/dec for the III-V and Ge FETs respectively. It should be noted, however, that the results for the III-V transistors referred to short-channel devices, unlike the long-channel Ge ones (800 nm long). Similar results have been also observed in other works for devices with different lengths. 70 nm long III-V FETs exhibiting SS= 96 mV/dec and 120 nm long devices with SS= 100 mV/dec have been reported in [83] and [82] respectively. Furthermore, 170 nm long Ge transistors have been reported in

[135], yielding a SS value of 116 mV/dec. Another observation made in Chapter 2 referred to the compatibility of the Ge-based heterogeneous S3D approach with existing widely used wafer technologies, thanks to the use of a SRB [43]. Indeed, a Ge active layer can be transferred with any widely available Si donor wafers [43], as opposed to [81–83], which require the use of smaller-diameter InP wafers. This has been among the main motivations for the adoption of Ge active layers in the in-house S3D process.

Despite the extensive work on the processing aspects of heterogeneous S3D integration, there has been limited research in its potential applications in circuits and systems, with the existing works focusing mainly on III-V materials. The implementation of photo-detecting applications in the top III-V layer has been proposed in [42]. The detection wavelength can range from infra-red to ultra-violet, depending on the selected III-V compounds. Top-tier photo-diodes were also demonstrated for detection in the visible spectrum [42]. Readout circuitry, which was not included in [42], could be implemented in the bottom tier with Si devices. III-V based S3D heterogeneous integration has been also proposed for MicroLED (Micro Light Emitting Diodes) displays by integrating active layers with various III-V compounds one on-top the other. Each layer can then contain diodes for light emission at a specific wavelength [42]. As for digital applications, S3D inverters with III-V transistors in the top-tier and SiGe ones in the bottom have been reported in [83]. They exhibited inverter characteristics for supply voltages as low as 250 mV, however, the results were limited to DC characterization only, with no time-domain analysis. The prospect of S3D RF applications with III-V FETs in the top-tier has been identified in [136], nonetheless, the analysis was based solely on demonstrating the high f_T that can be obtained with III-V FETs. However, III-V-on-insulator transistors that have been fabricated with a low-temperature process, suitable for S3D integration, exhibited rather low f_T values instead [82]. In particular, for devices with L= 120 nm, f_T was limited to 16.4 GHz. For comparison, the 150 nm long Si bulk transistors that served as the base process in the S3D PPDK (see Chapter 3 and 4), demonstrated a f_T= 74 GHz. The low f_T has been attributed to the relatively high source/drain resistance, which in turn is caused by the lack of any silicidation-like reactions.

Unlike the III-V-over-Si S3D integration, the potential of the Ge-over-Si approach for circuits and systems has not been investigated yet. Hence, the next sections focus on such an investigation and study. This study is not restricted to the capabilities of the in-house process. Instead, drawing from the measurement results of the in-house long-channel Ge FETs, as well as, state-of-the-art works in the field, the performance of short-channel Ge-FETs will be predicted with the help of TCAD simulations. The procedure that was followed can be summarized in the following steps.



(a) **(b)**

Figure 6.1: (a)Top-view and (b) cross-section of a measured in-house Ge-pFET with L= 800 nm

1. TCAD models have been calibrated to match the measurement results for the in-house Ge devices
2. The main issues that affect the performance of long-channel devices have been studied and solutions to solve them have been identified.
3. Taking into consideration these solutions, TCAD simulations have been re-run for short-channel FETs.
4. The TCAD simulation results have served as a basis for identifying the circuits that could benefit from Ge FETs over Si.
5. To facilitate circuit analysis, spice models have been calibrated to match the above TCAD simulations.

6.2 In-house Ge transistors

The topology of the measured 800 nm long Ge transistors is depicted in Fig. 6.1. The gate stack consists of 3 nm of thermally grown GeO_2 and 5 nm of alumina (Al_2O_3) [43], resulting in EOT $\approx$ 4.3 nm. The device's V_{TH} is set by the deposition of 12 nm of TiN over the alumina. No silicidation-like reaction of the source/drain/gate terminals is carried out, in order to comply with the low-temperature processing specifications. It is important to note that the BOX thickness, 145 nm, is different from the Chapter 3 value T_{BOX}= T_{ILD}= 1μm. This is because the measured transistors were formed by transferring and bonding a crystalline Ge layer over the BOX of a SOI wafer, not over the ILD of the in-house S3D process. The processing temperature of the measured transistors was limited to 560^{o}C and since the BOX thickness is large enough to negate the impact of back-gate biasing, the performance of the measured devices is expected to match

top-tier Ge FETs. The measurement results are plotted in Fig. 6.2 and compared against corresponding TCAD simulations. The drift-diffusion transport model was employed for the TCAD simulations.

Looking at Fig. 6.2(a), the on-current of the devices reaches I_{on} = 54 μA, whereas the off-current is I_{off} = 90 nA, yielding a ratio of I_{on}/I_{off} = 600. At high V_{GS}, the on-current is limited mainly by the source/drain resistance, R_s. The latter should be attributed to the lack of silicidation-like reactions and the low thermal budget in the top tier that results in incomplete dopant activation. Indeed, the concentration of activated dopants in the source/drain areas is limited to 3.8e18 cm^{-3}, as shown in Fig. 6.3. A value for $R_s \approx 6.5$ k$\Omega/\square$ has been considered for a good agreement between the measurement data and TCAD simulations . High R_s was also observed in [82, 137]. The relatively high I_{off} indicates a poor interface between the channel and the gate-stack with a high density of traps. Traps offer an additional conduction path, even when the transistor is off, which explains also the relatively large sub-threshold slope, SS = 170 mV/dec. To reproduce this value, acceptor-type traps were considered in the Ge/GeO_2 interface at midband with a state density of D_{it} = 7.5e12 $eV^{-1}cm^{-2}$. A uniform distribution for these traps was also assumed in the energy band (E_0 - 0.35 eV) < E < (E_0 + 0.35 eV), with E_0 the midgap level of Ge. Lastly, the constant mobility of Ge, i.e. the mobility prior to any degradation mechanism, was set to $\mu_{o,Ge}$ = 750 $cm^2V^{-1}s^{-1}$. This value is 58% larger than the corresponding value for Si ($\mu_{o,Si}$ = 475 $cm^2V^{-1}s^{-1}$), which is in agreement with [43], where 60% higher mobility was reported for Ge FETs, as compared to reference Si devices.

6.3 Short-channel Ge transistors

The results described so far indicate that the high source/drain resistance and the poor quality of the gate-channel interface present major roadblocks for the successful scaling of the in-house Ge transistors. The impact of R_s on the I_{on}, the transconductance, as well as the f_T of the transistors becomes more pronounced for scaled devices. The main solution that has been proposed to reduce R_s is the use of raised source/drain (RSD) structures [81, 82, 135, 138–140]. RSD structures are grown epitaxially with in-situ doping, which, however, increases the processing temperature, without a sufficient reduction in R_s, as illustrated in [82]. Recently, the formation of nickel germanides (NiGe) has been shown to reduce the R_s of Ge FETs [141]. In particular, sheet resistance values below 50 $\Omega/\square$ were achieved with the deposition of 8.4 nm of Ni and a rapid thermal anneal at 400^{o}C, a low-enough temperature to allow the adoption of this solution for top-tier Ge FETs. Hence, the coming analysis will consider a source/drain resistance equal to Rs = 50 $\Omega/\square$.

As for the traps in the channel/gate interface, employing the same gate-stack for short-channel Ge FETs is prohibitive, since scaling the transistors' length gives rise to further degradation in both I_{off} and SS. Interface trap densities lower than 5e11 $eV^{-1}cm^{-2}$ have been shown with a novel gate-stack consisting of GeO_2 and

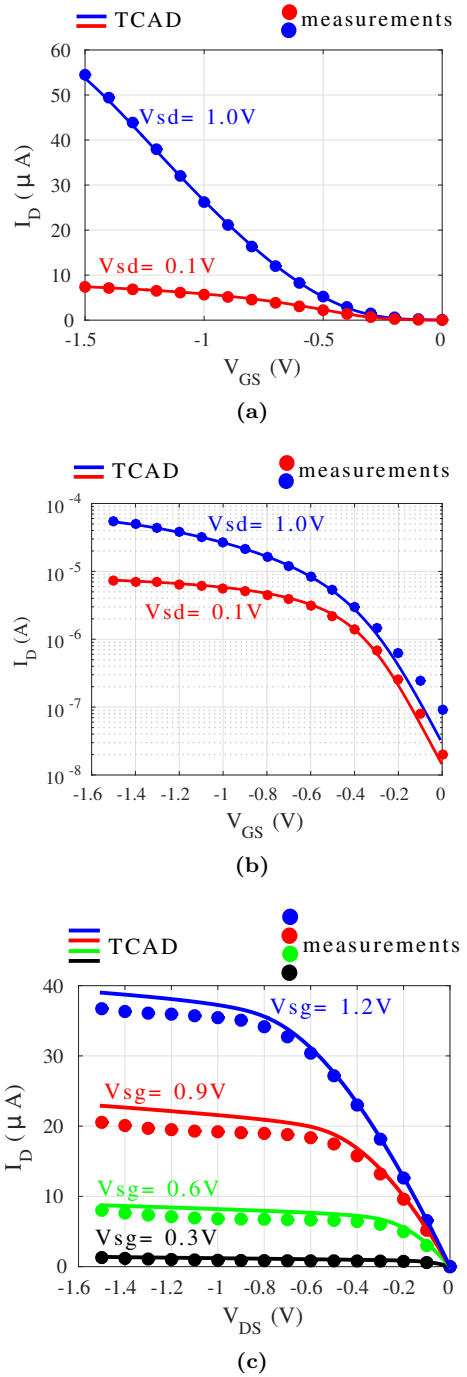


Figure 6.2: Measurement results for an in-house Ge transistor and corresponding TCAD simulations. I_D versus V_{GS} in (a) linear and (b) semi-logarithmic axis and (c) I_D versus V_{DS}.

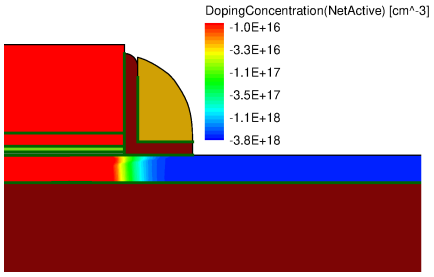


Figure 6.3: Net concentration of activated dopants in the channel

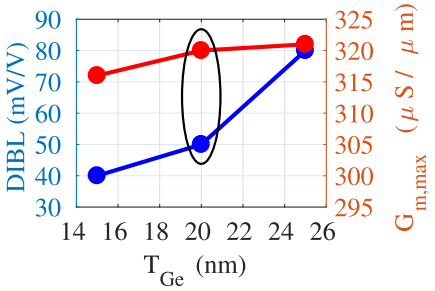


Figure 6.4: Dependence on DIBL and $G_{m,max}$ on the Ge layer's thickness, T_{Ge}.

Tm_2O_3 after one minute anneal at 500^{o}C at an O_2 ambient [142]. Thus, a value of D_{it} = 5e11 eV^{-1}cm^{-2} will be assumed henceforth.

With the aforementioned solutions for R_s and D_{it}, an ultra-short channel device would be a natural choice for a test-case. However, considering that the processing of Ge-FETs is still in research-phase, the length of the studied devices was set to 150 nm, to avoid any severe short-channel effects. To improve the gate control over the channel, the EOT of the gate stack has been also scaled to 3.3 nm, according to the corresponding value of a 150 nm commercial process, i.e. the base process in Chapter 3. To achieve this EOT, the gate stack was modified to $t_{GeO2} = 2.3$ nm and $t_{Al2O3} = 4$ nm. The thickness of the Ge layer has been also scaled to suppress short-channel effects. In general, the thinning of the active layer leads to less pronounced short-channel effects. This is clearly shown in Fig. 6.4. In the same figure, no severe degradation is observed for the the maximum transconductance, $G_{m,max}$: as T_{Ge} scales from 25 nm to 15 nm, $G_{m,max}$ drops from 321 μs/μm to 316 μS/μm. Therefore, the new value of the Ge thickness was set to $T_{Ge} = 20$ nm, in line with the capabilities of the in-house process. The simulation results are plotted in Fig. 6.5.

The Ge FETs are compared against TCAD simulations of Si p- and n-type

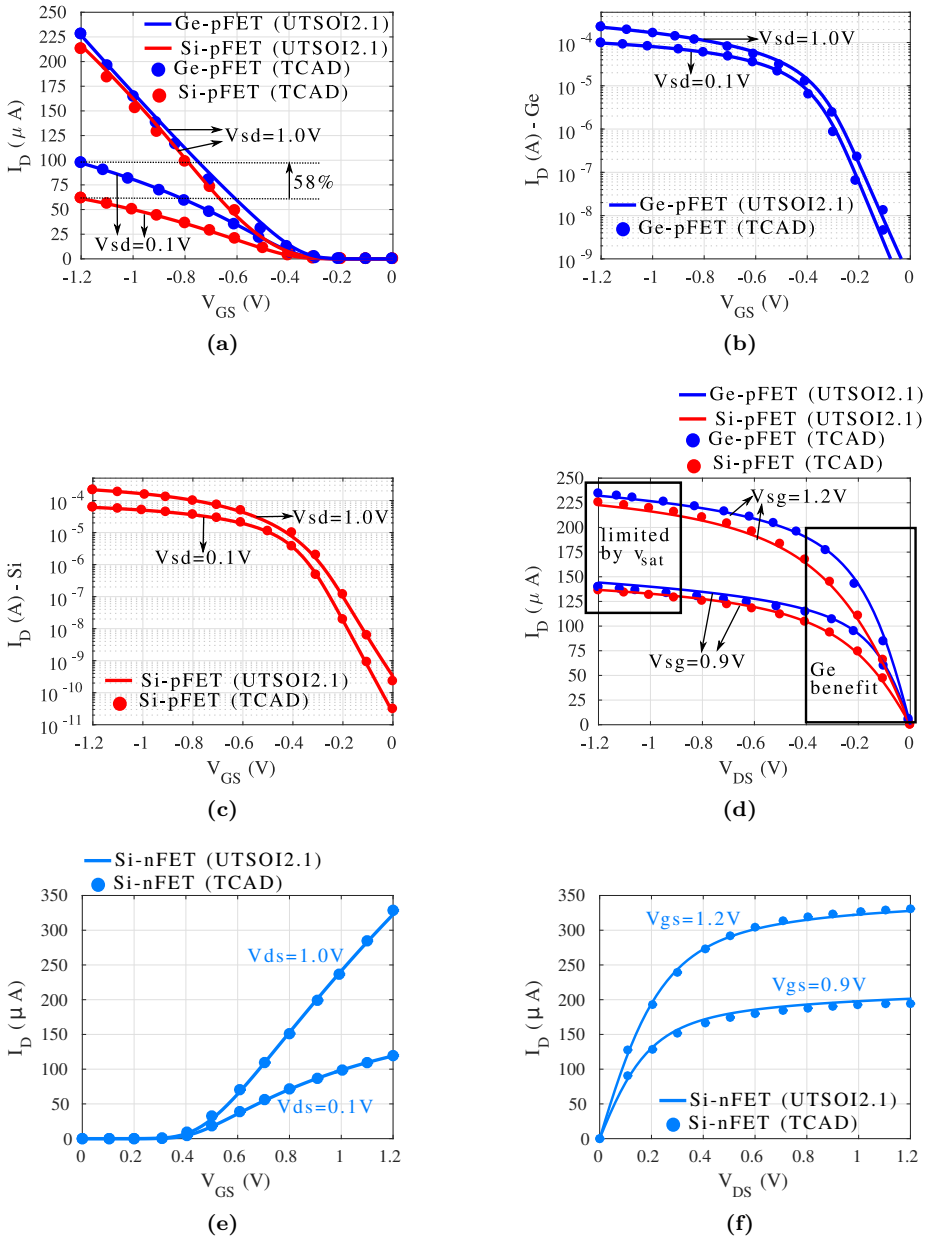


Figure 6.5: (a) Transfer characteristics of Ge and Si pFETs. TCAD and spice simulations for (b) Ge pFETs and (c) Si pFETs. (d) Output characteristics of the Ge and Si pFETs. (e) Transfer and (f) output characteristics of the Si nFET device.

Table 6.1: Comparison between the Ge pFET, the Si pFET and the Si nFET

Metric	Ge p-FET	Si p-FET	Si n-FET
SS (mV/dec)	90	75	70
DIBL (mV/V)	50	55	40
I_{on}/I_{off}	7e4	9e5	7e6

transistors with the same topology, i.e. the same length, EOT, T_{BOX} and $T_{Ge} = T_{Si} = 20$ nm. The source/drain resistance of the Si FETs was set to 50 Ω/μm, while no traps were considered for the channel/gate interface. The work function of the gate metal was calibrated to ensure the same threshold voltage for all three devices (in absolute value). To facilitate the study of potential applications of Ge FETs, UTSOI2.1 models have been calibrated to match the corresponding TCAD simulations. The results obtained with the UTSOI2.1 models are also included in Fig. 6.5.

The performance of these three devices is summarized in Table 6.1. The TmO_2-GeO_2 gate stack proposed in [142] leads to a spectacular improvement of the Ge device, with its SS being on par with the Si devices. Furthermore, the improvement in the SS and the reduced R_s obtained with the nickel-germanide lead to an increase in the I_{on}/I_{off} ratio, from 600 reported in [43] to 7e4. As for the short-channel effects, the scaling in EOT and T_{Ge} (T_{Si}) result in low DIBL for all three devices. The transfer characteristics in Fig. 6.5(a) indicate that for low V_{SD} values, i.e. when the transistor operates in the triode region ($V_{SD} = 0.1$ V), Ge p-FETs exhibit 58% higher current than the Si p-FETs, in line with the 58% higher Ge mobility. However, for the same biasing conditions, Si nFETs outperform Ge pFETs. In saturation ($V_{SD} = 1$ V), the higher current of Ge pFETs over their Si counterparts vanishes, with almost equal currents for both devices. Similar observations are drawn for the output characteristics in Fig. 6.5(d). The degradation of the Ge transistor's current in saturation is caused by velocity saturation, v_{sat}, as both Ge and Si devices exhibit similar values [143]. Hence, the most significant benefit of Ge pFETs over their Si counterparts occurs in the linear region, in which their higher dI_D/dV_D translates to a lower output impedance, R_{on}. Overall, applications that could benefit from a Ge-over-Si S3D integration scheme, are those at which the transistors do not operate continuously in saturation. Two examples of such applications, i.e. track-and-hold (TH) circuits and digital cells are examined in the following section. The intrinsic gain of the considered devices, A_v is plotted against the transistors' current in Fig. 6.6. In the sub-threshold region, the Ge-pFETs exhibit nearly double A_v compared to Si p-FETs and very close to the nFETs value. Therefore, a Ge-over-Si S3D approach, has also potential for ultra-low power applications. However, the intrinsic gain of Ge quickly rolls off in the saturation region, in which the nFETs outperform both the Si and Ge pFETs. A critical value exists for the biasing current, I_{crit}, for which Si and Ge pFETs exhibit the same

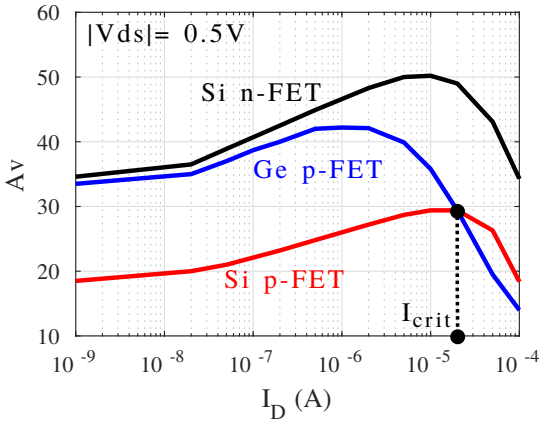


Figure 6.6: Intrinsic gain A_v of a Si nFET, a Si pFET and a Ge pFET. The dimensions of all three devices were L = 150 nm and W = 1 μm.

intrinsic gain. For currents larger than this value, the intrinsic gain of Ge pFETs drops faster than Si, and so top-tier Ge devices operating in saturation and biased at a current larger than the critical value should be avoided.

6.4 Applications Ge-Over-Si S3D Integration

As mentioned above, two types of circuits that could benefit from a Ge-over-Si S3D integration are THs and digital cells. The lower R_{on} of the Ge pFETs results in higher speed, which can be combined with the area reduction offered by S3D integration. The area reduction of S3D digital circuits has been extensively described in Chapter 2. The TH is a basic switch-capacitor circuit that tends to occupy a large area due to the large capacitors and the dummy structures used to meet stringent matching requirements. S3D integration could thus improve the area efficiency of switch capacitor circuits through a block-level partitioning, for instance, by placing other blocks under the capacitors.

For both the digital cells and the THs, nFETs are considered in the bottom tier and pFETs in the top.

6.4.1 Track-and-Hold Circuit

TH circuits are widely used in Analog-to-Digital Converters (ADCs). They sample an input signal and retain its value until its conversion to a digital word is completed. A TH typically consists of a switch and a hold-capacitor (C_{hold}), and a typical CMOS TH implementation is shown in Fig. 6.7. This sub-section compares

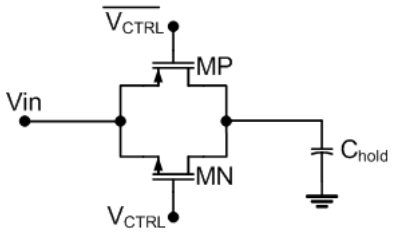


Figure 6.7: CMOS TH circuit with C_{hold} = 500 fF.

the performance of S3D THs for two cases: (a) with Si and (b) with Ge pFETs in the top-tier. This analysis is limited to the following metrics:

- $\mathbf{R}_{TH}$. It stands for the on-resistance of the switch. For the topology in Fig. 6.7, it is calculated by:

$$R_{TH} = R_{on,MP} // R_{on,MN} \tag{6.1}$$

- $\mathbf{f}_{s,max}$ is the maximum sampling frequency for which the TH tracks successfully the input signal. To estimate $f_{s,max}$, the worst case scenario for charging the capacitor C_{hold} is considered: charging it from the lowest input voltage to the maximum:

$$\Delta V_{load} = V_{in,PP}(1 - e^{-\frac{\Delta t}{\tau}}) \tag{6.2}$$

  with $\tau = R_{TH,max}C_{hold}$. Assuming that the hold capacitor is fully charged after approximately $\Delta t \approx 5\tau$, then, for a sampling signal with 50% duty cycle, the sampling period $T_s = 1/f_s$ needs to be larger than $2\Delta t$, and so:

$$f_{s,max} = \frac{1}{10 R_{TH} C_{hold}} \tag{6.3}$$

  Since $f_{s,max}$ is inversely proportional to R_{TH}, reducing R_{TH} is essential for high sampling rate applications.

- **Bandwidth in track mode ($\mathbf{BW}_T$)**. It refers to the maximum bandwidth of the input signal that can be effectively tracked when the switch is fully on. BW_T can be approximated by

$$BW_T = \frac{1}{2\pi R_{TH,max} C_{hold}} \tag{6.4}$$

  Equation (6.4) indicates that, similar to the $f_{s,max}$ case, minimizing R_{TH} improves the bandwidth of the tracked signal as well.

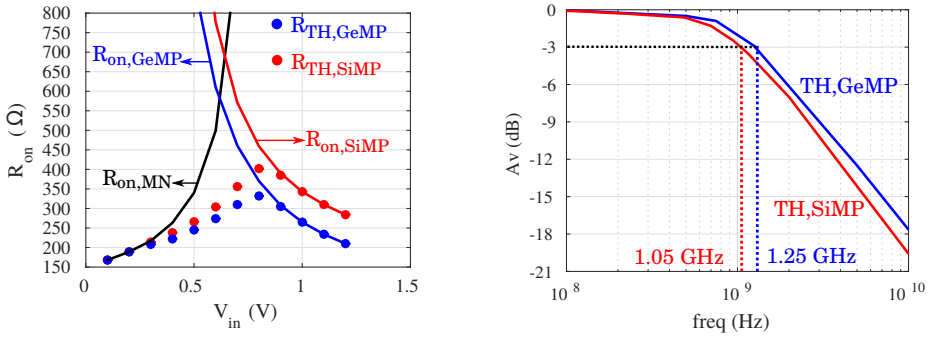


Figure 6.8: (a) R_{TH} for a Ge and a Si pFET (b) BW_T for a Ge and Si pFET.

A noise analysis of the S3D THs has not been considered, due to the lack of data on the noise of the Ge pFETs. However, the noise performance of Ge pFETs depends on the quality of their gate stacks (which affects the 1/f noise). The results for R_{TH} are plotted in Fig. 6.8(a). As expected, the output resistance of a Ge MP ($R_{on,GeMP}$) is lower than for the Si case ($R_{on,SiMP}$), leading to a lower $R_{TH,max}$ ($R_{TH,GeMP,max}$= 330 Ω, $R_{TH,SiMP,max}$= 400 Ω). The reduced value of R_{on} for the Ge case translates also to a 19% increase in BW_T (see Fig. 6.8(b)). As for $f_{s,max}$, substituting the extracted values in (6.3) results in 20% improvement in $f_{s,max,Ge}$ over $f_{s,max,Si}$ (600 MHz over 500 MHz). The performance of the TH for both cases of MP is summarized in Table 6.2. The use of Ge pFETs improves the speed of the TH, enabling it to operate at a higher sampling frequency and allowing higher bandwidth signals at its input. Furthermore, the S3D TH with Ge MP features a slight improvement in linearity over the Si-over-Si approach, as manifested by the higher IIP3. The impact of charge injection on the linearity of THs does not depend on the channel material of MP, as long as the EOT of the Ge and Si MPs are equal. Hence, this improvement in linearity should be attributed to the smaller variation in $R_{on,GeMP}$ when changing the input signal, as compared to $R_{on,SiMP}$ (see Fig. 6.8(a)). The improvement in speed and linearity for Ge-based S3D THs could make them suitable candidates for time-interleaved ADCs, which require high sampling frequency specifications for their input TH [144].

6.4.2 Digital Cells

Another circuit category that could benefit from the use of Ge-over-Si S3D integration is digital cells. S3D cells have been shown to trade-off area reduction for increased delays with respect to 2-D implementations [110]. This is mainly caused by the increased layout parasitics inside each cell. The use of Ge in the top-tier

Table 6.2: Performance comparison between a TH with a Ge and with a Si pFET

Metric	Ge pFET over Si nFET	Si pFET over Si nFET
$R_{TH,max}$	330 Ω	400 Ω
BW_T	1.25 GHz	1.05 GHz
$f_{s,max}$	600 MHz	500 MHz
IIP3 [1]	22 dBm	20 dBm

[1] measured with a 2 tones input signal (100 and 105 MHz), each at -6.5 dBm

could reduce this delay penalty, thanks to the reduced R_{on}, which in turn improves the delays related to the Pull Up Network (PUN). Therefore, a Ge-over-Si S3D integration scheme would result in S3D cells with lower rise time and the propagation delays related to the output pull-up. Ge pFETs, however, do not impact directly the operation of the Pull Down Network (PDN). Various S3D digital cells have been analyzed to quantify the impact of the Ge PUN on S3D cells. The results are compared against Si over Si S3D cells in Table 6.3. As expected, an average improvement of 14% in the cells' rise-time is observed for the Ge case, while the changes in the cells' fall time are insignificant. Since the reported propagation delay is averaged between the corresponding values related to the output pull-up and pull-down cases, the improvement in delay is less pronounced than for the rise time. The largest improvement is observed for NOR cells, the delay of which is dominated by the performance of the PUN (because of the cascoding of the pFETs). These observations could be extended to more advanced nodes as well. At such nodes, the onset of velocity saturation occurs at lower $|V_{DS}|$, and so the impact of the Ge's higher hole mobility on the saturation current is further minimized. However, this is not the the case in the triode region and so, the lower R_{on} of Ge pFETs, should be observed at scaled nodes as well.

6.5 Back-gate control of top-tier FETs

One of the benefits that FDSOIs bring to circuit design is the ability to control their threshold voltage at runtime through the back-gate terminal, as long as the thickness of the BOX is small enough. The applications of this Adaptive Back-gate Bias (ABB) feature of the Ultra Thin Body and BOX (UTBB) FDOIs range from compensating process and temperature variations to improving energy efficiency [145]. For instance, the ABB of UTBB FDOIs has been employed to improve the power consumption of a micro-controller built in a 28 nm UTBB FDSOI process [146]. This has been achieved by lowering the threshold voltage of the logic and memory transistors at the active-mode of the micro-controller, which, in turn, enabled the scaling of the supply voltage. Furthermore, the transistors' threshold voltage has been increased during the micro-controller's sleep-state to suppress leakage.

Table 6.3: Characteristics of the passive devices before and after stacking

Cells	**Average Delay[1](ps)**		**Rise Time[2](ps)**		**Fall Time[2](ps)**	
	Si / Si	Ge / Si	Si / Si	Ge / Si	Si / Si	Ge / Si
INV_X1	395	384	783	675	722	722
INV_X2	204	199	394	341	363	363
BUFF_X1	412	400	784	675	723	723
BUFF_X2	233	225	396	342	364	363
NAND_X1	494	483	790	681	1148	1148
NAND3_X1[3]	593	582	796	688	1585	1585
NOR_X1	520	473	1356	1042	725	725
NOR3_X1[3]	649	556	1952	1408	726	726

[1] average between the cases of rising and falling output
[2] measured between the 10% and 90% marks
[3] 3-input cell

However, ABB is not currently possible for the top-tier FETs in a S3D integration flow, owing to thick ILDs. ABB could be found particularly useful for Ge-over-Si S3D integration schemes. In particular, it could be applied to increase the threshold voltage of Ge FETs, V_{TH}, and thus reduce their notoriously high leakage for applications with low-leakage specifications. Furthermore, its application in suppressing process-variation could be found useful, owing to the novel gate stack and silicidation-like processes. Both process steps are under development and not fully mature and thus, they are expected to induce significant variations in a design.

In this work, a processing scheme has been conceived to enable effective back-gate biasing for top-tier Ge pFETs. It allows to control the Ge transistors' back gate from the bottom tier metals. The main processing steps of this scheme are as follows, and shown in Fig. 6.9.

- The process flow starts with a thick ILD that is deposited over the top-most metal layer of the bottom tier, MB (see Fig. 6.9(a))
- A contact-hole is then etched in the ILD to reach MB (see Fig. 6.9(b)).
- The next step is the metal filling of the contact-hole (see Fig. 6.9(c)). Care must be taken to guarantee the complete filling of the hole with full coverage and no voids.
- The metal needs to be etched away from all the areas outside the contact hole (see Fig. 6.9(d)). A CMP step could be also used to planarize the surface. This step, however, is the most critical one, as metal dishing caused by the CMP could result in variations in the BOX thickness.

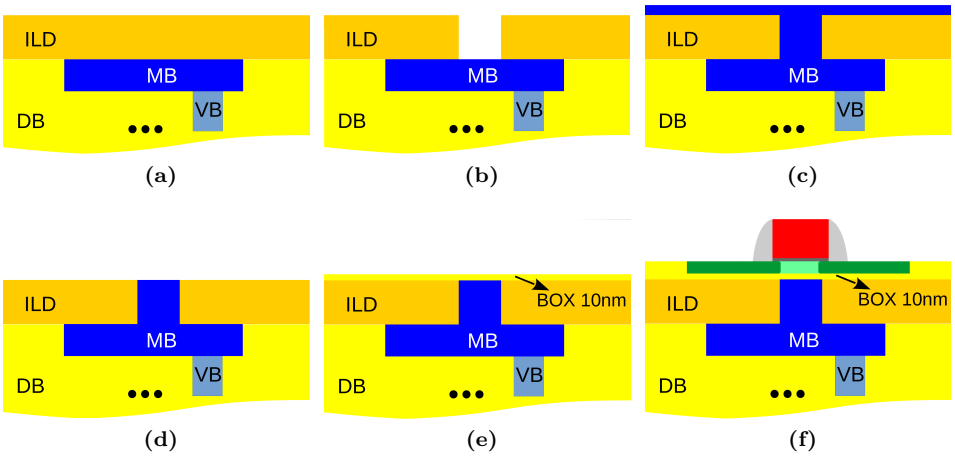


Figure 6.9: (a) Deposition of ILD over the last bottom tier metal, MB. (b) Etching of the ILD to form back-gate contact-hole. (c) Metal Filling of the contact-hole. (d) Etching and CMP of the metal from all the areas outside the contact-hole. (e) Formation of a 10 nm thick oxide, that will form the BOX of the top-tier FETs. (f) Processing of the top-tier FETs.

- A very thin SiO_2 layer is then deposited over the structure to serve as the BOX of the top tier transistors (see Fig. 6.9(e)). Its thickness in this work was assumed 10 nm, in line with current SOI processing capabilities [147].
- After the formation of the BOX, the Ge active layer can be transferred and bonded, followed by the processing of the top-tier FETs (see Fig. 6.9(f)).

To validate the effectiveness of this flow, the structure shown in Fig. 6.9(f) has been generated using Sentaurus Process from Synopsys® and simulated with Sentaurus Device. The topology of the Ge pFET is the same as in section 6.3. The simulated results for the transfer characteristics are plotted in Fig. 6.10, showing a 300 mV change in V_{TH} as the back-gate voltage sweeps from 0.6 V to -0.6 V. In particular a 0.6 V forward back-gate bias results in a 240 mV reduction in V_{TH}, whereas a 0.6 V reverse back-gate bias causes a 160 mV increase in V_{TH}. The large change in V_{TH} could be traded-off for a thicker BOX, which could better handle the metal dishing in the back-gate during the CMP step.

6.6 Summary

This chapter focused on the circuit design prospects of a Ge-over-Si S3D integration technology. An investigation study of the in-house Ge-over-Si S3D process has been

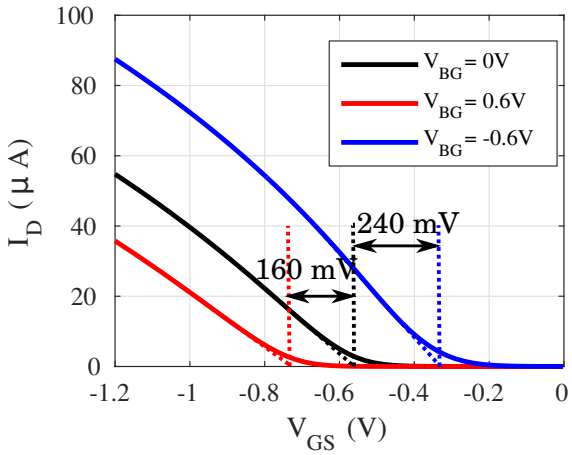


Figure 6.10: Transfer characteristics of a 150 nm Ge pFET for various back gate voltages V_{BG}. The back-gate biasing has been enabled through the devised process flow.

carried out and its current limitation to produce short-channel Ge devices has been identified. To solve these limitations, a number of solutions have been considered, and the performance of short-channel Ge FETs has been predicted with TCAD simulations. Track-and-hold circuits, as well as digital cells, have been identified as two types of circuits that could benefit from a Ge-over-Si S3D integration scheme in terms of improved speed and area efficiency. Lastly, a processing flow has been proposed to enable back-gate biasing in top-tier FETs. Doing so, the threshold voltage of top-tier Ge pFETs can be changed by up to 300 mV during runtime to boost the performance of specific circuits and compensate for process variations.

Chapter 7

Conclusions and Future Work

Sequential 3D (S3D) integration has been identified as a promising solution in the never ending quest for increasing the integration densities of ICs. This can be achieved by the sequential processing of device tiers with miniature-sized inter-tier vias. Furthermore, by stacking device tiers other than Si, additional functionalities can co-exist on the same die, i.e. logic, memory, radio, analog, etc. However, complex processing is required to harvest all these benefits. In particular, stringent requirements exist for the thermal budget of top-tier transistors. Failure to meet them will adversely impact the performance of both bottom tier transistors and interconnects. Hence, it comes as no surprise that extensive scientific attention has been directed to circumvent these processing issues. Building on these works, this thesis has investigated design methodologies and circuit techniques for high performance and area efficient S3D ICs. It also complements existing works that have focused on digital applications of S3D integration. However, unlike these works, the focus of this thesis has shifted away from digital applications. This has been motivated by the recent trends in the IC industry that favor hyper-connectivity and Internet of Things applications. Additionally, the performance variations between the top tier and bottom tier devices limit the effectiveness of the highly automated digital design flow.

Therefore, applications that could benefit from S3D integration needed to be identified. As a first step to achieve this, two S3D custom IC design platforms have been developed to allow the exploration of a wide range of applications, like analog, mixed-signal and radio. The first design platform aims at supporting the development of an in-house S3D process, as well as facilitating the close collaboration between process, device and circuit researchers. The second one aims at investigating potential S3D applications. Both design platforms include parameterized cells, physical verification flows and device models. Special care has been taken to ensure that the unique features of the top-tier devices (i.e. SOIs with a thick BOX) are considered in the device models. Furthermore, the device models have been calibrated to match the performance of measured devices as well as

TCAD simulations. In the scope of this thesis, a parasitic extraction flow for S3D ICs has been developed to study the impact of layout parasitics, and most importantly to identify potential issues related to the coupling of bottom-tier and top-tier geometries.

With the developed S3D design platforms, the potential of S3D RF/AMS circuits and systems has been identified. The main motivation behind this has been the FDSOI nature of the top-tier transistors and consequently, their better performance at high frequencies. Based on this observation, a frequency-based partition scheme has been proposed with high frequency blocks in the top-tier, implemented with FDSOI transistors and low-frequency ones in the bottom with bulk transistors. Among the various radio standards, the low-power short-range ones (i.e. Bluetooth, ZigBee, etc) appear to have the largest potential for S3D integration. As a proof of concept, a receiver front-end for the ZigBee standard has been designed in the S3D design platform. Its evaluation against the corresponding 2-D implementation has shown a 35% reduction in area and no performance degradation.

The study on the potential of S3D RF/AMS circuits would not be complete without analyzing the impact of S3D integration on inductors. Towards this, planar inductors in the top-most thick metal layer have been found as the most optimal configuration for S3D integration. Furthermore, if a S3D process is built over an existing one, then transferring an inductor from the initial process to the top-most metal layer of the derived S3D process, results in very small changes in its quality factor and inductance. Thus, for the most S3D cases, the inductor spice models of the initial process could be re-used. Patterned ground shields have been found instrumental at suppressing substrate noise coupling to top-tier S3D inductors. However, they do not appear to offer any benefits in terms of the inductors' quality factors. To improve the inductors' quality factors, the potential of shunting multiple top-tier metal layers has been also investigated, in order to reduce the inductors' series resistance. It was found that the stacked metal layers of the inductor give rise to substantial negative inductive coupling and proximity effects, causing a drop in both the inductance and the quality factors. Since every attempt to boost the inductors' quality factors was proved fruitless, the focus of this thesis shifted to ways of improving their area efficiency, namely by adding bottom tier blocks (analog and digital) under them. Towards this, a set of guidelines has been proposed to minimize the electromagnetic coupling between the inductor and the bottom tier blocks as well as to counteract any impact on the inductors' performance. These guidelines have taken into consideration the type of blocks in the bottom tier (i.e. analog, radio, or digital) and their operating frequency.

Lastly, the potential of heterogeneous S3D integration from a circuit design perspective has been also explored. The study has focused on a Ge-over-Si S3D integration flow, like the in-house S3D process. Because of the restriction of the in-house process to long-channel Ge devices, TCAD simulations have been employed to predict and analyze the performance of short-channel Ge transistors. The most recent breakthroughs in the processing of Ge transistors, as found in the corresponding literature, were incorporated in the analysis. The results indicated that,

despite the higher mobility of Ge transistors, their current in saturation is severely constrained by velocity saturation. Hence, to exploit the higher mobility of Ge devices, applications with transistors operating in the triode region were considered, in particular track-and-hold circuits and digital cells. To improve the performance of Ge top-tier transistors, a processing flow to enable back-gate biasing for the top-tier FETs was proposed, resulting in a 300 mV change in their threshold voltage as the back bias swept from -0.6V to +0.6V.

All in all, this thesis has achieved all the research objectives that were set in Chapter 1. However, under no circumstances it should be considered a complete work on S3D design methodologies. Some possible aspects to be considered in the future are:

- Investigating the impact of top-tier inductors on the performance of bottom tier transistors. Due to the current limitations of existing tools that do not allow electromagnetic and transistor-level co-simulations, the analysis in Chapter 5 was limited to bottom-tier resistors. Of course, the findings can be extended to transistors, assuming insignificant coupling between the channel area and inductors. However, to verify this assumption, methods to circumvent the current limitations of CAD tools need to be proposed, or preferably, actual S3D circuits with bottom tier blocks under inductors could be fabricated and measured.

- Building on the results reached for S3D RF/AMS circuits, the potential of S3D mixed-signal systems should be also explored. For instance, a complete receiver chain could be designed and validated for a complete study of S3D RF/AMS systems.

- The study of S3D heterogeneous integration needs to be extended to cover other approaches as well (i.e. memristors, Carbon nano-tubes, vertical FETs, etc). Furthermore, building on the prospects of S3D RF/AMS circuits that were demonstrated in this thesis, a noise analysis of Ge and III-V materials need to be carried out for a complete study of S3D heterogeneous integration.

Bibliography

[1] ITRS, "International Technology Roadmap for Semiconductors, Edition 2015, Beyond C-MOS," *Int. Technol. Roadmap Semicond.*, 2015.

[2] A. Abedin, L. Zurauskaite, A. Asadollahi, K. Garidis, G. Jayakumar *et al.*, "GOI fabrication for monolithic 3D integration," in *IEEE SOI-3D-Subthreshold Microelectron. Unified Conf. S3S*, 2017.

[3] A. Liscidini, M. Tedeschi, and R. Castello, "A 2.4 GHz 3.6mW 0.35mm 2 Quadrature Front-End RX for ZigBee and WPAN Applications," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Pap.* San Francisco, CA: IEEE, 2008, pp. 370 – 620.

[4] G. E. Moore, "Craming more components onto integrated circuits. Reprinted from Electronics, Volume 38, Number 8, April 19, 1965, pp.114 ff." *IEEE Solid-State Circuits Soc. Newsl.*, vol. 11, no. 3, pp. 33–35, 2006.

[5] S. Borkar, "Design challenges of technology scaling," *IEEE Micro*, vol. 19, no. 4, pp. 23–29, 1999.

[6] T. Ghani, M. Armstrong, C. Auth, M. Bost, P. Charvat *et al.*, "A 90nm," in *IEEE Int. Electron Devices Meet.*, 2003.

[7] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom *et al.*, "A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging," in *IEEE Int. Electron Devices Meet.*, 2007.

[8] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder *et al.*, "A 14nm Logic Technology Featuring 2nd-Generation FinFET Transistors, Air-Gapped Interconnects, Self-Aligned Double Patterning and a 0.0588mm2 SRAM cell size," in *IEEE Int. Electron Devices Meet.* San Francisco, CA: IEEE, 2014, pp. 3.7.1–3.7.3.

[9] L. W. Liebmann and R. O. Topaloglu, "Design and technology co-optimization near single-digit nodes," *IEEE/ACM Int. Conf. Comput. Des. Dig. Tech. Pap. ICCAD*, vol. 2015-Janua, no. January, pp. 582–585, 2015.

[10] B. Turkot, S. Carson, and A. Lio, "Continuing moore's law with EUV lithography," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 14.4.1–14.4.3, 2018.

[11] B. Turkot, S. L. Carson, A. Lio, T. Liang, M. Phillips *et al.*, "EUV progress toward HVM readiness," *Proc.SPIE*, vol. 9776, 2016.

[12] A. Loke, "Analog/Mixed-Signal Design in 7-nm CMOS and Beyond," in *IEEE Cust. Integr. Circuits Conf.*, 2018.

[13] D. Prasad, C. Pan, and A. Naeemi, "Impact of interconnect variability on circuit performance in advanced technology nodes," *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, vol. 2016-May, pp. 398–404, 2016.

[14] C. H. Jan, U. Bhattacharya, R. Brain, S. J. Choi, G. Curello *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 *Int. Electron Devices Meet.*, 2012.

[15] V. Roche, "Semiconductor innovation: Is the party over, or just getting started?" *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 61, pp. 8–11, 2018.

[16] J. Kim, K. Lee, D. Park, T. Hwang, K. Kim *et al.*, "Application of Through Mold Via (TMV) as PoP base package," *Proc. - Electron. Components Technol. Conf.*, pp. 1089–1092, 2008.

[17] I. Qin, O. Yauw, G. Schulze, A. Shah, B. Chylak *et al.*, "Advances in Wire Bonding Technology for 3D Die Stacking and Fan Out Wafer Level Package," in *Proc. - Electron. Components Technol. Conf.*, 2017.

[18] E. Beyne, "The 3-D Interconnect Technology Landscape," *IEEE Des. Test*, vol. 33, no. 3, pp. 8–20, 2016.

[19] S. V. Huylenbroeck, Y. Li, N. Heylen, K. Croes, G. Beyer *et al.*, "Advanced Metallization Scheme for 3x50um Via Middle TSV and Beyond," in *IEEE Electron. Components Technol. Conf.*, 2015.

[20] S. Van Huylenbroeck, M. Stucchi, Y. Li, J. Slabbekoorn, N. Tutunjyan *et al.*, "Small Pitch, High Aspect Ratio Via-Last TSV Module," in *IEEE Electron. Components Technol. Conf.*, 2016.

[21] U. Kang, H. J. Chung, S. Heo, D. H. Park, H. Lee *et al.*, "8 Gb 3-D DDR3 DRAM using through-silicon-via technology," *IEEE J. Solid-State Circuits*, vol. 45, no. 1, pp. 111–119, 2010.

[22] Z. Or-Bach, B. Cronquist, Z. Wurman, I. Beinglass, and A. Henning, "Precision bonders - A game changer for monolithic 3D," in *SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, Millbrae, CA, 2014.

[23] Z. Or-Bach, B. Cronquist, Z. Wurman, I. Beinglass, and A. K. Henning, "Modified ELTRAN®-A game changer for Monolithic 3D," in *IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, 2015.

[24] P. Batude, C. Fenouillet-Beranger, L. Pasini, V. Lu, F. Deprat *et al.*, "3DVLSI with CoolCube process: An alternative path to scaling," in *Symp. VLSI Technol.*, Kyoto, 2015, pp. T48–T49.

[25] P. Batude, T. Ernst, J. Arcamone, G. Arndt, P. Coudrain *et al.*, "3-D sequential integration: A key enabling technology for heterogeneous Co-integration of new function with CMOS," *IEEE J. Emerg. Sel. Top. Circuits Syst.*, vol. 2, no. 4, pp. 714–722, 2012.

[26] S. K. Samal, D. Nayak, M. Ichihashi, S. Banna, and S. K. Lim, "Monolithic 3D IC vs. TSV-based 3D IC in 14nm FinFET technology," *IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, 2016.

[27] B. De Salvo, "Brain-Inspired technologies: Towards chips that think?" *Dig. Tech. Pap. - IEEE Int. Solid-State Circuits Conf.*, vol. 61, pp. 12–18, 2018.

[28] L. Liebmann, J. Zeng, X. Zhu, L. Yuan, G. Bouche *et al.*, "Overcoming scaling barriers through design technology CoOptimization," in *Dig. Tech. Pap. - Symp. VLSI Technol.*, 2016.

[29] S. Panth, K. Samadi, Y. Du, and S. K. Lim, "Power-Performance Study of Block-Level Monolithic 3D-ICs Considering Inter-Tier Performance Variations," *ACM/EDAC/IEEE Des. Autom. Conf.*, 2014.

[30] S. Panth, K. Samadi, Y. Du, and S. K. Lim, "Shrunk-2-D: A Physical Design Methodology to Build Commercial-Quality Monolithic 3-D ICs," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 36, no. 10, pp. 1716–1724, 2017.

[31] Y. Chen and D. H. Kim, "A legalization algorithm for multi-tier gate-level monolithic three-dimensional integrated circuits," *Proc. - Int. Symp. Qual. Electron. Des. ISQED*, pp. 277–282, 2017.

[32] S. E. D. Lin and D. H. Kim, "Detailed-Placement-Enabled Dynamic Power Optimization of Multi-Tier Gate-Level Monolithic 3-D ICs," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 4, pp. 845–854, 2017.

[33] I. Hong and D. H. Kim, "Analysis of Performance Benefits of Multitier Gate-Level Monolithic 3-D Integrated Circuits," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 37, no. 8, pp. 1614–1626, 2018.

[34] J. Shi, D. Nayak, S. Banna, R. Fox, S. Samavedam *et al.*, "A 14nm FinFET transistor-level 3D partitioning design to enable high-performance and low-cost monolithic 3D IC," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 2.5.1–2.5.4, 2017.

[35] C. Yan and E. Salman, "Mono3D: Open Source Cell Library for Monolithic 3-D Integrated Circuits," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 65, no. 3, pp. 1075–1085, 2018.

[36] A. Guler, S. Member, and N. K. Jha, "Hybrid Monolithic 3-D IC Floorplanner," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 26, no. 10, pp. 1868–1880, 2018.

[37] H. Sarhan, S. Thuries, O. Billoint, and F. Clermidy, "An Unbalanced Area Ratio Study for High Performance Monolithic 3D Integrated Circuits," in *IEEE Comput. Soc. Annu. Symp. VLSI*, 2015.

[38] R. Boumchedda, J.-p. Noel, B. Giraud, K. C. Akyel, M. Brocard *et al.*, "High-Density 4T SRAM Bitcell in 14-nm 3-D CoolCube Technology Exploiting Assist Techniques," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, no. 8, pp. 2296–2306, 2017.

[39] J. Kong, Y. H. Gong, and S. W. Chung, "Architecting large-scale SRAM arrays with monolithic 3D integration," *IEEE/ACM Int. Symp. Low Power Electron. Des.*, 2017.

[40] W. H. Chen, C. F. Chen, Y. J. Chen, H. Y. Chiu, C. H. Shen *et al.*, "A Dual-Split-Controlled 4P2N 6T SRAM in Monolithic 3D-ICs With Enhanced Read Speed and Cell Stability for IoT Applications," *IEEE Electron Device Lett.*, vol. 39, no. 8, pp. 1167–1170, 2018.

[41] M. M. Shulaker, T. F. Wu, A. Pal, L. Zhao, Y. Nishi *et al.*, "Monolithic 3D integration of logic and memory: Carbon nanotube FETs, resistive RAM, and silicon FETs," *IEEE Int. Electron Devices Meet.*, 2014.

[42] S.-h. Kim, S.-k. Kim, J.-p. Shim, D.-m. Geum, G. Ju *et al.*, "Heterogeneous Integration Toward a Monolithic 3-D Chip Enabled by III-V and Ge Materials," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 579–587, 2018.

[43] A. Abedin, L. Zurauskaite, A. Asadollahi, K. Garidis, G. Jayakumar *et al.*, "Germanium on Insulator Fabrication for Monolithic 3-D Integration," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 588–593, 2018.

[44] J. Shi, D. Nayak, M. Ichihashi, S. Banna, and C. A. Moritz, "On the Design of Ultra-High Density 14nm Finfet based Transistor-Level Monolithic 3D ICs," in *IEEE Comput. Soc. Annu. Symp. VLSI*, 2016.

[45] F. Clermidy, O. Billoint, H. Sarhan, and S. Thuries, "Technology scaling: the CoolCube paradigm," in *IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf.*, 2015.

[46] D. Henry, X. Baillin, V. Lapras, M. H. Vaudaine, J. M. Quemper *et al.*, "Via first technology development based on high aspect ratio trenches filled with doped polysilicon," in *Electron. Components Technol. Conf.*, 2007.

[47] C. Laviron, B. Dunne, V. Lapras, P. Galbiati, D. Henry *et al.*, "Via first approach optimisation for through silicon via applications," in *Electron. Components Technol. Conf.*, 2009.

[48] E. Beyne, "Reliable via-middle copper through-silicon via technology for 3-D integration," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 6, no. 7, pp. 985–994, 2016.

[49] J. Knechtel, I. L. Markov, and J. Lienig, "Assembling 2-D Blocks Into 3-D Chips," *Comput. Des. Integr. Circuits Syst. IEEE Trans.*, vol. 31, no. 2, pp. 228–241, 2012.

[50] D. H. Kim, K. Athikulwongse, and S. K. Lim, "Study of through-silicon-via impact on the 3-D stacked IC layout," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 21, no. 5, pp. 862–874, 2013.

[51] K. J. Chui, H. Y. Li, K. F. Chang, S. Bhattacharya, and M. Yu, "A cost model analysis comparing via-middle and via-last TSV processes," *Proc. Electron. Packag. Technol. Conf. EPTC*, vol. 2016-Febru, pp. 5–8, 2016.

[52] E. Beyne, A. Jourdain, and A. Phommahaxay, "Thinning, Via Reveal, and Backside Processing - Overview," in *Handb. 3D Integr.*, P. Garrou, M. Koyanagi, and P. Ramm, Eds., 2014. [Online]. Available: doi:10.1002/9783527670109.ch15

[53] I. De Wolf, K. Croes, O. Varela Pedreira, R. Labie, A. Redolfi *et al.*, "Cu pumping in TSVs: Effect of pre-CMP thermal budget," *Microelectron. Reliab.*, vol. 51, no. 9-11, pp. 1856–1859, 2011.

[54] C. J. Wilson, I. De Wolf, B. Vandevelde, J. De Messemaeker, J. M. Ablett *et al.*, "Comparison of x-ray diffraction, wafer curvature and Raman spectroscopy to evaluate the stress evolution in Copper TSV's," *IEEE Int. Interconnect Technol. Conf. IITC 2012*, 2012.

[55] E. Beyne, J. De Messemaeker, and W. Guo, "Cu TSV Stress: Avoiding Cu Protrusion and Impact on Devices," in *Handb. 3D Integr.*, 2014, vol. 3, ch. 27, pp. 365–378. [Online]. Available: https://doi.org/10.1002/9783527670109.ch27

[56] A. Mercha, G. Van Der Plas, V. Moroz, I. De Wolf, P. Asimakopoulos *et al.*, "Comprehensive analysis of the impact of single and arrays of through silicon vias induced stress on high-k / metal gate CMOS performance," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 2.2.1–2.2.4, 2010.

[57] W. Guo, G. Van Der Plas, A. Ivankovic, V. Cherman, G. Eneman *et al.*, "Impact of through silicon via induced mechanical stress on fully depleted Bulk FinFET technology," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 18.4.1–18.4.4, 2012.

[58] W. Guo, V. Moroz, G. Van Der Plas, M. Choi, A. Redolfi *et al.*, "Copper through silicon via induced keep out zone for 10nm node bulk FinFET CMOS technology," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 340–343, 2013.

[59] C. Huang, Q. Chen, and Z. Wang, "Air-gap through-silicon vias," *IEEE Electron Device Lett.*, vol. 34, no. 3, pp. 441–443, 2013.

[60] R. Agarwal, W. Zhang, P. Limaye, R. Labie, B. Dimcic *et al.*, "Cu/Sn microbumps interconnect for 3D TSV chip stacking," *Proc. - Electron. Components Technol. Conf.*, vol. 1555, no. ii, pp. 858–863, 2010.

[61] D. Dutoit, E. Guthmuller, and I. Miro-Panades, "3D Integration for Power-Efficient Computing," in *Des. Autom. Test Eur. Conf. Exhib. (DATE), 2013*, 2013, pp. 779–784.

[62] P. Enquist, "Scalable direct bond technology and applications driving adoption," *2011 IEEE Int. 3D Syst. Integr. Conf. 3DIC 2011*, 2011.

[63] P. Batra, D. Latulipe, S. Skordas, K. Winstel, C. Kothandaraman *et al.*, "Three-dimensional wafer stacking using Cu TSV integrated with 45nm high performance SOI-CMOS embedded DRAM technology," *2013 IEEE SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. S3S 2013*, vol. 43, no. 1, pp. 7–8, 2013.

[64] D. Y. Chen, W. C. Chiou, M. F. Chen, T. D. Wang, K. M. Ching *et al.*, "Enabling 3D-IC foundry technologies for 28 nm node and beyond: Through-silicon-via integration with high throughput die-to-wafer stacking," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, pp. 353–356, 2009.

[65] T. Fukushima, E. Iwata, Y. Ohara, M. Murugesan, J. Bea *et al.*, "Multichip self-assembly technology for advanced die-to-wafer 3-D integration to precisely align known good dies in batch processing," *IEEE Trans. Components, Packag. Manuf. Technol.*, vol. 1, no. 12, pp. 1873–1884, 2011.

[66] T. Mourier, M. Gottardi, C. Ribiere, S. Minoret, P.-E. Philip *et al.*, "Advanced Metallizatisn Processes Integration as Manufacturing Worthy Solutions for $>$ 10:1 Aspect Ratio Mid-Process TSV," *2018 IEEE Int. Interconnect Technol. Conf.*, 2018.

[67] B. Rajendran, "Sequential 3D IC Fabrication-Challenges and Prospects," *VLSI Multi-level Interconnect Conf.*, 2006.

[68] P. Batude, M. Vinet, A. Pouydebasque, C. Le Royer, B. Previtali *et al.*, "Advances in 3D CMOS sequential integration," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, no. 110, pp. 345–348, 2009.

[69] C. Fenouillet-Beranger, B. Mathieu, B. Previtali, M. P. Samson, N. Rambal *et al.*, "New insights on bottom layer thermal stability and laser annealing promises for high performance 3D VLSI," in *Tech. Dig. - Int. Electron Devices Meet. IEDM*, no. February. San Francisco, CA: IEEE, 2015, pp. 27.5.1–27.5.4.

[70] A. Vandooren, L. Witters, J. Franco, A. Mallik, B. Parvais *et al.*, "Sequential 3D: Key integration challenges and opportunities for advanced semiconductor scaling," in *Int. Conf. IC Des. Technol.*, 2018.

[71] L. Brunet, P. Batude, C. Fenouillet-Beranger, P. Besombes, L. Hortemel *et al.*, "First demonstration of a CMOS over CMOS 3D VLSI CoolCube integration on 300mm wafers," 2016.

[72] A. Ayres, O. Rozeau, B. Borot, L. Fesquet, G. Cibrario *et al.*, "Guidelines on 3DVLSI design regarding the intermediate BEOL process influence," in *SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. (S3S), IEEE*. Rohnert Park, CA: IEEE, 2015, pp. 1–2.

[73] C. Fenouillet-Beranger, P. Acosta-Alba, B. Mathieu, S. Kerdilès, M. P. Samson *et al.*, "Ns laser annealing for junction activation preserving inter-tier interconnections stability within a 3D sequential integration," *2016 SOI-3D-Subthreshold Microelectron. Technol. Unified Conf. S3S 2016*, pp. 2016–2017, 2017.

[74] A. Vandooren, J. Franco, B. Parvais, Z. Wu, L. Witters *et al.*, "3-D Sequential Stacked Planar Devices Featuring Low-Temperature Replacement Metal Gate Junctionless Top Devices With Improved Reliability," *IEEE Trans. Electron Devices*, vol. 65, no. 11, pp. 5165 – 5171, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8487028/

[75] M. M. Shulaker, T. F. Wu, M. M. Sabry, H. Wei, H. P. Wong *et al.*, "Monolithic 3D Integration : A Path From Concept To Reality," in *Des. Autom. Test Eur. Conf. Exhib.*, 2015.

[76] M. M. Shulaker, K. Saraswat, and H. S. P. Wong, "Monolithic Three-Dimensional Integration of Carbon Nanotube FETs with Silicon CMOS," in *Symp. VLSI Technol.*, 2014.

[77] H. Wei, M. Shulaker, H. S. Wong, and S. Mitra, "Monolithic three-dimensional integration of carbon nanotube FET complementary logic circuits," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, 2013.

[78] N. Patil, A. Lin, E. R. Myers, H. S. Philip Wong, and S. Mitra, "Integrated wafer-scale growth and transfer of directional carbon nanotubes and misaligned-carbon-nanotube-immune logic structures," in *Dig. Tech. Pap. - Symp. VLSI Technol.*, 2008.

[79] X. Sun, C. D'Emic, C. W. Cheng, A. Majumdar, Y. Sun *et al.*, "High performance and low leakage current InGaAs-on-silicon FinFETs with 20 nm gate length," in *Dig. Tech. Pap. - Symp. VLSI Technol.*, 2017.

[80] P. Goley and M. Hudait, "Germanium Based Field-Effect Transistors: Challenges and Opportunities," *Materials (Basel).*, vol. 7, no. 3, pp. 2301–2339, mar 2014. [Online]. Available: http://www.mdpi.com/1996-1944/7/3/2301/htm

[81] V. Deshpande, H. Hahn, E. O'Connor, Y. Baumgartner, D. Caimi *et al.*, "Demonstration of 3-D SRAM Cell by 3-D Monolithic Integration of InGaAs n-FinFETs on FDSOI CMOS with Interlayer Contacts," *IEEE Trans. Electron Devices*, vol. 64, no. 11, pp. 4503–4509, 2017.

[82] V. Deshpande, V. Djara, E. O'Connor, P. Hashemi, K. Balakrishnan *et al.*, "DC and RF characterization of InGaAs replacement metal gate (RMG) nFETs on SiGe-OI FinFETs fabricated by 3D monolithic integration," *Solid. State. Electron.*, vol. 128, pp. 87–91, 2017.

[83] V. Deshpande, H. Hahn, V. Djara, E. O. Connor, D. Caimi *et al.*, "Hybrid InGaAs / SiGe CMOS Circuits with 2D and 3D Monolithic Integration," in *47th Eur. Solid-State Device Res. Conf.*, 2017.

[84] K. Garidis, G. Jayakumar, A. Asadollahi, E. Dentoni Litta, P. E. Hellström *et al.*, "Characterization of bonding surface and electrical insulation properties of inter layer dielectrics for 3D monolithic integration," *EUROSOI-ULIS 2015 - 2015 Jt. Int. EUROSOI Work. Int. Conf. Ultim. Integr. Silicon*, no. c, pp. 165–168, 2015.

[85] S. K. Samal, S. Panth, K. Samadi, M. Saeidi, Y. Du *et al.*, "Fast and Accurate Thermal Modeling and Optimization for Monolithic 3D ICs," *Proc. 51st Annu. Des. Autom. Conf. Des. Autom. Conf. - DAC '14*, pp. 1–6, 2014.

[86] C. Santos, P. Vivet, S. Thuries, O. Billoint, J. P. Colonna *et al.*, "Thermal performance of CoolCube monolithic and TSV-based 3D integration processes," *2016 IEEE Int. 3D Syst. Integr. Conf. 3DIC 2016*, 2017.

[87] S. K. Samal, D. Nayak, M. Ichihashi, S. Banna, and S. K. Lim, "Tier partitioning strategy to mitigate BEOL degradation and cost issues in monolithic 3D ICs," *Proc. 35th Int. Conf. Comput. Des. - ICCAD '16*, pp. 1–7, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2966986.2967080

[88] B. W. Ku, T. Song, A. Nieuwoudt, S. K. Lim, and M. View, "Transistor-Level Monolithic 3D Standard Cell Layout Optimization for Full-Chip Static Power Integrity," in *IEEE/ACM Int. Symp. Low Power Electron. Des.*, 2017.

[89] G. Cibrario, N. B. Salem, J. Lacord, K. Azizi-mourier, O. Rozeau *et al.*, "From 2D to Monolithic 3D Predictive Design Platform: An Innovative Migration Methodology for Benchmark Purpose," in *IEEE Int. 3D Syst. Integr. Conf.*, 2016.

[90] K. Chang, S. Sinha, B. Cline, R. Southerland, M. Doherty *et al.*, "Cascade2D: A Design-Aware Partitioning Approach to Monolithic 3D IC with 2D Commercial Tools," *Proc. 35th Int. Conf. Comput. Des. - ICCAD '16*, pp. 1–8, 2016. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2966986.2967013

[91] Y.-J. Lee, D. Limbrick, and S. K. Lim, "Power benefit study for ultra-high density transistor-level monolithic 3D ICs," in *Des. Autom. Conf. (DAC), ACM/EDAC/IEEE*, 2013, p. 1.

[92] K. Chang, K. Acharya, S. Sinha, B. Cline, G. Yeric *et al.*, "Impact and Design Guideline of Monolithic 3-D IC at the 7-nm Technology Node," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, no. 7, pp. 2118–2129, 2017.

[93] E. Baer, A. Burenkov, P. Evanschitzky, and J. Lorenz, "Simulation of process variations in FinFET transistor patterning," *Int. Conf. Simul. Semicond. Process. Devices, SISPAD*, no. Cd, pp. 299–302, 2016.

[94] X. Wang, D. Reid, L. Wang, C. Millar, A. Burenkov *et al.*, "Process informed accurate compact modelling of 14-nm FinFET variability and application to statistical 6T-SRAM simulations," *Int. Conf. Simul. Semicond. Process. Devices, SISPAD*, vol. 1, pp. 303–306, 2016.

[95] S. K. Saha, "Modeling process variability in scaled CMOS technology," *IEEE Des. Test Comput.*, vol. 27, no. 2, pp. 8–16, 2010.

[96] T. Hiramoto, A. Kumar, T. Mizutani, J. Nishimura, and T. Saraya, "Statistical advantages of intrinsic channel fully depleted SOI MOSFETs over bulk MOSFETs," *Proc. Cust. Integr. Circuits Conf.*, pp. 8–11, 2011.

[97] E. G. Ioannidis, C. G. Theodorou, S. Haendler, E. Josse, C. A. Dimitriadis *et al.*, "Impact of Source-Drain Series Resistance on Drain Current Mismatch in Advanced Fully Depleted SOI n-MOSFETs," *IEEE Electron Device Lett.*, vol. 36, no. 5, pp. 433–435, 2015.

[98] Y. S. Chauhan, S. Venugopalan, M. A. Karim, S. Khandelwal, N. Paydavosi *et al.*, "BSIM Industry standard compact MOSFET models," *Eur. Solid-State Circuits Conf.*, 2012.

[99] T. Poiroux, O. Rozeau, P. Scheer, S. Martinie, M. A. Jaud *et al.*, "Leti-UTSOI2 . 1 : A Compact Model for UTBB-FDSOI Technologies - Part I : Interface Potentials Analytical Model," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2751–2759, 2015.

[100] T. Poiroux, O. Rozeau, P. Scheer, S. Martinie, M.-a. Jaud *et al.*, "Leti-UTSOI2 . 1 : A Compact Model for UTBB-FDSOI Technologies - Part II : DC and AC Model Description," *IEEE Trans. Electron Devices*, vol. 62, no. 9, pp. 2760–2768, 2015.

[101] *QRC Techgen Reference Manual*. Cadence, 2017.

[102] *QRC Extraction Users Manual*. Cadence, 2017.

[103] Y. J. Lee and S. K. Lim, "Ultrahigh density logic designs using monolithic 3-D integration," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 32, no. 12, pp. 1892–1905, 2013.

[104] P. Chaourani, P.-E. Hellström, G. Malm, S. Rodriguez, and A. Rusu, "Towards Monolithic 3D Integration: A Design Flow," in *CDNLive, Cadence User Conf. EMEA*, 2016.

[105] P. Chaourani, P.-E. Hellström, R. Onet, S. Rodriguez, and A. Rusu, "Enabling Area Efficient RF ICs through Monolithic 3D Integration," in *Des. Autom. Test Eur. Conf. Exhib. - DATE*, 2017.

[106] M. G. Bardon, Y. Sherazi, P. Schuddinck, D. Jang, D. Yakimets *et al.*, "Extreme Scaling enabled by 5 Tracks Cells : Holistic design-device co-optimization for FinFETs and Lateral Nanowires," in *IEEE Int. Electron Devices Meet.*, 2016.

[107] J. Ryckaert, P. Raghavan, P. Schuddinck, H. B. Trong, A. Mallik *et al.*, "DTCO at N7 and beyond: patterning and electrical compromises and opportunities," in *Proc. SPIE, Des. Co-optimization Manuf. IX*, vol. 9427, 2015.

[108] C. G. Theodorou, E. G. Ioannidis, F. Andrieu, T. Poiroux, O. Faynot *et al.*, "Low-frequency noise sources in advanced UTBB FD-SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 4, pp. 1161–1167, 2014.

[109] C. Liu and S. Lim, "A design tradeoff study with monolithic 3D integration," *Qual. Electron. Des. (ISQED), Int. Symp.*, no. 404, 2012.

[110] C. Yan, S. Kontak, H. Wang, and E. Salman, "Open Source Cell Library Mono3D to Develop Large-Scale Monolithic 3D Integrated Circuits," in *IEEE Int. Symp. Circuits Syst.*, 2017, pp. 11–14.

[111] S. A. Panth, K. Samadi, Y. Du, and S. K. Lim, "Design and CAD methodologies for low power gate-level monolithic 3D ICs," *Proc. 2014 Int. Symp. Low power Electron. Des. - ISLPED '14*, vol. 1, pp. 171–176, 2014. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2627369.2627642

[112] B. Ku, P. Debacker, D. Milojevic, P. Raghavan, D. Verkest *et al.*, "Physical Design Solutions to Tackle FEOL/BEOL Degradation in Gate-level Monolithic 3D ICs," *Proc. Int. Symp. Low Power Electron. Des.*, vol. Part F1350, pp. 1–6, 2016.

[113] J. Franco, L. Witters, A. Vandooren, H. Arimura, S. Sioncke *et al.*, "Gate stack thermal stability and PBTI reliability challenges for 3D sequential integration: Demonstration of a suitable gate stack for top and bottom tier nMOS," in *IEEE Int. Reliab. Phys. Symp.*, 2017.

[114] A. Mallik, A. Vandooren, L. Witters, A. Walke, J. Franco *et al.*, "The impact of sequential-3D integration on semiconductor scaling roadmap," in *IEEE Int. Electron Devices Meet.*, 2017.

[115] A. Limaye and T. Adegbija, "A Workload Characterization for the Internet of Medical Things (IoMT)," *Proc. IEEE Comput. Soc. Annu. Symp. VLSI, ISVLSI*, vol. 2017-July, pp. 302–307, 2017.

[116] Z. Specification, "Ig ee pecification," 2012.

[117] A. Liscidini, A. Mazzanti, R. Tonietto, L. Vandi, P. Andreani *et al.*, "Single-Stage Low-Power Quadrature RF Receiver Front-End: The LMV Cell," *IEEE J. Solid-State Circuits*, vol. 41, no. 12, 2006.

[118] B. Razavi, *Behzad Razavi-RF Microelectronics, 2nd Edition(2011).pdf*. Prentice Hall, 2011.

[119] K. Benaissa, J. Y. Yang, D. Crenshaw, B. Williams, S. Sridhar *et al.*, "RF CMOS on high-resistivity substrates for system-on-chip applications," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 567–576, 2003.

[120] F. Gianesello, D. Gloria, C. Raynaud, S. Montusclat, S. Boret *et al.*, "On the design of high performance RF integrated inductors on high resistivity thin film 65 nm SOI CMOS technology," *2008 IEEE Top. Meet. Silicon Monolith. Integr. Circuits RF Systms - Dig. Pap. SiRF*, pp. 98–101, 2008.

[121] N. Li, K. Okada, T. Inoue, T. Hirano, Q. Bu *et al.*, "High-Q Inductors on Locally Semi-Insulated Si Substrate by Helium-3 Bombardment for RF CMOS Integrated Circuits," *IEEE Trans. Electron Devices*, vol. 62, no. 4, pp. 1269–1275, 2015.

[122] J. N. Burghartz and B. Rejaei, "On the design of RF spiral inductors on silicon," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 718–729, 2003.

[123] X. Sun, A. Rouhi Najaf Abadi, W. Guo, K. Ben Ali, M. Rack *et al.*, "Noise coupling between TSVs and active devices: Planar nMOSFETs vs. nFinFETs," *Proc. - Electron. Components Technol. Conf.*, vol. 2015-July, pp. 260–265, 2015.

[124] U. R. Tida, R. Yang, C. Zhuo, and Y. Shi, "On the Efficacy of Through-Silicon-Via Inductors," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 7, pp. 1322–1334, 2015.

[125] S. Mondal, S. B. Cho, and B. C. Kim, "Modeling and crosstalk evaluation of 3-D TSV-based inductor with ground TSV shielding," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 25, no. 1, pp. 308–318, 2017.

[126] P. Chaourani, S. Rodriguez, P. E. Hellstrom, and A. Rusu, "Inductors in a monolithic 3-D process: Performance analysis and design guidelines," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 27, no. 2, pp. 468–480, 2019.

[127] C. P. Yue and S. S. Wong, "On-Chip Spiral Inductors with Patterned Ground Shields for Si-Based RF IC's," *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 743–752, 1998.

[128] T. S. D. Cheung, J. R. Long, K. Vaed, R. Volant, a. Chinthakindi *et al.*, "Differentially-shielded monolithic inductors," *Proc. IEEE 2003 Cust. Integr. Circuits Conf. 2003.*, pp. 95–98, 2003.

[129] T. S. Cheung and J. R. Long, "Shielded Passive Devices for Silicon-Based Monolithic Microwave and Millimeter-Wave Integrated Circuits," *IEEE J. Solid-State Circuits*, vol. 41, no. 5, 2006.

[130] J. C. Rautio, J. D. Merrill, and M. J. Kobasa, "Efficient electromagnetic analysis of spiral inductor patterned ground shields," in *IEEE Int. Conf. Microwaves, Commun. Antennas Electron. Syst.*, 2013.

[131] C. Patrick Yue and S. Simon Wong, "Physical modeling of spiral inductors on silicon," *IEEE Trans. Electron Devices*, vol. 47, no. 3, pp. 560–568, 2000.

[132] H. Elwan, A. Tekin, and K. Pedrotti, "A differential-ramp based 65 dB-linear VGA technique in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2503–2514, 2009.

[133] S. Rodriguez, A. Rusu, L.-R. Zheng, and M. Ismail, "Digital calibration of gain and linearity in a CMOS RF mixer," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2008, pp. 1288–1291.

[134] P. Chaourani, D. Stathis, S. Rodriguez, P.-E. Hellstrom, and A. Rusu, "A Study on Monolithic 3-D RF/AMS ICs: Placing Digital Blocks Under Inductors," pp. 1–3, 2019.

[135] S. Yadav, K. H. Tan, A. Kumar, K. H. Goh, G. Liang *et al.*, "Monolithic Integration of InAs Quantum-Well n-MOSFETs and Ultrathin Body Ge p-MOSFETs on a Si Substrate," *IEEE Trans. Electron Devices*, vol. 64, no. 2, pp. 353–360, 2017.

[136] A. Vandooren, B. Parvais, L. Witters, A. Walke, A. Vais *et al.*, "3D technologies for analog/RF applications," *2017 IEEE SOI-3D-Subthreshold Microelectron. Unified Conf. S3S 2017*, vol. 2018-March, pp. 1–3, 2018.

[137] A. Vandooren, L. Witters, E. Vecchio, E. Kunnen, G. Hellings *et al.*, "Double-gate Si junction-less n-type transistor for high performance Cu-BEOL compatible applications using 3D sequential integration," in *IEEE SOI-3D-Subthreshold Microelectron. Unified Conf. S3S*, 2017.

[138] V. Deshpande, V. Djara, E. O'Connor, P. Hashemi, K. Balakrishnan *et al.*, "Advanced 3D Monolithic hybrid CMOS with Sub-50 nm gate inverters featuring replacement metal gate (RMG)-InGaAs nFETs on SiGe-OI Fin pFETs," *Tech. Dig. - Int. Electron Devices Meet. IEDM*, vol. 2016-Febru, no. 1, pp. 8.8.1–8.8.4, 2015.

[139] V. Deshpande, H. Hahn, E. O'Connor, Y. Baumgartner, D. Caimi *et al.*, "First Demonstration of 3D SRAM Through 3D Monolithic Integration of InGaAs n-FinFETs on FDSOI Si CMOS with Inter-layer Contacts," in *IEEE Symp. VLSI Technol.*, 2017, pp. 4503–4.

[140] A. Vandooren, J. Franco, B. Parvais, Z. Wu, L. Witters *et al.*, "3-D sequential stacked planar devices featuring low-temperature replacement metal gate junctionless top devices with improved reliability," in *IEEE Trans. Electron Devices*, vol. 65, no. 11, 2018, pp. 5165–5171.

[141] L. Jablonka, T. Kubart, D. Primetzhofer, A. Abedin, P.-E. Hellström *et al.*, "Formation of nickel germanides from Ni layers with thickness below 10 nm," *J. Vac. Sci. Technol. B*, vol. 35, no. 020602, 2017.

[142] L. Zurauskaite, L. Jones, V. Dhanak, I. Mitrovic, P.-E. Hellström *et al.*, "Investigation of Tm2O3 as a Gate Dielectric for Ge MOS Devices," *ECS Trans.*, vol. 86, no. 7, pp. 67–73, 2018.

[143] L. Pantisano, L. Trojman, J. Mitard, B. DeJaeger, S. Severi *et al.*, "Fundamentals and extraction of velocity saturation in sub-100nm (110)-Si and (100)-Ge," *Dig. Tech. Pap. - Symp. VLSI Technol.*, no. 110, pp. 52–53, 2008.

[144] B. P. Ginsburg and A. P. Chandrakasan, "500-MS/s 5-bit ADC in 65-nm CMOS with split capacitor array DAC," *IEEE J. Solid-State Circuits*, vol. 42, no. 4, pp. 739–747, 2007.

[145] M. Blagojevic, M. Cochet, B. Keller, P. Flatresse, A. Vladimirescu *et al.*, "A fast, flexible, positive and negative adaptive body-bias generator in 28nm FDSOI," *IEEE Symp. VLSI Circuits, Dig. Tech. Pap.*, 2016.

[146] D. Bol, M. Schramme, L. Moreau, T. Haine, P. Xu *et al.*, "A 40-to-80MHz Sub-4uW/MHz ULV Cortex-M0 MCU SoC in 28nm FDSOI with Dual-Loop Adaptive Back-Bias Generator for 20us Wake-Up From Deep Fully Retentive Sleep Mode," in *IEEE Int. Solid- State Circuits Conf.* -, 2019.

[147] S. Makovejev, S. Olsen, F. Andrieu, T. Poiroux, O. Faynot *et al.*, "On extraction of self-heating features in UTBB SOI MOSFETs," *Int. Conf. Ultim. Integr. Silicon, ULIS*, 2012.