Self-Heating Aware Design of ICs in Deep Sub-Micron FDSOI and Bulk Technologies

THÈSE N^{O} 8690 (2018)

PRÉSENTÉE LE 6 JUILLET 2018
À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DE SYSTÈMES MICROÉLECTRONIQUES
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Can BALTACI

acceptée sur proposition du jury:

Prof. A. P. Burg, président du jury
Prof. Y. Leblebici, Dr T. Toifl, directeurs de thèse
Prof. V. F. Pavlidis, rapporteur
Dr T. Ernst, rapporteur
Prof. C. Dehollain, rapporteuse

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

To my parents

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Prof. Yusuf Leblebici for giving me the opportunity to spend more than six years in an excellent academic atmosphere. His contributions to me are countless, not only as a supervisor but also as a person, he is an example in every respect.

I would also like to thank Dr. Thomas Toifl for sharing his valuable experience as my co-advisor and the thesis committee members Prof. Andreas P. Burg, Prof. Vasilis F. Pavlidis, Dr. Thomas Ernst and Prof. Catherine Dehollain for their contribution and insightful comments. I would like to thank Dr. Alain Vachoux, Dr. Alexandre Schmid, Tuğba Demirci and Slyvain Hauser for their technical support in LSM. I must also thank Samsung Flash Memory Team for the two-month I spent with them; it was a precious learning experience.

Thanks to Buğra, Kerem Kapucu and Kerem Seyid, I had amazing time outside work. The lunch breaks and the random events we had after work were priceless. I am also thankful to my friends Ahmet, Cenk, Çağlar, Erhan and Fatih. I would like to thank many Turkish people in Switzerland, Cem-Handan, Meriç-Nariye, Başak-İsmail, Ece and Dilan for their friendship.

I am thankful to my students Krishna Pradeep, Mustafa Beşirli, Nicolas Claude Frigerio and Cloé Sophie Liberté Roy for their contributions in my work. I also thank all the current and former LSM members: Mustafa, Selman, Şeniz, İrem, Cosimo, Jonathan, Navid, Elmira, Behnoush, Jury, Gain, Clemens, İlter, Alessandro, Kiarash, Armin, Vladan, Nicola, Radisav, Ömer, Gözen, Kadir, Duygu, Firat, Ayça for being the best colleagues.

Most importantly, I am what I am thanks to my parents Olcay and Eray. I am grateful to them for their endless support, patience and love.

Finally, I would like to thank Elsa. Without her, none of these would be possible.

Can Baltacı
Lausanne, May 2018

Abstract

Bulk CMOS technologies left the semiconductor market to the novel device geometries such as FDSOI and FinFET below 30 nm, mainly due to their insufficient electrical characteristics arising from different physical limitations. These innovative solutions enabled the ongoing device scaling to continue. However, the threshold voltage and the power supply values did not shrink with the device sizes, which caused an excessive amount of heat generation in very small dimensions. With the high thermal resistivity materials used in FDSOI and FinFET, the generated heat cannot leave the device easily, which is not the case in bulk. With all of these, modern geometries brought a major problem, which is the self-heating.

Due to self-heating effects (SHE), the temperature of a device rises significantly compared to its surroundings. Having very large local temperature brings important reliability issues. Moreover, the electrical behaviour of a device also changes dramatically when its temperature is very large. These facts bring the need of considering SHE and the temperature of each device separately. Nevertheless, in many of today's CAD tools, a single global temperature is applied to all of the devices. Even if some advanced simulation options are used, estimating the temperature of a device is not a simple task as it depends on many parameters.

The focus of this thesis is to show the significance of SHE in the design of ICs and provide self-heating aware design guidelines. In order to achieve this, different circuit implementations are studied by considering the SHE. The study consists of two main parts, which are the reliability of the high-speed digital circuits and the performance of analog blocks where noise is critical. Moreover, detailed device-level electro-thermal simulations are performed to explain the self-heating phenomena more in detail and to perform a comparison between bulk and FDSOI.

The digital part of the self-heating study is performed on two very high-speed full-custom 64-bit Kogge-Stone adders in 40 nm and 28 nm technologies. Thermal simulations are performed on these blocks to compare SHE in bulk and FDSOI geometries. The comparison of two implementations also provides the increasing significance of SHE with scaling. Extensive heating analyses are performed to find the most critical devices that are the primary heat generators. Design guidelines and solutions are proposed to flatten the temperature profiles in precharged and static logic implementations and to decrease the probability of electromigration.

The analog study of the work focuses on the thermal noise performance of LNAs and SHE on the flicker noise. Since thermal noise of a device linearly depends on the temperature, it is directly affected by SHE. To show the amount of SHE on the noise figure, three common gate cascode LNAs operating at 2 GHz with different device lengths are implemented in 28 nm FDSOI. The measurements show that the self-heating effects are clearly observed on the noise figure and the performance of the blocks deviate importantly from the simulations. Moreover, the self-heating effects are significantly more in short channel devices due to their large heat density. Similar experiments are also performed on different test structures in FDSOI at lower frequencies to observe SHE on flicker noise. The experiments show that flicker noise degrades at larger temperatures and more in short channel implementations.

Keywords: Self-Heating Effects, FDSOI, bulk, high speed digital, 64-bit adders, reliability, integrated inductors, low noise amplifiers, thermal noise, flicker noise

Résumé

Les technologies CMOS bulk ont laissé le marché des semi-conducteurs aux nouvelles géométries de dispositifs telles que FDSOI et FinFET en dessous de 30 nm, principalement en raison de leurs caractéristiques électriques insuffisantes résultant de différentes limitations physiques. Ces solutions innovantes ont permis la poursuite de la mise à l'échelle des dispositifs. Cependant, la tension de seuil et les valeurs d'alimentation n'ont pas rétréci avec la taille des dispositifs, ce qui a provoqué une génération excessive de chaleur dans de très petites dimensions. Avec les matériaux à haute résistivité thermique utilisés dans FDSOI et FinFET, la chaleur générée ne peut pas quitter le dispositif facilement, ce qui n'est pas le cas en bulk. A cause de tout cela, les géométries modernes ont apporté un problème majeur : l'auto-échauffement.

En raison des effets d'auto-échauffement, la température d'un dispositif augmente considérablement par rapport à son environnement. Avoir de très grandes températures locales apporte d'importants problèmes de fiabilité. De plus, le comportement électrique d'un dispositif change également de façon spectaculaire lorsque sa température est très élevée. Ces faits apportent le besoin de considérer les effets d'auto-échauffement et la température de chaque dispositif séparément. Néanmoins, dans la plupart des outils de CAO actuels, une seule température globale est appliquée à tous les dispositifs. Même si certaines options de simulation avancées sont utilisées, l'estimation de la température d'un dispositif n'est pas une tâche simple car elle dépend de nombreux paramètres.

L'objectif de cette thèse est de montrer l'importance des effets d'auto-échauffement dans la conception des circtuis intégrés et de fournir des directives de conception prenant en compte l'auto-échauffement. Pour ce faire, différentes implémentations de circuits sont étudiées en considérant les effets d'auto-échauffement. L'étude se compose de deux parties principales, à savoir la fiabilité des circuits numériques haute vitesse et la performance des blocs analogiques où le bruit est critique. De plus, des simulations électrothermiques détaillées au niveau du dispositif sont effectuées pour expliquer plus en détail les phénomènes d'auto-échauffement et pour effectuer une comparaison entre le FDSOI et le bulk.

La partie numérique de l'étude d'auto-échauffement est réalisée sur deux additionneurs

Kogge-Stone 64 bits à très haute vitesse, entièrement personnalisés, dans des technologies 40 nm et 28 nm. Des simulations thermiques sont réalisées sur ces blocs pour comparer les effets d'auto-échauffement dans les géométries bulk et FDSOI. La comparaison de deux implémentations fournit également l'importance croissante des effets d'auto-échauffement avec la mise à l'échelle. Des analyses d'échauffement approfondies sont effectuées pour trouver les dispositifs les plus critiques qui sont les principaux générateurs de chaleur. Des lignes directrices et des solutions de conception sont proposées pour aplatir les profils de température dans les implémentations logiques préchargées et statiques, et pour diminuer la probabilité d'électromigration.

L'étude analogique du travail se concentre sur la performance de bruit thermique des LNAs et des effets d'auto-échauffement sur le bruit de scintillation. Puisque le bruit thermique d'un dispositif dépend linéairement de la température, il est directement affecté par des effets d'auto-échauffement. Pour montrer la quantité des effets d'auto-échauffement sur la figure de bruit, trois LNAs de cascode à grille commune fonctionnant à 2 GHz avec des longueurs de dispositifs différentes sont implémentés en FDSOI 28 nm. Les mesures montrent que les effets d'auto-échauffement sont clairement observés sur le chiffre de bruit et que les performances des blocs s'écartent fortement des simulations. De plus, les effets d'auto-échauffement sont significativement plus importants dans les dispositifs à canaux courts en raison de leur grande densité de chaleur. Des expériences similaires sont également effectuées sur différentes structures de test en FDSOI à des fréquences plus basses pour observer les effets d'auto-échauffement sur le bruit de scintillation. Les expériences montrent que le bruit de scintillation se dégrade à des températures plus élevées et encore plus dans des implémentations à canaux courts.

Mots clés : Effets auto-échauffants, FDSOI, bulk, numérique haute vitesse, additionneurs 64 bits, fiabilité, inductances intégrées, amplificateurs à faible bruit, bruit thermique et bruit de scintillation

Contents

Acknowledgements **i**

Abstract (English/Français) **iii**

Table of Contents **ix**

List of Figures **xi**

List of Tables **xvii**

1 Introduction **1**
- 1.1 Operation at a Different Temperature 1
  - 1.1.1 Heating or Self-Heating? 2
  - 1.1.2 Observing Self-Heating Effects 3
  - 1.1.3 Why a Single Device? 5
  - 1.1.4 Self-Heating Related Issues 6
  - 1.1.5 Scope of This Work 7
- 1.2 State of the Art 7
- 1.3 Key Contributions of This Thesis 10
- 1.4 Thesis Outline 11

2 Device Level Analysis of Self-Heating **13**
- 2.1 Introduction 13
- 2.2 Theoretical Background and Related Work 14
  - 2.2.1 Block Level Thermal Simulations 14
  - 2.2.2 Device Level Thermal Simulations 15
- 2.3 Details of Our Device Level Simulations 17
- 2.4 Comparison of FDSOI and Bulk 19
  - 2.4.1 Physical Structure and Thermal Boundary Conditions 19
  - 2.4.2 Results and Discussions 21
- 2.5 Experiments on FDSOI 25
  - 2.5.1 Effect of Drain Voltage 25
  - 2.5.2 Effect of Gate Voltage 27
  - 2.5.3 Effect of Gate Length 29
  - 2.5.4 Effect of BOX Thickness 31

2.5.5 Effect of Drain and Source Contact Thermal Resistance 31
2.6 Conclusion 32

3 Thermal Aware Design of High Performance Digital Circuits 33
3.1 Introduction 33
3.2 Method for Block Level Thermal Simulations 35
3.2.1 Creation of Heat Maps 36
3.2.2 Thermal Model 37
3.2.3 Creation of Temperature Maps 39
3.3 Design of 64-bit Parallel Prefix Adders for Thermal Analysis 39
3.3.1 Architecture 40
3.3.2 Performance 45
3.4 Comparison of Bulk and FDSOI in 40 nm and 28 nm Nodes 47
3.4.1 Heating 47
3.4.2 Temperature 51
3.4.3 Simulation Time 64
3.5 Circuit Level Analysis of Hot-Spots 65
3.5.1 Heat Density of Individual Devices 65
3.5.2 Self-Heating of Devices with Different Functions 67
3.6 Conclusion 72

4 Self-Heating Effects on the Noise Performance of FDSOI MOSFETs 73
4.1 Introduction 73
4.2 Thermal Noise 74
4.2.1 Theoretical Background and Related Work 75
4.2.2 Thermal Noise Considering Self-Heating 76
4.3 Flicker Noise 82
4.3.1 Theoretical Background and Related Work 82
4.3.2 Flicker Noise Considering Self-Heating 84
4.4 Conclusion 85

5 Self-Heating Aware Design of LNAs with Short Channel Devices 87
5.1 Integrated Inductor Design 87
5.1.1 Undesired Effects 88
5.1.2 Lumped Model of an Integrated Inductor 89
5.1.3 Parametric Cell (pCell) for Integrated Inductors 94
5.2 Common Gate Cascode LNA 99
5.2.1 Design Parameters 99
5.2.2 Optimum Bias Point for Lowest Noise Figure 110
5.3 Measurements 115
5.3.1 Results 117
5.3.2 Comments 120
5.4 Conclusion 120

6 Flicker Noise Measurements 121
6.1 Test Blocks 121
6.1.1 Practical Considerations 123
6.1.2 Input Referred Noise 126
6.2 Measurement Results 128
6.2.1 Measurement Procedure 130
6.2.2 Results of Experiment-A 136
6.2.3 Results of Experiment-B 139
6.3 Conclusion 142

7 Conclusion 143

Bibliography 156

Curriculum Vitae

List of Figures

1.1 Independent of the activity of any other devices and the ambient temperature, the characteristics of M0 changes while passing from the state shown on (a) to the one shown on (b) due to its self-heating. The influence of self-heating effects depends on the temperature difference $T_1 - T_0$. 3
1.2 Performance and reliability issues created by self-heating effects. Black lines encircle the items that are the focus of this work. 6
1.3 Block level thermal simulations performed on different processors in SOI technologies. 8
1.4 Device level electro-thermal simulations performed in different technology nodes and geometries. 9

2.1 Doping profiles of (a) the FDSOI and (b) the bulk device structures. The positive doping values are used to denote n-type doping, and negative values for the p-type doping. 20
2.2 Comparison of simulated 40 nm bulk and 28 nm FDSOI MOSFETs and the design kit models in terms of their I_D-V_G characteristics. 21
2.3 Spatial heat generation profile of 40 nm (a) bulk and (b) FDSOI devices. 22
2.4 Spatial temperature profile of 40 nm (a) bulk and (b) FDSOI devices. 22
2.5 Heat generation and temperature profiles of 40 nm (a) bulk and (b) FDSOI at y = -1 nm. 22
2.6 Lateral electric field in 40 nm (a) bulk and (b) FDSOI MOSFETs at y = -1 nm. . . 22
2.7 Spatial heat generation profile of 28 nm (a) bulk and (b) FDSOI devices. 23
2.8 Spatial temperature profile of 28 nm (a) bulk and (b) FDSOI devices. 23
2.9 Heat generation and temperature profiles of 28 nm (a) bulk and (b) FDSOI at y = -1 nm. 23
2.10 Lateral electric field in 28 nm (a) bulk and (b) FDSOI MOSFETs at y = -1 nm. . . 23
2.11 (a) heat generation, (b) lateral component of the electric field, (c) temperature and (d) conduction band energy inside the channel for different values of V_D while V_G is constant at 0.6 V. 26
2.12 Simulation results for V_D swept from 0 V to 1 V while V_G = 0.6 V. 26
2.13 (a) heat generation and (b) temperature inside the channel for different values of V_G while V_D is constant at 0.5 V. 28
2.14 Simulation results for V_G swept from 0 V to 1 V while V_D = 0.5 V. 28

2.15 Spatial distribution of heat generation for different V_G values from 0.2 V to 1.5 V while V_D is kept constant at 0.5 V. 29
2.16 (a) Maximum lateral electric field, (b) maximum heat generation, (c) maximum temperature and (d) full bias drain current with respect to gate length for the 28 nm FDSOI technology. 30
2.17 (a) Heat generation and (b) temperature inside the channel for different the gate length values in 28 nm FDSOI technology. 30
2.18 (a) Temperature and (b) heat flow with respect to BOX thickness. 31
2.19 (a) Heat generation and (b) temperature inside the channel for different thermal boundary conditions for drain and source. 32

3.1 Thermal analysis flow and data exchange between the used tools. 35
3.2 Summation of the independent heat sources located in the same square. 37
3.3 Thermal models of FDSOI (left) and bulk (right). 38
3.4 The architectural block diagram of the implemented 64-bit Kogge-Stone parallel prefix adders with the radix-4 and sparsity-4 options. The input/output nodes, signal names and the hardware blocks on the critical path are indicated with red colour. 40
3.5 Schematics of N-domino logic PG-Generator gates in (a) 40 nm and (b) 28 nm. The devices at the bottom are the clocked footing devices [1]. 41
3.6 Schematics of NP-domino logic P-Merge gates used in (a) 40 nm and (b) 28 nm. 43
3.7 Schematics of NP-domino logic G-Merge gates used in (a) 40 nm and (b) 28 nm. 43
3.8 The number and type of the devices on the critical path of (a) the 40 nm and (b) the 28 nm adder. 43
3.9 The block diagram of the 4-bit CSA. The $S_1(i)$ and $S_0(i)$ signals at the outputs of the two 4-bit ripple carry adders are the sum signals which are generated for the two cases where C_i is equal to 1 and 0 respectively. 45
3.10 The layouts and floorplans of the implemented 64-bit adders in (a) 40 nm and (b) 28 nm. Thermal simulations are performed on these areas. 45
3.11 (a) The die micrograph, (b) zoomed micrograph and (c) the test setup of the implemented 64-bit adder in 40 nm. Thermal simulations are performed on the 64-bit adder core block. 46
3.12 Thermal images of the 40 nm adder taken during tests. (a) the visible image of the board where the thermal images are taken, (b) thermal image just before the addition begins, (c) thermal image just after the addition begins and (d) 20 seconds after the addition begins. 47
3.13 Heat maps of the 40 nm adder for different pitch values. Heat density unit is W/cm^2. 48
3.14 Heat maps of the 28 nm adder for different pitch values. Heat density unit is W/cm^2. 49

3.15 Maximum and minimum heat density curves for heat maps with different resolutions. Maximum heat density saturates as the pitch value becomes comparable to the minimum length of the specified technology, whereas the minimum heat density drops to 0 W/cm^2 as resolution increases. 50
3.16 Temperature maps of the 40 nm adder in bulk geometry for different pitch values. 52
3.17 Temperature maps of the 40 nm adder in FDSOI geometry for different pitch values. 53
3.18 Temperature maps of the 28 nm adder in bulk geometry for different pitch values. 54
3.19 Temperature maps of the 28 nm adder in FDSOI geometry for different pitch values. 55
3.20 Maximum and minimum temperature values for all temperature maps with heating inputs of different resolutions. 56
3.21 Heat map of different blocks in 40 nm with their individual heating (x and y in μm). 57
3.22 Heat maps of (a) 40 nm and (b) 28 nm technologies focused on the hot-spots. . 60
3.23 Temperature maps of (a) 40 nm and (b) 28 nm FDSOI technologies focused on the hot-spots. 60
3.24 Temperature maps of (a) 40 nm and (b) 28 nm bulk technologies focused on the hot-spots. 60
3.25 Heat maps of (a) 40 nm and (b) 28 nm for observing extreme case temperature profiles. 63
3.26 Temperature profiles of (a) 40 nm and (b) 28 nm due to the heating inputs of Figure 3.25. 63
3.27 Temperature profiles along the cut-lines (Figure 3.26) taken at the centre of the devices. The red thick lines show the location of the devices on Figure 3.26. . . 63
3.28 Simulation time of bulk and FDSOI with respect to pitch. In this analysis the resolution of the heating input is swept while the resolution of the thermal simulation is fixed at 2048 × 2048. 64
3.29 Heat density distribution of the transistors in the 64-bit adder circuits. 66
3.30 The heat density distribution after increasing the width of the high heat density devices by a factor of 2 and 3. 66
3.31 Standard cell of a *Generate-Merge* gate with different size pre-charge devices. . 67
3.32 The schematic that illustrates the devices belonging to different groups. The logic gate in the centre corresponds to the domino logic implementation of Generate-Merge function with N-type logic tree. 68
3.33 Distribution of heat density for different group of devices in the 64-bit adders implemented in (a) 40 nm and (b) 28 nm. 69
3.34 Time domain power dissipation density waveforms of devices from different groups. 70
3.35 The heat density distribution after increasing the width of the PI and PO devices by a factor of 2 and 3. 71

4.1 The thermal model that provides the temperature rise of each device due to their self-heating [2]. 74
4.2 Thermal noise current and temperature of 30 nm, 60 nm and 300 nm nMOS devices as V_{DS} is changed. The continuous curves correspond to the case where self-heating effects are considered and the dashed curves correspond to the case where they are ignored. 77
4.3 Error in thermal noise current (a) with respect to V_{DS} for devices with different gate lengths from 30 nm to 300 nm and (b) with respect to gate length for V_{DS} = 0.2 V, 0.4 V, 0.6 V, 0.8 V, 1.0 V, 1.2 V. For V_{DS} = 0 V, there is no error. 78
4.4 Model of the noisy device with a fictitious input referred noise voltage source $\overline{V_{n,i}^2}$ 79
4.5 (a) drain current, (b) transconductance, (c) temperature rise due to self-heating, (d) the calculation error in drain current and transconductance when self-heating effects are not considered. 80
4.6 (a) input referred thermal noise voltage and (b) its calculation error when self-heating effects are ignored. The devices are biased with V_G = 0.6 V. 81
4.7 (a) Flicker noise voltage normalized with respect to effective device area and (b) further normalized with respect to overdrive voltage. 83
4.8 (a) flicker noise voltage at the gate at 20 MHz according to (4.8) and (b) its calculation error when self-heating effects are ignored. The devices are biased with V_G = 0.6 V. 85

5.1 Loss mechanisms of an integrated inductor. 88
5.2 Simple lumped models of an integrated inductor for different cases: (a) ideal, (b) first order lumped model where only the winding series resistance is taken into account, (c) more detailed model where other high frequency losses are considered. 90
5.3 Parallel *RLC* representations of the models on (a) Figure 5.2b and (b) Figure 5.2c 91
5.4 Behaviours of $|Z_{L1}|$, R_P and Q_S with respect to the inductor series resistance R_S for f_0 = 10 GHz, L = 5 nH and C = 50.7 fF. 92
5.5 Behaviours of $|Z_L|$, R_{LP} and Q_L with respect to parallel high frequency resistance R_{HF} for f_0 = 10 GHz, L = 5 nH, R_S = 10 Ω and C = 50.7 fF. 93
5.6 Layout screen-shots and cross-section views of the (a) two-stacked and (b) symmetric inductor pCells synthesized in 28 nm FDSOI technology. 95
5.7 Comparison of the behaviours of the RF simulation results and lumped model of (a) a two-stack and (b) a symmetric integrated inductor. 97
5.8 Example of the integrated stacked inductor drawing and its adjustable parameters. 98
5.9 Behaviours of the inductor parameters as (a) the shield pitch of the inductor was changed while shield-width/shield-pitch ratio was kept constant at $\mathrm{W_{sh}/P_{sh}}$ = 0.6, (b) shield width was changed while shield pitch was kept constant at $\mathrm{P_{sh}} = 5\mu\mathrm{m}$. 99
5.10 Schematic of the common gate cascode LNA 100

5.11 Small-signal AC model used for finding the impedance seen from the source of a MOSFET while the drain is connected to an arbitrary impedance. 100
5.12 Total input impedance (dashed lines) and equivalent parallel resistance (continuous lines) of three different LNAs with different gate lengths 30 nm, 45 nm and 180 nm (a) while looking into the source of M1 (b) seen from the PCB. 104
5.13 Tuning the input impedance by considering the off-chip parasitics: wirebonding inductance, L_{wb} and parasitic capacitance of the PCB, C_{PCB}. 104
5.14 Small-signal AC model used for finding the impedance seen from the drain of a MOSFET while the source is connected to an arbitrary impedance. 105
5.15 Small-signal AC model of the common gate cascade LNA at f_0. 107
5.16 Small-signal AC model of the common gate cascade LNA at f_0 with the independent noise sources. 108
5.17 Noise figure under different V_{B2} levels when the supply voltage is set to 3 V for three different the common gate cascode LNA implementations with gate lengths of (a) 30 nm, (a) 45 nm, (a) 180 nm. Continuous lines correspond to the case where self-heating effects are included and dashed lines are for the case where self-heating effects are excluded. (d) the calculation error when self-heating effects are not included. 112
5.18 (a) temperature rise of M1 (continuous) and M2 (dashed) as V_{B2} changes, (b) noise figure vs temperature rise of M1. 113
5.19 ΔT profile of the common gate cascode LNA with $L = 30$ nm for the bulk thermal properties under different bias conditions provided by V_{B2} while $V_{DD} = 3$ V. The observable devices are M2 and M1 from left to right. The spatial unit is μm. . . 114
5.20 ΔT profile of the common gate cascode LNA with $L = 30$ nm for the FDSOI thermal properties under different bias conditions provided by V_{B2} while $V_{DD} = 3$ V. The observable devices are M2 and M1 from left to right. The spatial unit is μm. 115
5.21 (a) micrograph of the taped-out die where the three LNA implementations are indicated with the channel length values, (b) RF test board, (c) bonded die and (d) the measurement set-up. 116
5.22 Screen-shot of the N8975A screen, where the noise figure and gain of the 45 nm LNA can be seen for the 1 GHz - 3 GHz band. 117
5.23 Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 180$ nm. 118
5.24 Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 45$ nm. 118
5.25 Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 30$ nm. 119

6.1 Main parts of the two test structures for observing the flicker noise of a single MOSFET (M1). (a) experiment-A: common source and (b) experiment-B: common source with a cascode device. 122

6.2 Simple schematic representation of the complete test system for the designs of Figure 6.1. 124
6.3 (a) simple schematic of source follower and (b) its small-signal AC model. . . . 125
6.4 Small signal AC models of the complete system of experiment-A. 127
6.5 Small signal AC models of the complete system of experiment-B. 127
6.6 (a) Micrograph of the taped-out die, (b) bonded die: gold wire-bonding is used for better electrical conductivity, (c) RF test board and (d) the measurement set-up. Many power supplies and multi-meters were necessary to provide the bias and supply voltages (V_I, V_O, V_B, $V_{DD,A}$, $V_{DD,B}$, $V_{DD,S}$) and measure different currents for each data point. 129
6.7 Configurations of FSP for different steps of the gain measurement. Bias-tees are indicated with BT. 131
6.8 Configurations of FSP for different steps of the noise measurement. 132
6.9 Noise power spectral densities at different stages of the set-up for obtaining the flicker noise voltage of M1 from the measured noise power. 133
6.10 (a) external SMA 50 Ω feed-through termination, (b) configuration of FSP for tuning the output resistance to 50 Ω and (c) measured gain and output impedance values. 135
6.11 Drain current of M1. 137
6.12 Power dissipation, P, of M1 ($V_{DS,M1} \times I_{D,M1}$). 137
6.13 Gain, A_p, of M1 at 20 MHz. 137
6.14 Output noise power spectral density, $S_{Pn,o}$, of M1 at 20 MHz. 137
6.15 Temperature rise, ΔT, of M1 due to its self-heating (simulation). 138
6.16 Percentage of the flicker noise of M1 in the total noise at 20 MHz (simulation). . 138
6.17 Input referred noise power spectral density, $S_{Pn,i}$, of the entire design at 20 MHz. 138
6.18 The power spectral density of flicker noise voltage, S_{Vnf1}, of M1 at 20 MHz. . . 138
6.19 Normalized flicker noise voltage power spectral densities of the 100 nm length and the 60 nm length devices for high heating case. The values of the horizontal axis are shifted by the smallest V_{DS} value in the measurement set. 139
6.20 Drain current of M1. 140
6.21 Power dissipation, P, of M1 ($V_{DS,M1} \times I_{D,M1}$). 140
6.22 Gain, A_p, of M1 at 20 MHz. 140
6.23 Output noise power spectral density, $S_{Pn,o}$, of M1 at 20 MHz. 140
6.24 Temperature rise, ΔT, of M1 due to its self-heating (simulation). 141
6.25 Percentage of the flicker noise of M1 in the total noise at 20 MHz (simulation). . 141
6.26 Input referred noise power spectral density, $S_{Pn,i}$, of the entire design at 20 MHz. 141
6.27 The power spectral density of flicker noise voltage, S_{Vnf1}, of M1 at 20 MHz. . . 141
6.28 Normalized flicker noise voltage power spectral densities of the 100 nm, 60 nm and 30 nm length devices for high heating case. The values of the horizontal axis are shifted by the smallest V_B value in the measurement set. 142

List of Tables

2.1 Duality between electrical and thermal circuits at steady state 15
2.2 Parameters of the bulk and the FDSOI devices of Figure 2.1 20

3.1 Thermal properties of FDSOI and bulk in 40 nm and 28 nm technologies. 38
3.2 Comparison of the implemented adders in 40 nm and 28 nm technologies. . . . 45
3.3 Comparison of the published 64-bit adders. †The adder in [3] is a 32-bit implementation. 46
3.4 Parameters of the devices on Figure 3.25 62
3.5 Maximum and mean heat density values and the number of devices from different device groups in 40 nm. 69
3.6 Maximum and mean heat density values and the number of devices from different device groups in 28 nm. 70
3.7 Change of thermal parameters in scaling from 40 nm to 28 nm. 72

4.1 Parameters for the devices with short, medium and long channels. 77

5.1 Adjustable parameters and their explanations of the designed inductor pCells that are shown on Figure 5.8. Important note: all dimensions are shrinked by 0.9 in fabrication. 96
5.2 Extracted lumped circuit parameters of the two-stack and symmetric integrated inductors. 96
5.3 Input resistance values of LNAs with different gate lengths and their contributors from different parts of the circuit at f_0. 104
5.4 Design parameters of three different common gate cascode LNAs implemented in 28 nm FDSOI technology. 111
5.5 The measured I_{REF} and the applied V_{B1} values for each experiment. The units of I_{REF} and V_{B1} are μA and mV respectively. 117

6.1 Properties and bias voltages of different samples used in experiment-A and experiment-B. 129

1 Introduction

Starting from the early 1960s, the integrated circuit technology has maintained a continuous advancement throughout its 60-year history. The number of devices per unit area in an IC was doubled every two years as predicted by G. E. Moore [4] primarily thanks to *device scaling*. With device scaling, the area of a single MOSFET was roughly halved from one technology node to the next one. Having more and smaller devices in integrated circuits enabled the designers to implement more complex systems with higher throughput, thanks to the increased clock frequency. As from mid-1990s, traditional device scaling started to encounter physical limitations such as velocity saturation, mobility degradation, sub-threshold and gate leakage, punch through, increased parasitics and so on. Innovative solutions like pocket and halo implantation, shallow trench isolation (STI), usage of high-k materials as gate dielectric enabled the ongoing device scaling to continue. However, the power density started to increase from one technology to another by breaking the Dennard scaling, which states that as the transistors are reduced in size, their power density stays constant [5]. Approaching to the late 2000s, bulk CMOS technology reached its limits and alternative device geometries such as Fully Depleted Silicon on Insulator (FDSOI) and FinFET devices took its place. The new device geometries improved the electrical characteristics of the devices; however, they brought the necessity to use thermally poor materials around them. While enabling the device scaling to go on, all of these novelties brought another formidable problem of excessive heating and operation at elevated temperatures.

1.1 Operation at a Different Temperature

Many of the physical parameters of solid-state devices depend on temperature. In general, increasing the temperature changes these parameters in the undesired direction. The operation of semiconductors materials and solid-state devices at different temperatures have been investigated for a long time. One can find well established and experimentally proven accurate models in the literature [6, 7, 8, 9][1]. To observe a change in the temperature of a solid-state

---

[1] In this report, we do not provide a detailed explanation of the physics of temperature dependent properties unless it is necessary to clarify a point, since it is not the main scope of this work.

device (a system in general), it has to interact with heat. Generally speaking, the temperature of an enclosed system increases only in two separate cases or with the combination of two. First, the temperature of a system increases if there is a positive heat flow into this system and no heat sources or absorbers are located in the same enclosed volume. In case this is an electronic system such as an IC, an example of this case can be placing it inside a heater without connecting any power supplies. Second, the temperature of a system also increases if a heat source that is located in this system generates heat that is greater than zero and no heat passes through the boundaries of this system. Returning back to the same IC, connecting it to a power supply and placing it to a thermal insulator test environment can be an example for this case. While the first case is not mandatory in many cases, the second one is an inevitable result of the desired operation of an IC. In this work, we skip the first case and focus on the second one, where the elevated temperature of an electronic system is due to the heating resulting from its operation.

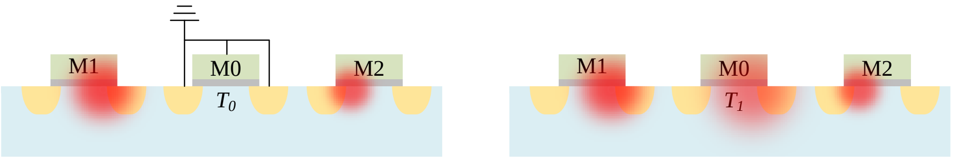
1.1.1 Heating or Self-Heating?

Heating[2] in electronic systems is the process where the kinetic energy of charge carriers is transferred to the body of the conductor through different scattering mechanisms. An IC can be composed of millions of transistors. According to the processed data, each transistor might conduct different amount of current where some of them might not be conducting any current at all. Depending on their activity, each conducting transistor contribute differently to the overall heating of the IC. Because of the overall heating of all transistors, the temperature values at different points of the IC would take values that are larger than the state where the IC is not connected to a power supply. Due to the temperature dependence of transistors' different parameters, their electrical characteristics are directly or indirectly affected depending on their temperature. In advanced technologies, the temperature at different points in the ICs might reach to extremely large values (more than 100 C), which might bring significant reliability and performance issues. This brings the problem of heating in IC design where the source is the combination of *all* constitutive transistors.

When we focus on a *single* transistor in an IC, its temperature would take a value depending on the total heating of the other transistors and the heat generated by itself. Under the same operating conditions where all the other devices generate the same amount of heat, the temperature of a transistor in its conducting state would be greater than its non-conducting state. The temperature difference between these two cases results from the **Self-Heating** of this transistor (Figure 1.1). Self-heating would bring an additional effect on the operation of a transistor on top of the heating effects of the other devices. Formally, we define **Self-Heating Effects (SHE)** as the combination of all the modifications in the physical properties of a transistor due to the elevation of its temperature arising from the heating that is only originating from its own electrical activity. In this work, our goal is to focus on the self-heating rather than the total heating of ICs or operation of ICs at high temperature environments.

---

[2] Also called **Joule heating**, **Ohmic heating** or **resistive heating**



(a) M0 is not generating any heat.

(b) M0 is generating heat.

Figure 1.1 – Independent of the activity of any other devices and the ambient temperature, the characteristics of M0 changes while passing from the state shown on (a) to the one shown on (b) due to its self-heating. The influence of self-heating effects depends on the temperature difference $T_1 - T_0$.

1.1.2 Observing Self-Heating Effects

The definition of self-heating effects contains two states where the transistor is electrically inactive and active. To observe the amount of self-heating effects, one should perform the necessary measurements on a transistor in these two states and report the values of each parameter for both cases. The difference of a parameter in two states would give information about how much that parameter is affected from self-heating effects. More formally, self-heating effects bring an additional shift in a parameter with a coefficient of

$$k_p = \frac{p|_{SH} - p|_{NO}}{p|_{NO}} \tag{1.1}$$

where $p|_{NO}$ is the value of a parameter with *no* self-heating effects and $p|_{SH}$ is its value when self-heating effects are present[3]. To illustrate the situation, let us assume we are interested in the self-heating effects on the transconductance of a transistor under a certain gate and drain bias. In this case, p in (1.1) is replaced by the transconductance, g_m. Hence, the value of transconductance under self-heating effects is

$$g_m|_{SH} = (1 + k_{g_m}) g_m|_{NO}. \tag{1.2}$$

From a temperature point of view, $g_m|_{NO}$ is the transconductance value at T_0, $g_m|_{SH}$ is the transconductance value at T_1 and $T_1 - T_0$ is the temperature rise due to self-heating. $g_m|_{SH}$ can be simply observed by applying the desired bias levels with power supplies and performing the related measurements. When it comes to measure $g_m|_{NO}$, we cannot simply *switch* the self-heating effects *off* and repeat the same experiment while the desired voltage levels are still being applied by the external sources. The self-heating is naturally there and it cannot be removed without removing the external power. However, by switching off the power supplies, we cannot bring the device to necessary bias conditions in which we are interested. Nevertheless, the temperature of the device with no self-heating, T_0, can be measured simply by applying no power. Therefore, the temperature of the transistor can be externally pushed to T_0 during its desired operation via heat sinks. In this case, the measured transconductance

---

[3] In the following chapters, k_p in (1.1) is also referred as the **Error** for quantifying the miscalculation when self-heating effects are not considered.

would give the value of the sought $g_m\big|_{NO}$.

The amount of self-heating effects can be theoretically reported by applying the previously explained methods. Nevertheless, there are additional practical issues in measuring the temperature. In commercially available advanced technologies, the channel length of a transistor can be as small as 10 nm by 2018. Moreover, the transistors are buried under many interconnect and dielectric layers. These facts create issues for performing an accurate temperature measurement of a transistor. By removing the back end of line (BEOL) and some of the front end of line (FEOL) layers, the transistor can be revealed. However, this modification would change the temperature of the transistor due to the modified thermal geometry by removing different layers. Even if this does not bring a significant change, one should make sure that the measurement technique does not disturb the temperature of a nanometer scale region. In addition to these, it is reported in many sources and shown in our work that the temperature of a transistor shows very large variations in its volume. For example, the temperature difference between the drain and the source of a device might be some tens of Kelvins [10, 11]. This means that the temperature in a short channel device might change by more than 5 K in one nanometer distance. Consequently, measuring a single temperature value and assuming that it is the temperature of the entire transistor is not sufficient. Considering these, bringing the device temperature down to T_0 would also be practically impossible. Even if the average temperature of a device can be set to T_0 with a cooling system, there would still be a non-uniform temperature profile in the device due to its non-uniform heating and different heat diffusion paths.

Because of the mentioned difficulties and mainly because the self-heating cannot be practically removed, measuring the self-heating effects is a complicated task. Yet, there has been different studies on observing self-heating effects experimentally. One of the most common methods is the pulsed I-V technique [12]. The idea is to apply a sharp voltage or current pulse and to perform a measurement immediately after. This way, the temperature shift due to self-heating can be prevented to an extent. This depends on how fast the measurement setup can obtain a proper data point and the thermal time constant of the measured device. Different authors reported two distinct behaviours between the pulsed and the DC measurements [13, 14, 15, 16, 17] for SOI and FinFET devices. However, the amount of self-heating is still questionable at the instant where the first measurement data is sampled. Especially in nanometer scale devices, the hot-spot is extremely small. This decreases the thermal time constant significantly and the level of self-heating at the measurement moment becomes more questionable.

There has been also many attempts to gather information about the self-heating effects by performing device level simulations [10, 11, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31]. The authors tried to estimate the temperature rise due to self-heating effects. Different sources reported different temperature values depending on the used techniques. The advantage of simulations is to have the chance to create the scenario where self-heating effects are disabled. However, one needs to consider detailed mathematical models for all physical phenomena which are actively taking part under the applied conditions (size of the geometry, ambient temperature, boundary conditions, applied voltage levels, material type etc.). As the dimensions get smaller (less than the mean free paths of quantized energy particles such as phonons and electrons or comparable with the lattice constant of the conducting medium), it becomes necessary to consider the quantum effects, which contain very complex mathematical models. For this reason, very detailed simulations might take very long times to converge and the results might still be questionable. Today, there is still an ongoing research on performing more accurate electro-thermal simulations.

1.1.3 Why a Single Device?

One might naturally wonder if the contribution of a single device out of millions is important to investigate or how much the temperature of an IC increases by only one device. It is true that turning a single device on and off would not have a big impact on the overall system if the entire chip is considered. However, the focus in the self-heating study is not the influence of a single device on the entire chip. The critical question is how much the operation of a device is influenced due to its own heating, excluding the heating of the other transistors.

The self-heating related problems became more an issue especially after the introduction of the modern MOSFET device geometries like FDSOI and FinFET, and new dielectric materials [10]. Previously, it was reported that the peak temperature of the FDSOI devices is located close to the drain end of the device [19, 24] and the maximum temperature value in FDSOI FETs is found to be much higher than the one in the conventional bulk MOSFETs [32]. The higher maximum temperature of the FDSOI and FinFET structures is mainly due to the thermal behaviour of its constitutive materials. SiO_2 is one of these materials which is used on the sides of the channel for isolation [26, 33, 34]. The thermal conductivity of the SiO_2 isolation layer is two orders of magnitude lower than the thermal conductivity of the bulk Si. Moreover, in these structures, the channel is no longer a part of the bulk; it is rather a thin layer of Si. The thermal conductivity of the Si thin film, where the devices are generating heat, is one order of magnitude less than the thermal conductivity of bulk Si [24]. Additionally, the boundary between Si and SiO_2 creates a temperature jump, hence a finite interface thermal resistance, [35, 36], which is equal to the thermal resistance of a SiO_2 layer with a thickness of 20 nm [18]. Due to the mentioned facts, the dissipated power in FDSOI devices does not find a high conductance diffusion path. Because of this, in FDSOI and FinFET technologies, the generated heat turns into temperature in nanometer scale local spots with much smaller dimensions than the ones in bulk technologies. For this reason, contrary to the traditional bulk CMOS technologies where temperature is assumed a global variable, it is necessary to consider the self-heating effects and the temperatures of each device separately to obtain reliable results and better yield.

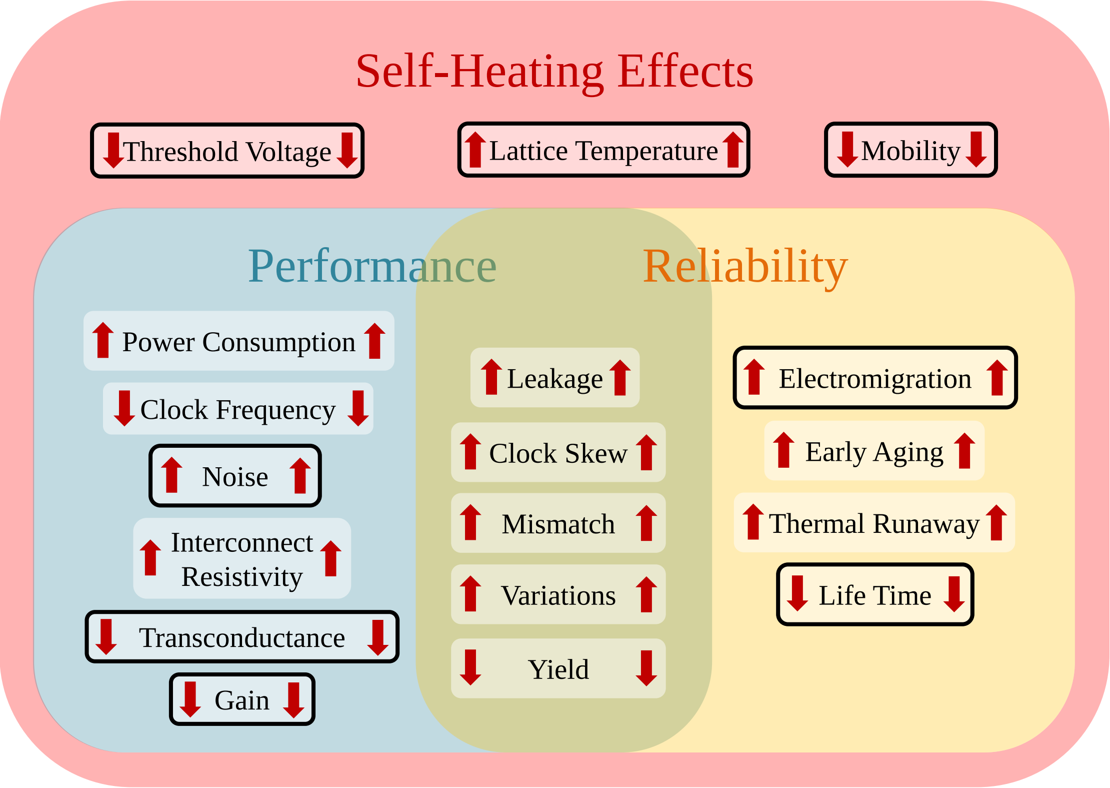


Figure 1.2 – Performance and reliability issues created by self-heating effects. Black lines encircle the items that are the focus of this work.

1.1.4 Self-Heating Related Issues

As it has already been explained, the primary outcome of the self-heating is the increased temperature of the device. Two most well known device parameters that are directly affected by the increased temperature are the mobility and the threshold voltage. Mobility decreases at elevated temperatures mainly due to increased phonon scattering rate. Threshold voltage also decreases with temperature due to the temperature dependence of the flatband voltage [37]. These modifications in the main device parameters create threats in the performance and the reliability of ICs (Figure 1.2).

Due to the decreased mobility, the speed of the devices, consequently the maximum clock frequency decreases. Lower threshold voltage increases the sub-threshold current significantly due to its negative exponential dependence on the threshold voltage [38]. Larger sub-threshold current means leakage. Due to increased leakage, the power consumption also increases [39, 40]. Higher power consumption brings higher temperature and this might result in thermal runaway where the die fails due to the uncontrolled increase in the temperature. Although thermal runaway does not happen, the chip might settle down to a higher temperature, which would degrade the performance as well as the reliability of the chip [41, 42]. Electromigration phenomena is another reliability problem related to temperature where the metal interconnects are broken due to diffusion or flow of atoms under very high current densities at high temperatures [43, 44]. The difference between the heating values of devices might create an uneven temperature profile. This results in different behaviours of the devices with the same properties and temperature dependent variations throughout the die. If these temperature variations are observed on some devices where matching is important, mismatch issues like larger offset voltage might occur. Nonhomogeneous temperature distribution also results in different conductance values of the interconnect lines due to the temperature dependence of the metal resistivity [45, 46, 47]. Consequently, the delay times of balanced interconnects might differ, which might cause serious clock skew issues [48, 49]. Larger average electron kinetic energy increases the noise, which results in lower SNR and dynamic range. All of the mentioned problems show that having a reliable and high performance chip is not possible without considering the self-heating effects and the thermal behaviour of a design.

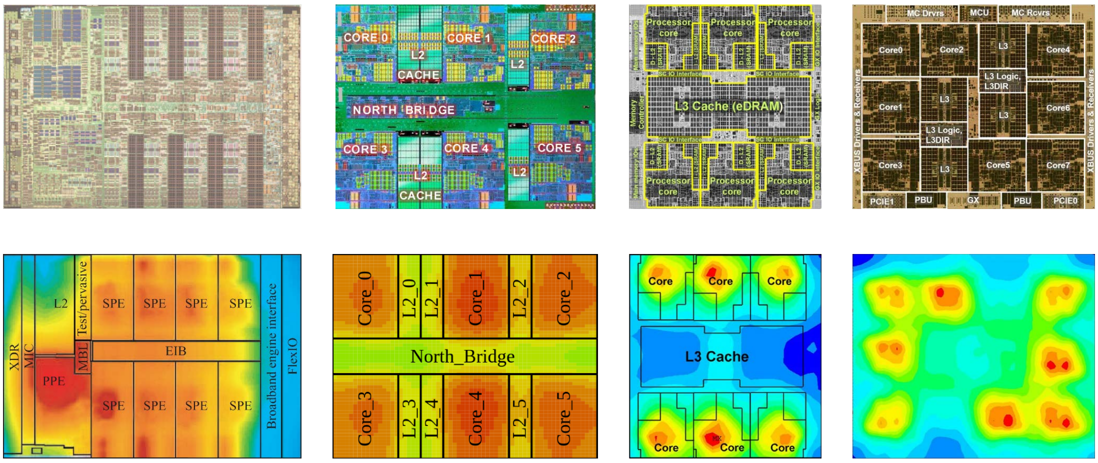
1.1.5 Scope of This Work

The primary target of this work is to show the importance of considering self-heating effects of individual MOSFETs during the design cycle in order to get maximum performance and reliability. To demonstrate this, we analyze the self-heating effects on digital and analog blocks. In the digital side of the study, we investigate the reliability of high performance arithmetic blocks by performing their detailed thermal analysis. In the analog part of the study, we investigate the noise performance of devices and low noise amplifiers (LNAs). To understand the self-heating effects in more detail for different cases like technology (bulk, FDSOI), thermal boundary conditions, channel length etc., we also take advantage of device level electro-thermal simulations. Finally, we demonstrate the findings of our work by electrical measurement results.

While performing our device and block level thermal simulations, we mainly focus on the self-heating effects of MOSFET devices rather than other dissipative elements inside a chip such as interconnects and monolithic resistors. The heating of interconnects has already been widely studied and different publications can be found on thermal modelling, simulations and the effect of technology scaling [46, 48, 50, 51, 52, 53, 54, 55]. Different solutions are proposed mainly to increase the electromigration reliability and decrease the temperature dependent interconnect delay uncertainty [56, 57, 58, 59]. In our work, we exclude the heating effects of different stacks of interconnects higher than the contact and first metal layer, and assume that the upper part of the chip is thermally isolated, in order to observe the heating of only MOSFET devices and simplify our analysis, unless otherwise is stated.

1.2 State of the Art

Thermal simulations have been performed in different scales from chip and block level down to device level. Chip level thermal simulations are usually necessary during the design of high performance multi-core processors that are composed of billions of transistors. In these



(a) Cell/B.E. processor in 90 nm SOI [60]

(b) AMD Opteron 6172 processor in 45 nm SOI [61]

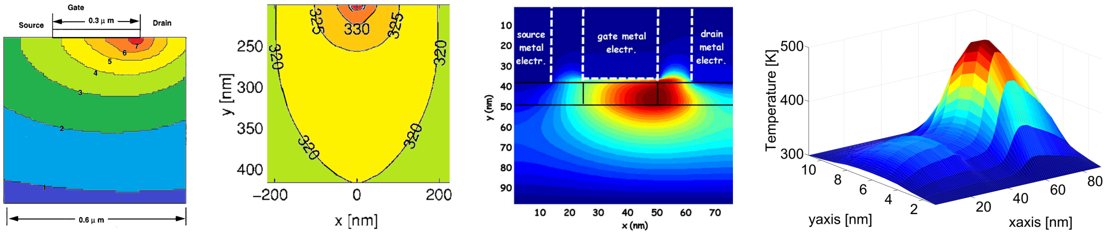
(c) IBM zEC12 processor in 32 nm SOI [62]

(d) IBM System Z processor in 22 nm SOI [63]

Figure 1.3 – Block level thermal simulations performed on different processors in SOI technologies.

analysis, first the power dissipation and the heat generation of high and low workloads of the chip is obtained by the electrical simulations. The extracted power traces are used during the thermal simulations in order to extract the temperature maps (Figure 1.3). The simulations are performed with steady state or transient heat diffusion theory since the grain size of the simulation is relatively large (tens of micrometers). The obtained temperature maps are useful in terms of observing the maximum and mean temperature values, temperature gradients and re-optimizing the floor-plan of the design in order to decrease the peak temperature and the temperature differences between different cores.

Different gruops have performed chip level thermal simulations with different resolutions and the obtained results were used for different solutions to decrease the possible thermal issues. In [60], the thermal simulations were performed in 100-micron-range length scale. The generated thermal maps for different workloads (Figure 1.3a) were analyzed to improve the design and floor-plan of the chip and distributing the thermal sensors, which are responsible of monitoring the temperatures of the critical regions and controlling the thermal management unit (TMU). Similar to other designs, the TMU controls the external cooling mechanisms and can interrupt the the processor and chip clocks externally if the maximum temperature exceeds certain threshold values. In [62], the temperature data (Figure 1.3c) obtained from thermal simulations was verified by the measurements. For that a dedicated thermal test vehicle built for calibrating the parameters of the thermal model. The processor chip contained arrays of sensors for local thermal and power measurements and calibration. This test vehicle allowed testing of alternative packaging options and materials, as well as allowing early reliability testing with accelerated power and temperature conditions. The obtained temperature data was used to increase the reliability by limiting the DC and RMS current values extensively in all interconnects. The thermal analysis in [63] was performed with a relatively larger resolution



(a) 240 nm Bulk [65]

(b) 90 nm Bulk [66]

(c) 25 nm FDSOI [23]

(d) 20 nm FinFET [33]

Figure 1.4 – Device level electro-thermal simulations performed in different technology nodes and geometries.

compared to other works. New methodologies were developed to analyze the thermal aspects of the design at a variety of length scales from the gate level, all the way up to the chip level. To avoid micro hot-spots at the individual gate level, caused by device self-heating, high switching factor nets were identified during functional simulation. Gates driving these nets had their maximum output load capacitances reduced, and were spaced apart from other gates driving such nets in order to avoid excessive heating. In addition the design was broken into small tiles, with total current through low-level power vias calculated, looking for local regions of high power density. Power dissipation was then rolled up to the chip level, for a detailed thermal analysis including package and system effects. Another work, where the design is thermally analyzed at the standard cell level performed by Chen *et al.* [64]. In their work, they proposed a cell-homogenization technique to decrease the computational load of the thermal simulation without influencing the accuracy significantly. The results provide the hot-spot locations with a granularity that is comparable to the size of the logic gates.

The main drawback of large scale thermal analysis is the fact that the nanometer scale hot-spots due to the self-heating of individual transistors are not detected due to the low resolution of the simulation. In order to understand the necessity of more detailed thermal analysis, it is required to consider the individual devices in different technologies, which can be observed by device level electro-thermal simulations. With the device level electro-thermal simulations, one can detect the nanometer scale hot-spots inside the device geometry. Figure 1.4 shows the temperature profiles of the device level simulations proposed by different authors [23, 33, 65, 66]. The increasing trend of the device peak temperature with the device scaling suggests that one needs to consider the self-heating of individual devices in order to better optimize the sensitive blocks. However, estimating the device temperature with a good precision necessitates detailed physical models and the proper simulation approach especially for dimensions that are comparable to the mean free length of different quantized entities. On the other hand, more detailed simulations increase the computational complexity and the results show the information for a single device rather than a block or the entire chip. We provide more details about the state-of-the-art device level electro-thermal simulations in Chapter 2.

The last thermal simulation example is the multiscale approach [67, 68, 69]. With this approach one can obtain the thermal behaviour of a large scale block while considering the nanometer scale hot-spots due to the self-heating of individual devices. In this approach, the entire chip is thermally simulated with heat diffusion equation at a coarser grain size, while Boltzmann-Transport equation is used for obtaining the temperature values of the individual devices. This approach provides a large scale temperature map including the nanometer-scale local hot-spots. However, due to the detailed models and large size of the blocks it is computationally heavy.

In our block level thermal simulations, we analyze the effect of heating resolution on the simulation accuracy by performing simulations with a wide resolution range from block size down to device size in bulk and FDSOI technologies. On the other hand, we verify the obtained results by comparing the block level thermal simulation results with more detailed device level electro-thermal simulations. The details of our simulations are provided in Chapter 2 and Chapter 3.

1.3 Key Contributions of This Thesis

The main contribution of this thesis can be classified into three different categories, where the self-heating effects are investigated in **device level**, and self-heating aware design approaches are applied on the reliability of **digital circuits** and noise performance of **analog circuits**.

During the investigation of self-heating effects with device level electro-thermal simulations,

- A detailed FDSOI and bulk geometry comparison of short channel devices in different technology nodes is performed. **The importance of the geometry on the amount of temperature increase inside a single device is demonstrated.** With the simulations performed on different technology nodes, it is shown how critical the self-heating effects become from one technology node to another in FDSOI and bulk geometries, and the future trajectory of self-heating effects on upcoming technology nodes is estimated.
- While performing device level simulations, the FDSOI devices are analyzed from a circuit design perspective and **important guidelines that are useful in the design of analog and digital blocks considering the self-heating effects are provided. The influence of the device size and its operation region (linear or saturation) on its heating and temperature profile is demonstrated**, which should be considered in circuit design to minimize the self-heating effects and maximize the reliability and performance.

In the digital part of our work,

- The results of the device level electro-thermal simulations are applied on the design of very high-speed digital circuits. Thermal simulations are performed on a much larger scale and comparisons are made between FDSOI and bulk in addition to the investigation of technology scaling on the temperature profile and reliability of high performance digital circuits.

- Appropriate techniques for performing thermal simulations of large-scale blocks in FDSOI and bulk geometries are shown by providing temperature maps and the necessary computational resources to get reliable results.
- A very detailed heat generation analysis is performed considering each transistor of the implemented high-speed digital blocks. With the obtained results, **reliability guidelines are provided to increase the lifetime of these circuits** by proper design of certain group of devices.

In the analog part of our work,

- **The self-heating effects on the thermal noise of FDSOI devices are shown for the first time** to the best of our knowledge by performing electrical measurements. The self-heating aware design of low noise amplifiers are explained in detail. Proper sizing and biasing techniques are given to get the lowest noise figure while being aware of the fact that the self-heating effects are present.
- Integrated inductors with different geometries are designed in 28 nm FDSOI technology. To have an evaluation of the high frequency losses and a handy design of integrated inductors, **a figure of merit is presented by using a simple lumped circuit.**
- Electrical measurements are also performed on different amplifier structures in 28 nm FDSOI technology to show **the self-heating effects on the flicker noise.**

1.4 Thesis Outline

This thesis is organized as follows: Chapter 2 gives a device level analysis of self-heating effects. In this chapter, our approach in device level electro-thermal simulations is explained along with other approaches to model the physical phenomena responsible of self-heating effects in a device. The significance of self-heating effects in bulk and FDSOI devices and the effect of device scaling are investigated by performing detailed device level electro-thermal simulations. The influence of biasing conditions and different regions of operation (linear and saturation) on the self-heating effects are analyzed to be applied on the design of more complex circuit blocks.

In Chapter 3, self-heating effect on the reliability of high-speed digital circuits are examined. The full-custom design of two high performance 64-bit Kogge-Stone adders in 40 nm and 28 nm technologies are given with their thermal simulations in FDSOI and bulk geometries. A bulk-FDSOI and a technology scaling comparison are performed by using these implementations. An extensive analysis is performed on the designs to detect the devices that threaten the reliability the most. Finally, some design guidelines are proposed to increase the lifetime of high performance digital implementations in FDSOI.

Chapter 4 explains how the thermal and flicker noise are affected by self-heating effects. This chapter provides a theoretical basis and motivation for the implementations and performed measurements in Chapter 5 and Chapter 6.

In Chapter 5, detailed guidelines for the thermal aware design of LNAs are provided. Simulations and electrical measurements show the dependency of the noise figure on the self-heating effects by comparing LNA implementations with different gate lengths. Additionally, the design of integrated inductors for RF applications are explained in detail.

Chapter 6 focuses on the self-heating effects on the flicker noise. The details of the implemented test structures to measure the flicker noise of FDSOI devices are explained. At the end of the chapter, the measurement results are provided with the explanations of the influence of self-heating effects on the performance.

Finally, Chapter 7 concludes the thesis.

2 Device Level Analysis of Self-Heating

In this chapter, we provide the details and result of our device level electro-thermal simulations performed on bulk and FDSOI devices. The results of this chapter will be used in the later chapters to increase the reliability and the performance of larger scale blocks.

2.1 Introduction

Device level simulations are very useful tools in understanding the influence of various parameters on different measures. Especially for the temperature, which is not estimated in many circuit level simulators, electro-thermal device level simulations are the only option in order to obtain the real behaviour. In this chapter, we explain the methodology of our device level electro-thermal simulations and present the results obtained on different FDSOI and bulk devices. Our main targets and motivations in performing electro-thermal device level simulations are

- to understand the self-heating phenomena in more detail
- to make comparison between different geometries and technologies
- to observe the effect of scaling
- to identify the thermal trends by sweeping a certain parameter and finding the optimum point

so that the obtained knowledge can be applied on large scale circuits that will be presented in the following chapters.

The chapter is organized as follows. Section 2.2 describes the details of device level electro-thermal simulations and different techniques and attempts made by different authors, Section 2.3 provides the details of our block level thermal simulations, Section 2.4 provides a comparison between bulk and FDSOI, Section 2.5 shows the influence of different parameters on the self-heating of FDSOI devices. Finally, the summary of the chapter is presented in Section 2.6.

2.2 Theoretical Background and Related Work

The self-heating in semiconductors is caused by the interactions between the carriers accelerated by the applied electric field and the lattice. For the MOSFET devices, these interactions mainly occur inside the channel where the carriers are transported from the source to the drain. The kinetic energy of the carriers is transferred to the lattice by scattering with acoustic and optical phonons [18, 24]. Depending on the heat conduction properties of the materials and the type of the scattering mechanism, the transferred energy (i.e. generated heat) diffuses and leaves the channel with a certain rate. On the other hand, the amount of the excess energy that stays inside and the vicinity of the channel results in increased temperature of the device and its surrounding.

From a simulation point of view, one can split the thermal effects into two main parts: heat conduction and generation [70]. For both of them, there are various physical models. The selection of the physical model is quite crucial because the dimension of the simulated geometry and granularity of simulation has a significant influence on the reliability of the results. For example, neglecting the recombination heat might result in an incorrect estimation of the temperature profile of a device in a device level simulation, while considering the quantum effects might cause a very large simulation time for the thermal simulation of a large digital circuit in a block level simulation. Therefore, someone should be careful while selecting the proper physical models for both heat conduction and heat generation depending on the size of the simulated volume and the expected accuracy. In the following, we describe different physical models and their suitability for different dimensions.

2.2.1 Block Level Thermal Simulations

Fourier's law of heat conduction, which gives the classical continuum heat diffusion equation in a material, is a useful tool for estimating the resulting temperature distribution profiles at large scales through heat conduction [71].

$$\rho c_p \frac{\partial T}{\partial t} = \kappa \nabla^2 T + q \tag{2.1}$$

In the above equation T is the temperature [K], κ is the thermal conductivity [$\mathrm{Wm^{-1}K^{-1}}$], ρ is the volumetric mass density [$\mathrm{kg/m^3}$], c_p is the specific heat [J/kg-K] and q is the power density [$\mathrm{W/m^3}$] of the heat generators. At the steady state conditions, the time derivatives in (2.1) vanishes and the equation becomes

$$0 = \kappa \nabla^2 T + q. \tag{2.2}$$

For modelling the heat generation term in (2.2), the most direct approach is the Drift-Diffusion (DD) formulation

$$q = \mathbf{J} \cdot \mathbf{E} \tag{2.3}$$

Table 2.1 – Duality between electrical and thermal circuits at steady state

Electrical	Unit	Thermal	Unit
Potential Difference	V	Temperature Difference	K
Electric Current	A	Heat flow	W
Voltage Source	-	Fixed Temperature	-
Current Source	-	Heat Generator	-
Resistance	Ω	Thermal Resistance	K/W

where **J** is the current density [A/m^2] and **E** is the electric field [V/m].

The demonstrated physical models for heat conduction and heat generation contain equations with continuous functions. To perform thermal simulations on a computer, (2.2) and (2.3) can be integrated with respect to space inside a proper volume to discretize the simulation space. The integration volume determines the resolution of the simulation and integrated values are similar to the lumped circuit components. Hence, one can model the thermal geometry as a lumped electrical circuit [42] where the dualities can be seen on Table 2.1.

Most of the chip and block level thermal simulation approaches are based on the use of (2.2) and (2.3) [72, 73, 74, 75]. This approach could provide good estimation about the average temperature values of the constitutive blocks of a chip when the resolution is not so high [73, 74, 75]. However, having low resolution fails to detect the locations of nanometer scale hot-spots inside the channel which are present especially in circuits implemented in FDSOI technology [72]. A more elaborated modelling approach is used by Hassan *et al.* [68, 69]. In their work, heat equation derived from Fourier's Law is solved for larger scales to find the average temperature of the blocks and Boltzmann Transport Equations (BTE) are used to get the nanometer scale hot-spots. In our block level thermal simulations, we benefited from the HotSpot [76] tool where the heat conduction is modeled by (2.1). For bulk and FDSOI technologies, we provided different geometries and modified the thermal conductance of different materials and boundaries to account for the thin film thermal conductance and interface thermal resistance. More detail explanation of our block level simulation approach is given in Chapter 3.

2.2.2 Device Level Thermal Simulations

For the cases where the device size are comparable to the mean free path of different quentized entites (≈ 290 nm for phonnons and ≈ 20 nm for electrical charge carriers) (2.1) and (2.3) cannot be used as they fail to take into account the granularity of heat conduction in semiconductors as explained previously. A more correct picture of the phonon transport at very small length scale (shorter than the acoustic phonon mean free path, ~300 nm [77]) can be expressed by the Boltzmann Transport Equation [18],

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f = \left.\frac{\partial f}{\partial t}\right|_{coll} + \left.\frac{\partial f}{\partial t}\right|_{g} \tag{2.4}$$

where $f(\mathbf{r},\omega,t)$ is the phonon distribution function [m^{-3}] and $\mathbf{v}$ is the phonon velocity [m/s]. The first term in (2.4), which is due to the phonon collisions, can be approximated using the relaxation time approximation as,

$$\left.\frac{\partial f}{\partial t}\right|_{coll} = \frac{f_0 - f}{\tau_{ph}} \tag{2.5}$$

where $f_0 = 1/(exp(\hbar\omega/k_B T) - 1)$ is the equilibrium Planck distribution at temperature T and τ_{ph} is the average phonon scattering time [sec]. (2.4) can be integrated over the phonon frequency and density of states to get,

$$\frac{\partial u}{\partial t} + \mathbf{v}\cdot\nabla u = \frac{u_0 - u}{\tau_{ph}} + q \tag{2.6}$$

in terms of the phonon energy density, u [J/m^3], and heat generation term, q [18].

For the device level simulations, the heat generation term is modelled differently in each method employed. The simplest approach is to use the Drift-Diffusion approximation shown by (2.3). By including the generation and recombination rates, (2.3) can be written as [10],

$$q = \mathbf{J}\cdot\mathbf{E} + (R - G)(E_g + 3k_B T) \tag{2.7}$$

where R and G are the recombination and the generation rates [m^{-3}s^{-1}] respectively. This approach is not suitable for nano-scale devices, as it does not take into account the microscopic non-locality of the phonon emission near the peak of the electric field, and thereby result in a displacement of a few electron mean free path lengths between the maximum electric field point and the complete transfer of the energy to the lattice. A more sophisticated approach is used in the Hydrodynamic model [18],

$$q = \frac{3}{2}k_B\frac{n(T_e - T_L)}{\tau_{e-L}} + (R - G)\left[E_g + \frac{3}{2}k_B(T_e + T_L)\right] \tag{2.8}$$

where T_e is the electron temperature, T_L is the lattice temperature and τ_{e-L} is the energy relaxation time [sec] and k_B is the Boltzmann constant. This method has better resolution of the non-locality of the phonon emission near the electric field peak, but has problems due to the simplifications of using only a single carrier temperature and energy relaxation times. In addition, this approach ignores the differentiation between the different types of the phonons (acoustic and optical) emitted. The most accurate method to approach this situation, which is capable of taking into consideration the entire phonon dispersion spectra is the Monte-Carlo method [18],

$$q \sim \frac{1}{t_{sim}}\Sigma(\hbar\omega_{ems} - \hbar\omega_{abs}) \tag{2.9}$$

where t_{sim} is the simulation time. As can be expected, a more complex method is more computationally expensive than a simple one.

Different authors have attempted to approach this trade-off in different ways. Fiegna *et al.* [20, 21] calibrated a Drift-Diffusion model at different temperatures by comparison with a full-band self-consistent MC simulator for a 3-D electron gas by modifying the mobility parameters in the DD model to match the isothermal transfer and drain characteristics obtained from the two simulators. On the other hand, Sadi *et al.* [22] tried to iteratively couple a 2D Monte Carlo simulator with a 2D steady state heat diffusion equation (HDE) solver to study electro-thermal effects in a Si/SiGe MODFET. Both these approaches used the steady state heat diffusion formulation, which as discussed before is incapable to capture the granularity of the phonon dispersion. The group of Vasileska *et al.* [23, 24, 25, 26] tried to solve the coupled electron-optical phonons-acoustic phonons-heat bath problem using energy balance equations derived from the Boltzmann equations that were coupled with a Monte Carlo electric simulation. Later works tried to concentrate on the inclusion of the differentiation between the different types of the phonons. Narumanchi *et al.* [27] came up with a BTE model, which incorporated ballistic term for the acoustic modes, but ignored the optical modes. Wang [28] developed a method for computing the scattering rates for the different phonon modes from the perturbation theory. A simpler anisotropic relaxation time phonon BTE model was developed by Ni [29], where the relaxation time is a function of the wave vector. Misawa *et al.* [30] reported the development of a Monte Carlo approach with different time scales for the electron and phonon transport, in which the group velocities of the phonons were calculated from the dispersion curves reported by Azuhata *et al.* [31].

Our device level simulations were performed using the Sentaurus TCAD tools provided by Synopsys. To keep the simulation times reasonable, we used the hydrodynamic simulation approach to perform the electro-thermal simulations. The temperatures of both the n-type (electrons) and p-type (holes) carriers were calculated in addition to the temperature of the lattice. For heat conduction, we used the energy balance equations that are derived from Boltzmann Transport Equation, the details of which are given in Section 2.3.

2.3 Details of Our Device Level Simulations

Our device level simulations were performed using the Sentaurus TCAD tools. The device structure and doping were defined using the Sentaurus Structure Editor. The simulations were performed in the Sentaurus Device tool, which is in effect a numerical solver. The predominant method of solving used in our simulation was the hydrodynamic simulation model with the temperature equation implemented by means of energy balance equations. Both the carrier temperatures were calculated in addition to the lattice temperature. Sentaurus device uses a simpler formulation which follows the work of Bløtekjær [78] and Stratton [79], but without any convective terms. The energy balance equations for the electrons (2.10), the holes (2.11) and the lattice (2.12) are derived from the different moments of the Boltzmann Transport Equation [80].

$$\frac{\partial W_n}{\partial t} + \nabla \cdot \mathbf{S_n} = \mathbf{J_n} \cdot \nabla E_C + \left.\frac{dW_n}{dt}\right|_{coll} \tag{2.10}$$

$$\frac{\partial W_p}{\partial t} + \nabla \cdot \mathbf{S_p} = \mathbf{J_p} \cdot \nabla E_V + \left.\frac{dW_p}{dt}\right|_{coll} \tag{2.11}$$

$$\frac{\partial W_L}{\partial t} + \nabla \cdot \mathbf{S_L} = \left.\frac{dW_L}{dt}\right|_{coll} \tag{2.12}$$

The W_n, W_p and W_L are the electron, hole and lattice energy densities. The energy flux terms are formulated as $\mathbf{S_n}$, $\mathbf{S_p}$ and $\mathbf{S_L}$. The current densities of electrons and holes are $\mathbf{J_n}$ and $\mathbf{J_p}$ and the last terms denote the energy exchange between the carriers and the lattice due to collisuons. For the solution of the current density and the potential profiles along the device geometry Poisson and continuity equations are used together with the energy balance equations [80]. The energy balance formulation shown by (2.10), (2.11), and (2.12) is very similar to the energy conservation equations used by Vasileska [23, 24] except that energy densities of the acoustic and the optical phonons are not separately formulated. Finally, the lattice temperature profile is obtained by (2.13) where c_L is the specific heat of the Silicon lattice.

$$T = \frac{W_L}{c_L} \tag{2.13}$$

We activated the effective bandgap narrowing model to include the effects of temperature and doping on the bandgap of Silicon. The Philip's unified mobility model proposed by Klaassen [81] was used with saturation of mobility at higher fields. Only Shockley-Read-Hall recombination was considered in our simulations. Direct band recombination is not expected to occur in an indirect band gap semiconductor like Silicon. Similarly, the Auger recombination was ignored, as the doping densities in the channel was not very high. Impact ionization was also ignored, as the devices considered are mainly to work in the sub 1V regime, where impact ionization is not expected to occur [82].

For FDSOI devices, where the channel is located inside a Silicon thin film, it is important to consider the effect of the film thickness on the thermal conductivity. The thermal conductivity of the material is decreased mainly due to the phonon-boundary scattering when the thickness is smaller than the phonon mean free path length. We used a modified version of the approach of Sondheimer [83] which assumes that the phonon boundary scatterings are purely diffusive. For a semiconductor film of thickness t, where the z-axis is perpendicular to the plane of the film located between $z = 0$ to $z = t$, the thermal conductivity of the film for $0 < z < t$ can be expressed as [25],

$$\kappa(z) = \kappa_0(T) \int_0^{\pi/2} \sin^3\theta \left\{ 1 - \exp\left(-\frac{t}{2\lambda(T)\cos\theta} \right) \cosh\left(\frac{t-2z}{2\lambda(T)\cos\theta} \right) \right\} d\theta \tag{2.14}$$

where $\lambda(T)$ is the temperature dependent mean free path of phonons in bulk given by,

$$\lambda(T) = \lambda_0 \left(\frac{300}{T}\right) \text{nm} \tag{2.15}$$

where λ_0 is the room temperature mean free path and equal to 290 nm. On the other hand, Sentaurus only takes into account the temperature dependence, given by [80]

$$\kappa_0(T) = \frac{1}{a + bT + cT^2} \tag{2.16}$$

where $a = 0.03$ cm-K-W^{-1}, $b = 1.56 \times 10^{-3}$ cm-W^{-1} and $c = 1.65 \times 10^{-6}$ cm-K^{-1} – W^{-1}. This is similar to the expression used by Vasileska *et al.* [25]. Finally, the average thermal conductivity of the thin film can be obtained as,

$$\kappa_{av} = \int_0^t \kappa(z) dz \tag{2.17}$$

This integrated expression, for the required thickness, was used to modify the value of the constants in (2.16) for our simulations.

2.4 Comparison of FDSOI and Bulk

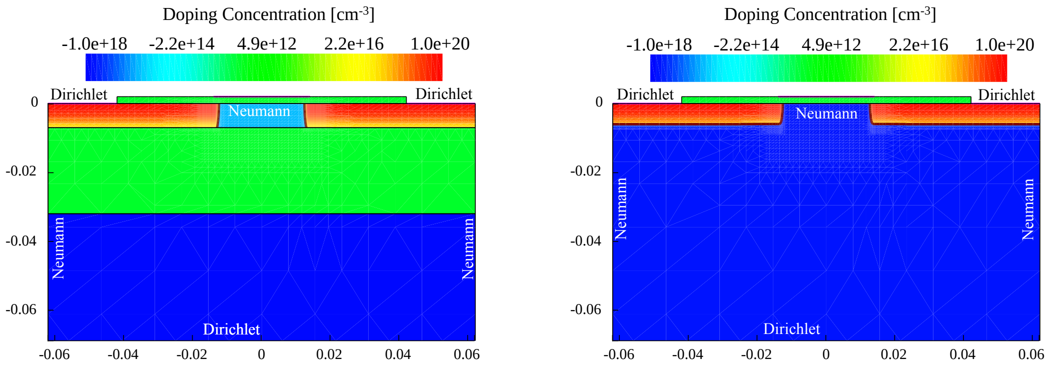
In this section, the thermal behaviours of FDSOI and bulk geometries are compared by presenting the results of our device level simulations. The simulations were performed in 40 nm and 28 nm technology nodes. In total four different devices were simulated. These devices are

- 40 nm Bulk
- 40 nm FDSOI
- 28 nm Bulk
- 28 nm FDSOI

While modelling the device geometries, the technology parameters of commercially available 40 nm bulk and 28 nm FDSOI nodes were used. The details of the device structure and process parameters are explained in the following section, which is followed by the simulation results and discussions.

2.4.1 Physical Structure and Thermal Boundary Conditions

The physical structure of the simulated bulk and FDSOI devices can be seen on Figure 2.1 for the 28 nm case. The device cross sections look quite similar in 40 nm except the dimensions. The other parameters of all devices (dimensions, doping etc.) can be observed on Table 2.2.



(a) FDSOI

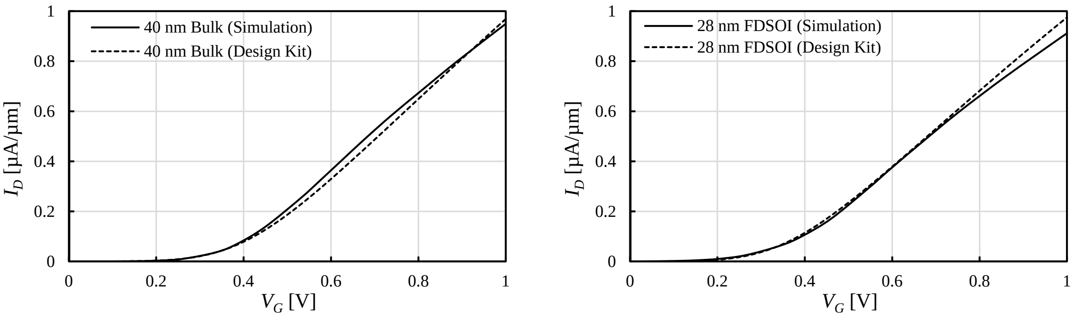
(b) Bulk

Figure 2.1 – Doping profiles of (a) the FDSOI and (b) the bulk device structures. The positive doping values are used to denote n-type doping, and negative values for the p-type doping.

Table 2.2 – Parameters of the bulk and the FDSOI devices of Figure 2.1

Parameter	**Bulk**	**FDSOI**	**Unit**
	40 nm / 28 nm	40 nm / 28 nm	
Gate oxide thickness	2	2	nm
Thin film Si thickness	-	10 / 7	nm
BOX thickness	-	50 / 25	nm
Substrate thickness (under Gate)	200 / 100	200 / 100	nm
Source/Drain contact width	60 / 40	60 / 40	nm
Gate to contact distance	40 / 28	40 / 28	nm
Source/Drain peak n-doping	1×10^{20}	1×10^{20}	cm^{-3}
Channel p-doping	1×10^{18}	1×10^{16}	cm^{-3}
Substrate p-doping	1×10^{18}	1×10^{18}	cm^{-3}
$\mathrm{Si-SiO_2}$ interface thermal resistance	2×10^{-8}	2×10^{-8}	$\mathrm{m^2KW^{-1}}$

As we are interested in the thermal behaviour of a single device, it is imperative to provide reasonable thermal boundary conditions at the device level. For the block level thermal simulations, which will be explained in Chapter 3 in detail, it is assumed that the upper part of the circuit (metal routing and SiO_2 isolation) is thermally insulated due to low thermal conductance. Eventually, the heat removal from the entire chip is mainly provided through the substrate, heat spreader and the heat sink (Figure 3.3b). However, when a single device is considered, this assumption loses its validity since there might be significant heat transport between the devices through the upper part of the active region especially via metal interconnects. Consequently, the possible heat transport paths on the upper part of the device should also be considered and modelled. These two different boundary conditions for the block level and the device level thermal simulations might seem contradictory at first glance. However, they are valid assumptions since the transported heat between the devices through the upper part of the chip will be removed out mainly through the bottom heat sink rather than the upper SiO_2 isolation which has a very low thermal conductivity.



(a) 40 nm bulk

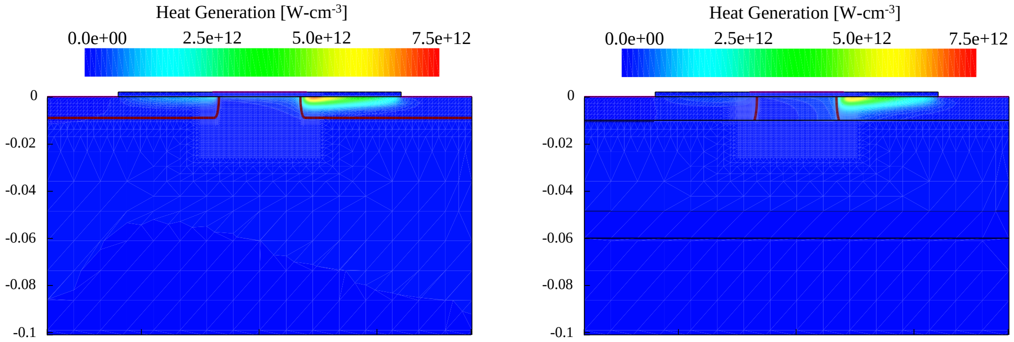
(b) 28 nm FDSOI

Figure 2.2 – Comparison of simulated 40 nm bulk and 28 nm FDSOI MOSFETs and the design kit models in terms of their I_D-V_G characteristics.

While choosing the thermal boundary conditions, each contact and side wall of the device were considered separately. Since the source and the drain contacts are implemented with metal, their thermal conductance is high. The thermal conductance at substrate contact is also high due to the high thermal conductance of bulk Silicon. Therefore, it was assumed that the source, drain and substrate contacts are the main heat removal paths. Consequently, the temperature at these contacts are fixed to a constant value by applying Dirichlet boundary conditions. The constant temperature at the source, drain and substrate contacts were fixed to the room temperature, while the temperature in the active region of the device might differ depending on the local heat generation points. On the other hand, the thermal conductance of the gate is low due to the SiO_2 isolation layer and the interface thermal resistances. Therefore, it was assumed that the gate is thermally insulated and the heat flow through gate contact is set to 0 W/cm^2 by applying Neumann boundary conditions which is in accordance with [18]. Finally, for the side walls Neumann boundary conditions with 0 W/cm^2 heat flow is applied since it is assumed that the simulated device is next to a similar device with similar heat generation. This assumption would usually be the case both in digital logic gates where continuous diffusion is used and analog devices implemented with multi-fingers.

2.4.2 Results and Discussions

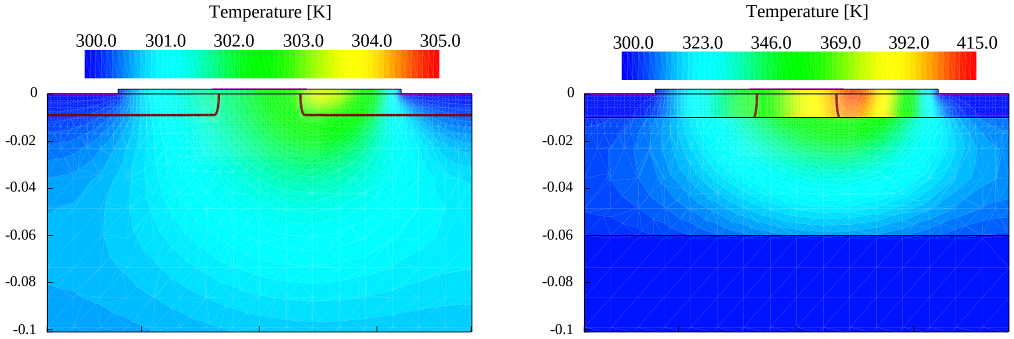
2D quasi-stationary simulations were performed for both bulk and FDSOI devices. A nominal operating region with V_{DS}=V_{GS}=V_{DD} = 1.0 V was selected to study the spatial spread of the heat generation and the lattice temperature in both the devices. Both the devices are in saturation at this biasing. On Figure 2.2, the I_D-V_G characteristics of the simulated devices are compared with the models of the provided design kits of 40 nm bulk and 28 nm FDSOI technologies. It can be seen that the results of the device level simulations and the models match quite well. However, for the 28 nm FDSOI device, the drain current starts to deviate slightly from the model provided by the manufacturer. This is due to the large temperature rise in the FDSOI device at large bias conditions and the resulting reduction in mobility.



(a) Bulk

(b) FDSOI

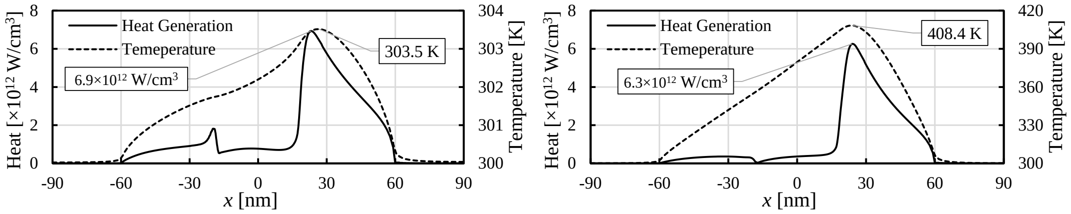
Figure 2.3 – Spatial heat generation profile of 40 nm (a) bulk and (b) FDSOI devices.



(a) Bulk

(b) FDSOI

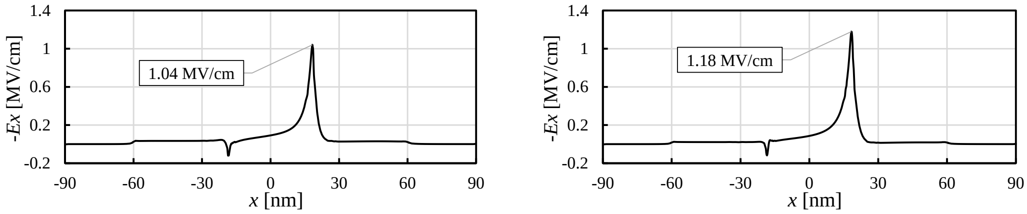
Figure 2.4 – Spatial temperature profile of 40 nm (a) bulk and (b) FDSOI devices.



(a) Bulk

(b) FDSOI

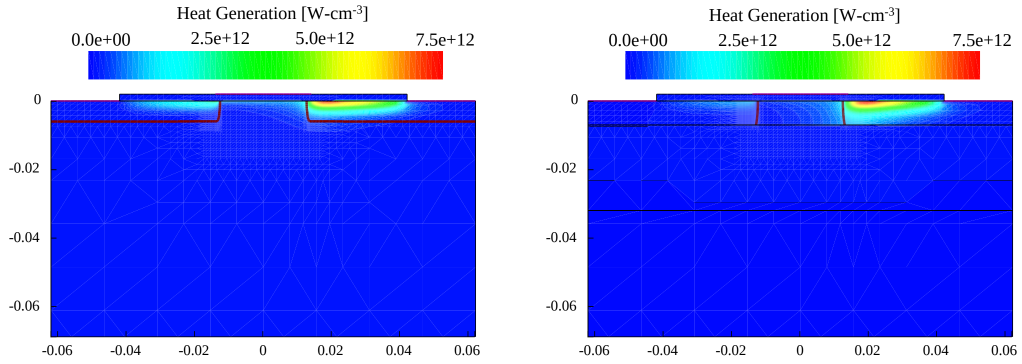
Figure 2.5 – Heat generation and temperature profiles of 40 nm (a) bulk and (b) FDSOI at $y = -1$ nm.



(a) Bulk

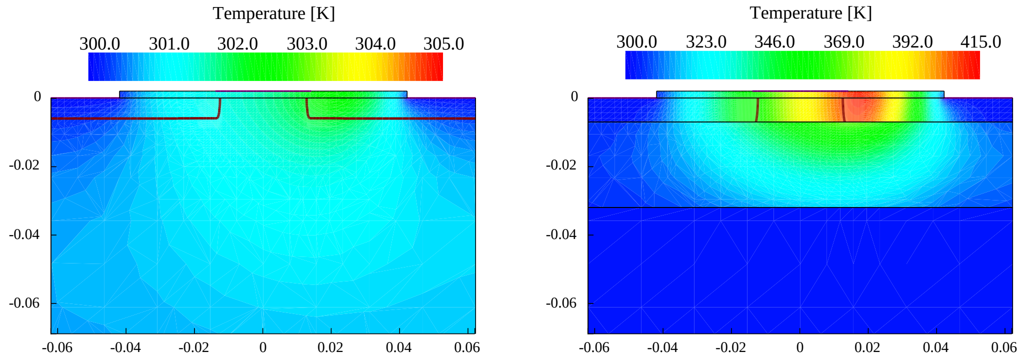
(b) FDSOI

Figure 2.6 – Lateral electric field in 40 nm (a) bulk and (b) FDSOI MOSFETs at $y = -1$ nm.



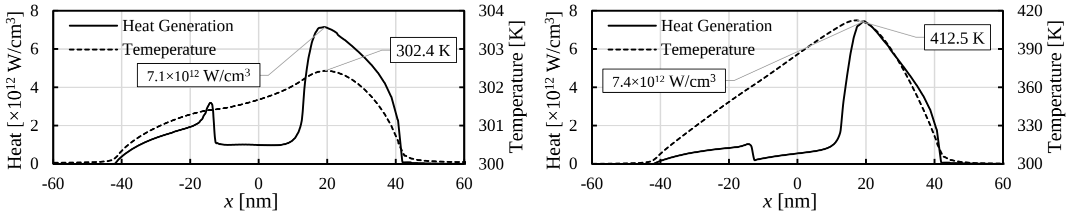
(a) Bulk (b) FDSOI

Figure 2.7 – Spatial heat generation profile of 28 nm (a) bulk and (b) FDSOI devices.



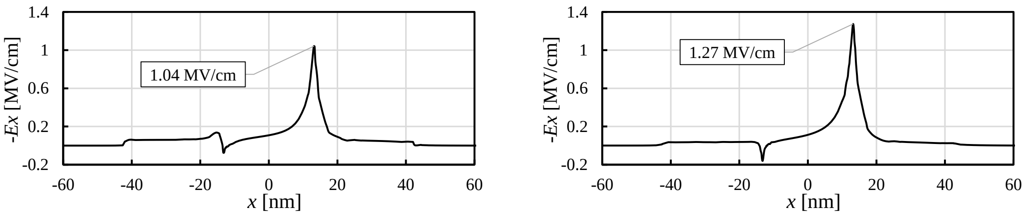
(a) Bulk (b) FDSOI

Figure 2.8 – Spatial temperature profile of 28 nm (a) bulk and (b) FDSOI devices.



(a) Bulk (b) FDSOI

Figure 2.9 – Heat generation and temperature profiles of 28 nm (a) bulk and (b) FDSOI at y = -1 nm.



(a) Bulk (b) FDSOI

Figure 2.10 – Lateral electric field in 28 nm (a) bulk and (b) FDSOI MOSFETs at y = -1 nm.

Heat Generation

The mechanism of heat generation inside a semiconductor device has already been discussed in detail. The heat generation is dependent on the carrier flow and the scatterings in the device and is not expected to vary much with the device structure under similar biasing conditions. The spatial spreads of the heat generation inside all the devices at a biasing of V_{DS} = V_{GS} = V_{DD} = 1.0 V, as obtained from the simulations, is illustrated on Figure 2.3 and Figure 2.7 for the 40 nm and the 28 nm technologies respectively. Additionally, Figure 2.9 and Figure 2.5 show the heat generation profiles in the channel just below the gate. The heat generation is similar when bulk and FDSOI devices are compared. They are also similar for the two technology nodes except that the peak heat generation is slightly higher in 28 nm. There are some intriguing observations that can be made from these plots. Firstly, a great percentage of the heat generation is occurring close to the drain contacts in both devices which is in line with the reported data [18, 25, 26]. The situation can be explained by observing the lateral component of the electric field in the channel (E_x), which is shown on Figure 2.6 and Figure 2.10 for different cases. The electric field reaches its maximum value around the pinch-off point, close to the drain side of the channel. The size of this region, where the electric field is maximum, is quite smaller than the mean free path of the electrons. Hence, in this region the electrons are mostly transported in a quasi-ballistic manner by gaining a significant amount of energy. After being transported into the drain region, the electrons transfer the gained energy via electron-phonon scattering close to the drain contact of the device. This results in a heat generation profile concentrated close to the drain of the device. This situation can be better understood by comparing the electric field and heat generation profiles for each case. It can be seen that the heat generation increases quite rapidly at the point where electric field reaches its maximum. However, the peak value of the heat generation is slightly on the right side of the electric field maximum due to the ballistic transport of the carriers. The heat generation is laterally more widely distributed than the electric field, which shows a very sharp peak characteristic. It is high even at the points close to the drain contact. This shows that there are significant electron-phonon interactions very close to the drain contact. Another observation to be made is regarding the heat generation at the source side of the channel. All devices have a small heat generation at the source side. However, the bulk devices show higher and wider heat generation compared to the FDSOI devices. Consequently, the heat generation in FDSOI device is even more concentrated on the drain side then the bulk device, which would result in a higher temperature of the FDSOI device.

Temperature

As discussed before, the lattice temperature inside the device is dependent both on the heat generation in the device and the heat transfer from the device. For this reason, the temperature profile is expected to be quite dependent on the device structure and the thermal characteristics of the materials in addition to the heat generation profile. The lattice temperature distributions inside all devices at a biasing of V_{DS} = V_{GS} = V_{DD} = 1.0 V, as obtained from the simulations, are illustrated in on Figure 2.4 and Figure 2.8. The lattice temperature distribution is drastically different when bulk and FDSOI are compared. For the bulk device, the maximum temperature rise inside the device is around 3 K in both technologies, which is almost negligible. On the other hand, in the FDSOI device, the maximum temperature rise values are alarming 108 K and 112 K for 40 nm and 28 nm respectively. The location of the lattice temperature peak is close to drain, similar to the location of the maximum heat generation as expected. The proximity of the temperature peak to the drain contact would cause the heat leave the device mostly through the drain contact. This would increase the temperature of the drain contact, hence the probability of electromigration.

The large difference in the maximum temperature for the bulk and the FDSOI cases can be explained by the different physical device structures. The buried oxide layer not only isolates the channel electrically from the rest of the device, it also isolates it thermally. As it is clearly observable, this isolation proves to be very costly from the thermal point of view, as the substrate plays a much bigger role in the heat removal from the device than the other contacts. In addition, the channel of the FDSOI is a thin film layer of Silicon, which has a lower thermal conductivity compared to bulk Silicon, as we had discussed before. The peak temperature is larger in 28 nm FDSOI than the 40 nm FDSOI even though the isolation is thicker in 40 nm. The increase in the peak temperature from one technology node to another is due to the increased heat density and it shows the impact of scaling on the self-heating. In addition to this, when the overall heating of a chip is considered, one would observe a larger average temperature due to the larger density of devices in smaller technologies.

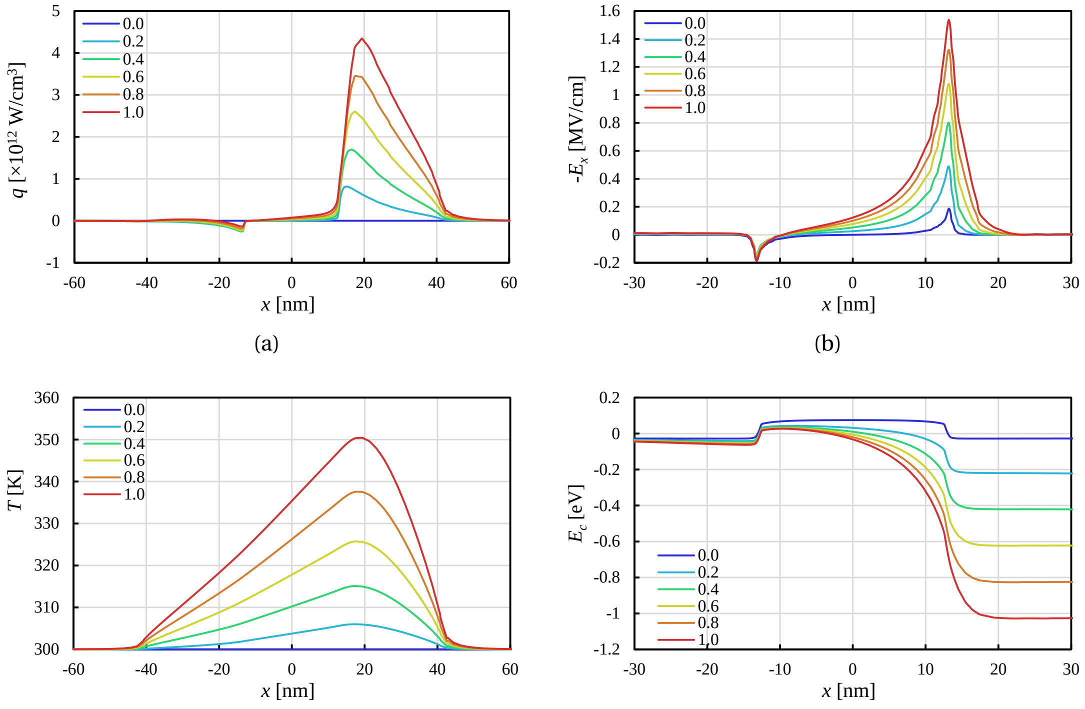
The large difference in the temperature rise between the bulk and the FDSOI devices is very similar to the results obtained from the block level thermal simulations, which are explained in detail in Chapter 3. However, the maximum temperature rise in the device level simulations is much higher than the one of the block level simulations. This is mainly because the devices are fully conducting in the device level simulations. However, none of the devices in the block level case is operating with 100% activity rate.

2.5 Experiments on FDSOI

By performing a comparison with the device level electro-thermal simulations, it has been shown that the self-heating effects are much more prominent in FDSOI than bulk. For this reason, we continue presenting our device level thermal simulation results of FDSOI. In this section, we focus on how different parameters affect the temperature rise in FDSOI devices.

2.5.1 Effect of Drain Voltage

To observe the effect of drain voltage, the 28 nm gate length FDSOI nMOS was simulated under a fixed gate bias of 0.6 V and different drain voltage levels from 0 V to 1.0 V. The same boundary conditions were applied as in Section 2.4. The heat generation, temperature,



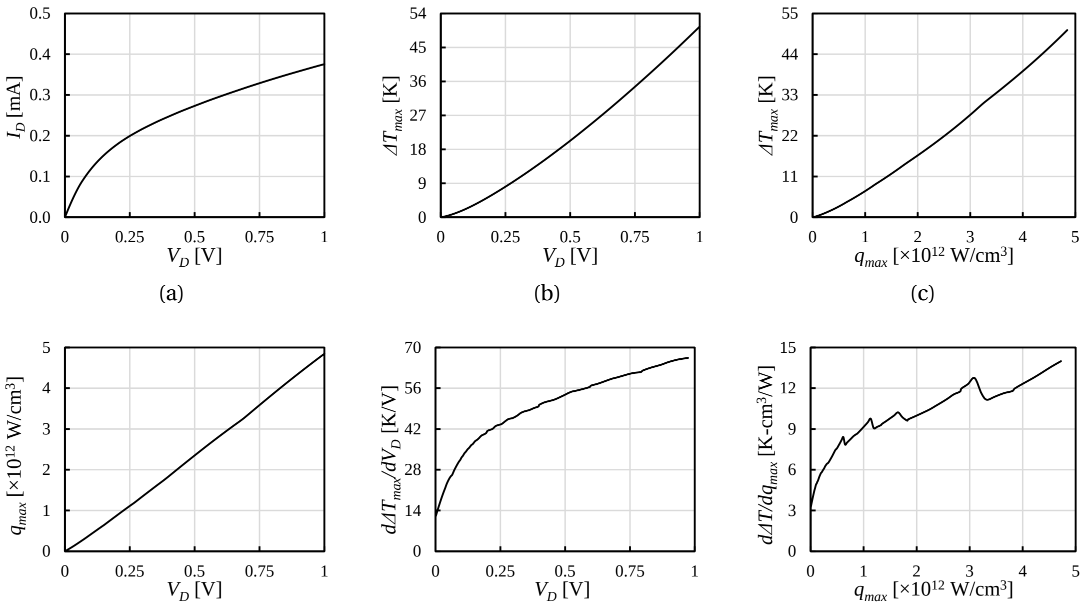
(a)

(b)

(c)

(d)

Figure 2.11 – (a) heat generation, (b) lateral component of the electric field, (c) temperature and (d) conduction band energy inside the channel for different values of V_D while V_G is constant at 0.6 V.



(a)

(b)

(c)

(d)

(e)

(f)

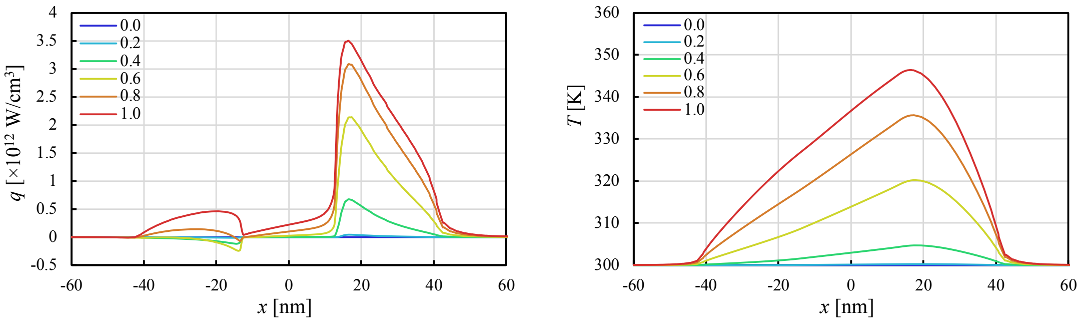
Figure 2.12 – Simulation results for V_D swept from 0 V to 1 V while V_G = 0.6 V.

electric field and the conduction band energy profiles in the channel can be seen on Figure 2.11 for different drain voltages. Additionally, Figure 2.12 shows the relations between drain voltage, drain current, maximum heat generation and maximum temperature inside the device. Different observation can be made by analysing these data. First, Figure 2.11a shows that the maximum heat generation increases linearly with V_D, which is an expected result since the power dissipation increases. The linear dependency can be observed better on Figure 2.12d. Nevertheless, the dependency of maximum temperature on V_D is not as linear as the maximum heat generation (Figure 2.11b). It is rather quadratic, which can be also seen on Figure 2.12b. As the device enters in the saturation, the heat generation shifts slightly more on the drain side due to the high-energy carriers and it becomes more localized. On the other hand, in linear region, the energy of the carriers are less and heat generation is more spread in the channel. This situation makes saturation region more dangerous in terms of having a high temperature localized hot-spot. To observe the influence of saturation region operation, we also plot the derivative of maximum temperature with respect to drain voltage (Figure 2.12e) and maximum heat generation (Figure 2.12f). It can be seen in both figures that the increase in the temperature for a unit increment in V_D and q_{max} is smaller in linear region and it increases as the device is more in saturation. Another observation can be made regarding the negative heating phenomena. A negative heating (cooling) point can be observed on the source side. This is due to the potential barrier at the junction between the source and the channel [18], which can be observed on Figure 2.11d.

The observations indicate that the self-heating effects are stronger in deep saturation operation. The hot-spot's being very close to the drain terminal makes the drain contact quite critical in terms of electromigration in FDSOI. This effect will be even more prominent in the future technologies due to the shorter distances between the drain contact and gate capacitor.

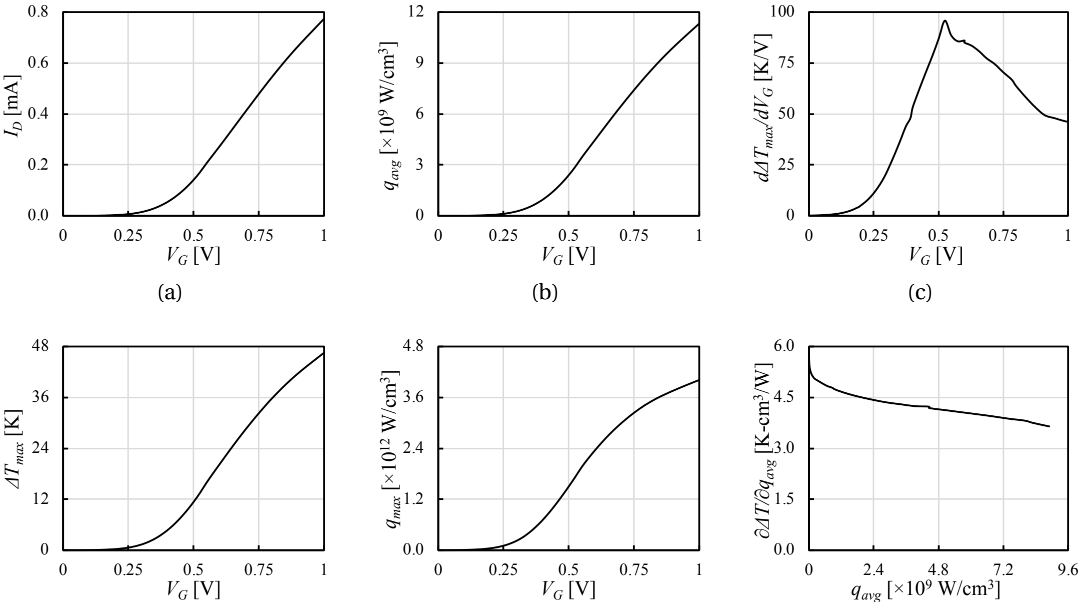
2.5.2 Effect of Gate Voltage

The gate voltage of the 28 nm FDSOI device was swept from 0.0 V to 1.0 V while the drain voltage was kept constant at 0.5 V. The heat generation and the temperature inside the channel can be observed on Figure 2.13 for different gate voltages. Additionally, Figure 2.14 shows the relations between drain voltage, drain current, maximum and average heat generation and maximum temperature inside the device. The average heat generation increases relatively linearly once the device turns on, which can be observed on Figure 2.14b. This is an expected result because the drain voltage does not change and the drain current increases linearly with the gate voltage (Figure 2.14a). Nevertheless, the temperature does not increase as fast as the average heating. In fact, it increases quite quickly for low gate bias where the device is in saturation and slows down once the saturation level decreases and the device goes into linear region. This can be also observed on Figure 2.14c and 2.14f where the rate of change of the temperature with respect to the gate voltage and the average power dissipation can be observed. The slowdown in the temperature rise of the device can be explained by observing the maximum heat generation (q_{max}) on Figure 2.14e. It can be seen that q_{max} increases



(a) (b)

Figure 2.13 – (a) heat generation and (b) temperature inside the channel for different values of V_G while V_D is constant at 0.5 V.

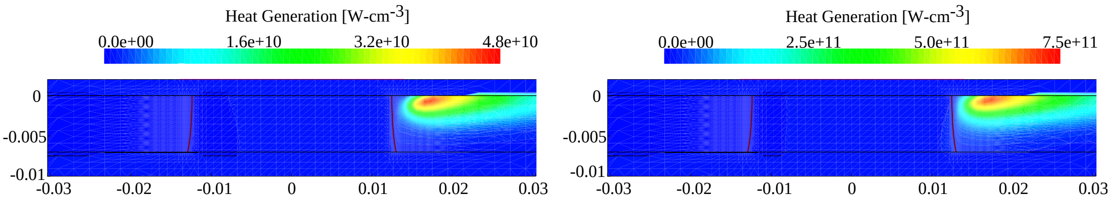


(a) (b) (c)

(d) (e) (f)

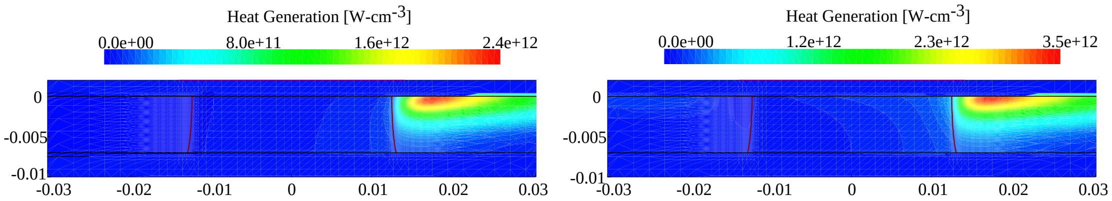
Figure 2.14 – Simulation results for V_G swept from 0 V to 1 V while V_D = 0.5 V.

quickly in saturation and its rate of change decreases as the device goes into linear region contrary to the average heating. This is the in the same direction as our previous explanations where the heating is more localized in saturation and more distributed in the linear region. The situation can be observed also on Figure 2.15, where the spatial heat generation distribution at the channel is shown for different gate bias values. It can be seen that heat generation in the channel is almost zero compared to the drain side for the V_G = 0.2 V case. On the other hand, there is significant heating in the channel and the source for V_G = 1.0 V, where the device is in linear region. This can be also observed on Figure 2.13a, where the heating in the channel gradually increases as the gate voltage increases and the device goes into linear region. For observing an extreme case, V_G = 1.5 V was also simulated and the resulting heat generation is



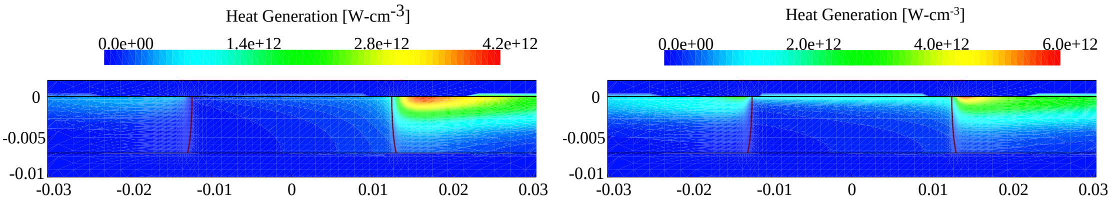
(a) $V_G = 0.2$ V

(b) $V_G = 0.4$ V



(c) $V_G = 0.6$ V

(d) $V_G = 0.8$ V



(e) $V_G = 1.0$ V

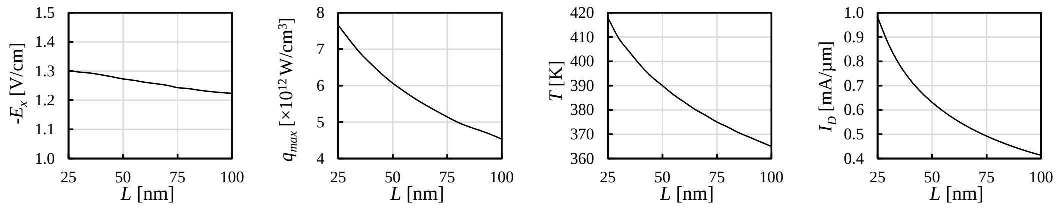
(f) $V_G = 1.5$ V

Figure 2.15 – Spatial distribution of heat generation for different V_G values from 0.2 V to 1.5 V while V_D is kept constant at 0.5 V.

plotted on Figure 2.15f, where the device is in very deep linear region. Comparing this result with Figure 2.15a shows the localization of the hot-spot in saturation region. Consequently, similar to our previous explanations, deep saturation operation results in a localized maximum heat generation and this has an effect on the maximum temperature since the temperature peak is observed very close to the peak heat generation in FDSOI.

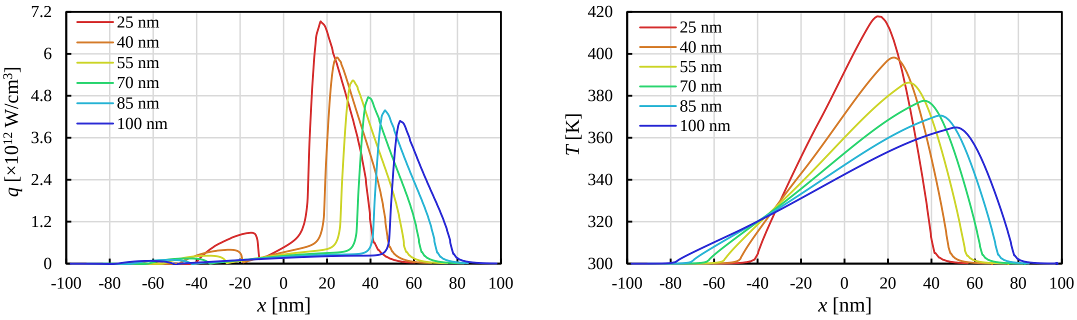
2.5.3 Effect of Gate Length

We have previously mentioned that the large value of the electric field in the pinch-off region is the most fundamental reason of having a localized hot spot on the drain end of the device. In order to understand the effect of gate length on the electric field in the pinch-off point, different devices with gate lengths from 25 nm to 100 nm were simulated. Figure 2.16a shows how the maximum lateral electric field changes with respect to the device gate length. It can be seen that the maximum electric field decreases slightly with the gate length; however, it has only a quite negligible dependency on the gate length. This is because once the device is in saturation; the voltage drop on the pinch-off region is independent of the device size as it primarily depends linearly on the applied drain voltage. Therefore, increasing the gate length has no significant effect in terms of spreading the local heating spot. Nevertheless, Figure 2.16b and Figure 2.16c show that the maximum heat density and the peak temperature



(a) (b) (c) (d)

Figure 2.16 – (a) Maximum lateral electric field, (b) maximum heat generation, (c) maximum temperature and (d) full bias drain current with respect to gate length for the 28 nm FDSOI technology.

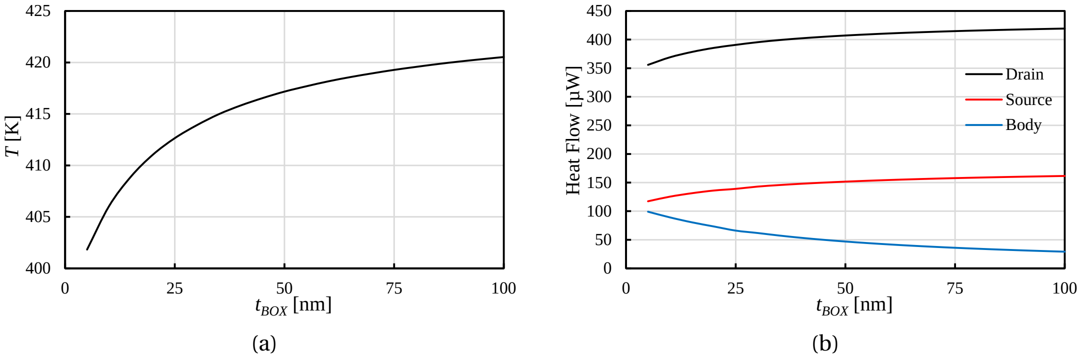


(a) (b)

Figure 2.17 – (a) Heat generation and (b) temperature inside the channel for different the gate length values in 28 nm FDSOI technology.

decreases significantly as the gate length increases. The reason can be mainly explained by observing the full bias drain current on Figure 2.16d. It can be seen that I_D has a strong dependency on q_{max}. This is because the amount of the carriers that pass the channel per unit time decreases as the gate length increases. Although the average kinetic energy of the carriers that enter into drain is independent of the channel length, their number decreases with increased gate length. Consequently, the maximum heating and the peak temperature decreases.

One might also claim that increasing the gate length should further decrease the temperature since the device area becomes larger and the heat spreads in a larger volume. This statement is true for bulk but not exactly the case for FDSOI. Figure 2.17 shows the heat density and the temperature inside the channel. It can be seen that, the peak temperature is still very close to the drain due to large thermal resistance surrounding the heating point. Consequently, increasing the area has less effect on decreasing the device *peak* temperature in FDSOI. Yet, it would decrease the average temperature of a relatively large block or the entire chip.



(a)

(b)

Figure 2.18 – (a) Temperature and (b) heat flow with respect to BOX thickness.

2.5.4 Effect of BOX Thickness

In order to observe the effect of BOX thickness, the 28 nm FDSOI MOSFET was simulated by sweeping the BOX thickness from 10 nm to 100 nm. It was observed that the maximum and the average heat generation stays constant independent of the BOX thickness. On the other hand, the maximum temperature increases with the BOX thickness, which can be observed on Figure 2.18a. This is an expected result since BOX creates a large thermal resistance under the channel. However, the increase in the temperate is more rapid for smaller BOX thickness values and it saturates, as the BOX thickness is too large. This can be understood by observing the heat flow through different terminals, which are shown on Figure 2.18b for drain, source and body terminals. It can be seen that the heat flow through the body decreases as the BOX thickness increases. On the other hand, the heat starts to prefer the drain and the source more. For very large BOX thickness values, the heat flow through body becomes negligible and all of the generated heat diffuses out through drain and source. At this point, the temperature rise of the device saturates and becomes independent of the BOX thickness.

2.5.5 Effect of Drain and Source Contact Thermal Resistance

During the device level simulations, it was assumed that the temperature of the drain and the source terminals are fixed at 300 K (i.e. Dirichlet boundary conditions are applied at drain and source). However, in a real device, the drain and the source are connected to an interconnect via metal contacts. In this section, thermal simulations are repeated by assuming that the drain and source are accessed via vertical metal contacts of different lengths. This way, the temperature of the drain and the source contacts can be observed and a more realistic estimation of the peak temperature inside the channel can be obtained. The heat generation and the temperature of the channel for different contact lengths are shown on Figure 2.19. The heat generation for different cases is quite similar since the device geometry and the bias conditions are kept constant. On the other hand, the temperature rises significantly with the increased contact length. Moreover, the temperature of the drain is larger than the source due to the location of the hot-spot. This makes the drain contact more critical in terms of having

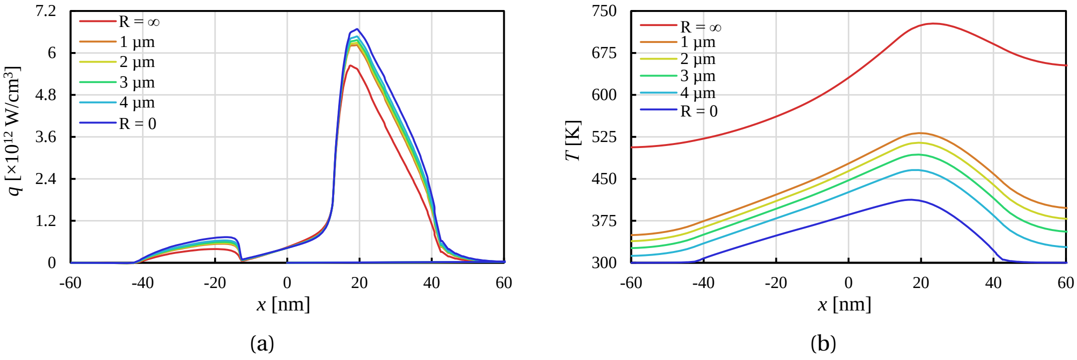


Figure 2.19 – (a) Heat generation and (b) temperature inside the channel for different thermal boundary conditions for drain and source.

an electromigration event. The red curve on Figure 2.19 corresponds to the case where the contact resistances were set to infinity. In this case, there is no heat flow through drain and source. It can be seen that the peak temperature can go up to 700 K. This is not a realistic case, but it points out the significance of the heat removal through the drain and the source in FDSOI. It can be also seen that the heat generation in this extreme case is significantly reduced, which is primarily due to increased mobility and the resulting lower current density.

2.6 Conclusion

Device level electro-thermal simulations were performed on bulk and FDSOI MOSFETs. It was demonstrated that the self-heating effects are much more prominent in FDSOI devices than bulk. It was shown that the saturation region is more critical than the linear region in terms of creating a local hot spot, which is primarily due to the large electric field closely located on the drain side of the device. It was explained that the hot-spot at the drain end makes the drain contact more critical in terms of electromigration probability.

3 Thermal Aware Design of High Performance Digital Circuits

In Chapter 2, we have performed detailed electro-thermal simulations on a *single* device and shown that the temperature of an FDSOI device can be significantly larger than a bulk one. We continue our thermal analysis on much *larger scale* circuits. In this chapter, we mainly focus on high performance digital circuits in FDSOI and bulk in order to gain deeper insight in

- the temperature profiles of high speed digital circuits in bulk and FDSOI,
- the influence of scaling on self-heating effects in advanced technologies and
- the most critical group of devices in terms of long-term reliability.

3.1 Introduction

Probably, the most critical reliability issue of deep sub-micron high performance digital circuits is the electromigration [44]. Electromigration is the phenomenon where the interconnects fail due to an open or short resulting from the gradual movement of the conducting atoms under very large current densities. According to Black's equation [43], the mean time to failure, MTF, is determined by

$$MTF = \frac{1}{AJ^2} e^{\left(\frac{\phi}{k_B T}\right)} \tag{3.1}$$

where T is the temperature of the interconnect, J is the current density inside the interconnect, A is a constant which contains a factor involving the cross-sectional area of the interconnect, ϕ is the activation energy and k_B is the Boltzmann constant. (3.1) indicates that the interconnect that is responsible of the failure should have a combination of the highest temperature and the largest current density. Firstly, the temperature dependency of MTF indicates that the interconnect should be located close to a spot with very large heating. Secondly, due to the current density dependency of MTF, the interconnect should have a small cross section area. In Chapter 2, we have shown that the generated heat in FDSOI cannot easily diffuse into the substrate due to the large thermal resistance created by the buried oxide. In the absence of a direct thermal contact to the highly conductive substrate, the generated heat seeks other alternative paths with relatively small thermal resistance. In this situation, the heat diffuses out through the drain and the source terminals due to their direct connection to the highly conductive metal contacts. This primarily increases the temperature of the contacts rather than the higher level metals. Moreover, due to their small cross section, their current density would also be larger than other interconnects. Therefore, focusing on contacts could be the first step of finding the most likely point of electromigration. In Chapter 2, we have shown that the point where the largest heating occurs in short channel devices is close to the drain end of the device. Because of this, the temperature of a drain contact in FDSOI is significantly larger than the temperature of same device's source contact. Therefore, the failure of an IC in FDSOI will be probably due to electromigration occurring in the drain contact of a device that has the largest heating and current density combination. To prevent that, it is necessary to find the most critical devices and modify their design in order to limit their excessive power dissipation and flatten the overall heating. To do this, one needs to detect the most critical group of devices in the entire circuit. However, in today's technologies, it is possible to integrate billions of devices in an IC. Therefore, a detailed heating analysis is a very demanding task. Nevertheless, the failure will probably happen in a high performance time critical block with heavy load. Even in a high performance block, only some of the devices will be the most critical ones depending on their function and activity rate. Hence, by understanding the correlation between the amount of heating and the specific functions of different devices, one can reduce the search zone significantly.

In our work, we have intended to understand which group of devices are the most critical ones in terms of electromigration probability in high performance digital circuits. For that reason, we have decided to focus on parallel prefix adders since they are one of the time critical blocks of high throughput digital circuits. In order to understand also the effect of scaling on self-heating and hot-spot temperatures, we have designed and implemented two 64-bit parallel prefix adders in commercially available 40 nm and 28 nm technologies [34]. For obtaining the heat generation and temperature maps, thermal simulations are performed on the two blocks by using the HotSpot tool [84]. During the simulations, both FDSOI and bulk geometries are used separately in order to see the significance of FDSOI in terms of self-heating effects. After the thermal simulations, the devices situated on the hot-spot locations are found and examined. It is observed that self-heating in FDSOI is much more prominent compared to bulk. The local hot-spots in FDSOI have sizes comparable to the size of the devices and their temperature is by far larger than bulk since the generated heat is directly translated into temperature. Consequently, the highest temperature values occur on the devices that have the highest power density. We have categorized each device according to their function and observed the heating values of each category. It has been shown that some group of devices that perform the same function are the most prominent heat generators [85]. Finally, a solution for decreasing the temperature of the hot-spots is proposed. It is shown that the peak temperature of the design in FDSOI can be decreased significantly with a cost of an insignificant increase in the area and parasitic capacitances.

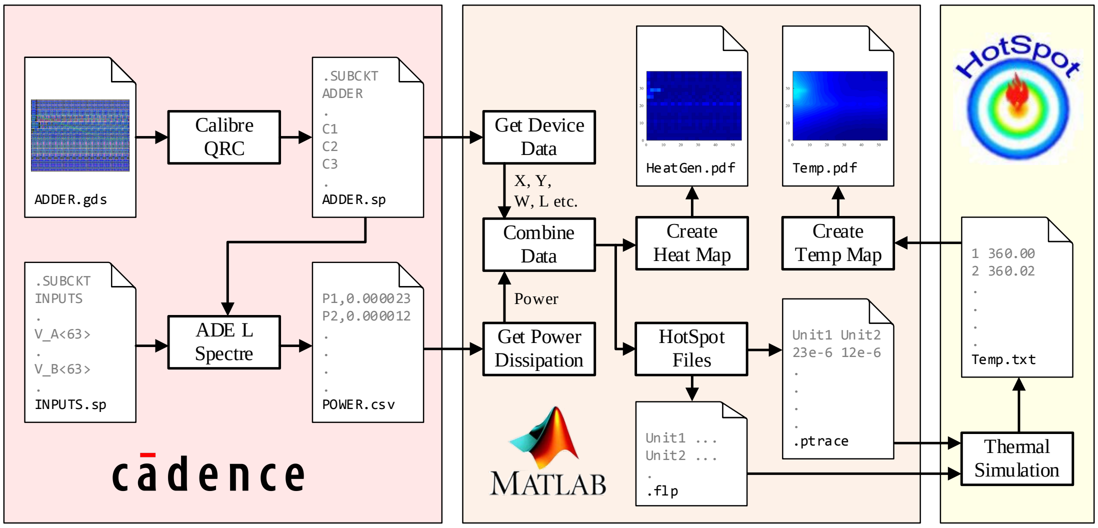


Figure 3.1 – Thermal analysis flow and data exchange between the used tools.

This chapter is organized as follows. In Section 3.2, the method and the tolls that are used in our block level thermal simulations are described. In Section 3.3, the architecture and the performance parameters of the implemented 64-bit parallel prefix adders are provided. In Section 3.4, bulk and FDSOI thermal simulation results and the temperature profiles of the designed 64-bit adders are shown. In Section 3.5, different groups of devices and the heating values of each group are given. Finally, the chapter is closed with the summary of the work and the conclusions in Section 3.6.

3.2 Method for Block Level Thermal Simulations

Detailed thermal analyses are performed on the implemented blocks by considering the thermal properties of both bulk and FDSOI device geometries for different cases. For that, Cadence, Matlab and HotSpot tools are used. The details and the data exchange between these tools are summarized on Figure 3.1. Firstly, electrical simulations are performed in Cadence Spectre circuit simulator by applying series of randomly generated input vectors to the inputs of the net-list in which the extracted interconnect parasitics are included. That way, the power dissipation data of each device in the entire adder and their corresponding coordinates on the layout are obtained. These data are used in Matlab to create **heat maps** for observing the spatial heating of the blocks. The thermal simulations are performed in HotSpot thermal simulator. For that, detailed **thermal models** with different layers are introduced for bulk and FDSOI geometries of 40 nm and 28 nm technologies. The **heating input**[1] of the thermal simulations are generated in Matlab and transferred to HotSpot. Finally, resulting temperature data of HotSpot is processed in Matlab to extract **temperature maps**; which is followed by

---

[1] A **heat map** refers to the two dimensional colour plots that show the heating information of a block; whereas, a **heating input** refers to the heat generation input of a thermal simulation with the same heating information and spatial distribution. Hereby, the terms might be interchangeably used throughout the text.

some other manipulations in order to obtain additional information such as temperature gradients, locations of hot-spots, maximum and minimum temperature values etc.

3.2.1 Creation of Heat Maps

During the thermal simulations of the 64-bit adder blocks that are implemented in 40 nm and 28 nm technologies, it is assumed that the only heat generators are the nMOS and the pMOS devices. The heating of the interconnect lines are ignored, although they contribute to the heat generation due to their finite conductance. This assumption is acceptable when the length of the interconnect is not longer than a certain amount [50], which is the case when the dimensions of examined test blocks are considered for the used technology nodes. To create the spatial heating data, the instantaneous power dissipation and heat generation[2] waveform of each device is extracted by

$$P_n(t) = I_{D,n}(t) \times V_{DS,n}(t) = \int\limits_{V_n} q(t)dv \tag{3.2}$$

where P_n is the instantaneous power dissipation (or heat generation) [W] of device n with an effective volume of V_n [m^3], $I_{D,n}$ is its drain current [A], $V_{DS,n}$ is the voltage drop between its drain and the source terminals [V] and q is the instantaneous heat generation at a point [W/m^3] according to Drift-Diffusion formulation (2.3). Having the instantaneous heat generation in hand, the average heat generation is calculated by

$$\overline{P_n} = \frac{1}{t_{sim}} \int_0^{t_{sim}} I_{D,n}(t) \times V_{DS,n}(t)dt \tag{3.3}$$

where $\overline{P_n}$ is the average heat generation [W] of device n throughout the simulation time t_{sim} [sec]. In addition to the average heating values of each device, their physical coordinates are extracted from the layout so that the horizontal and vertical coordinate locations (x_n and y_n) of each heat generator and their average heating values are combined together. The heat maps are created with a grid array rather than introducing each device separately. For that, the entire area of the simulated digital block is partitioned into squares for constructing a grid with a constant pitch size of X_p [μm] (i.e. spatial granularity of the simulation grid). Each square in the grid is defined by horizontal i and vertical j indices where i and j are integer numbers which are greater than zero. The set of i and j values are defined depending on the selected pitch value X_p and the dimensions (X and Y) of the digital block according to

$$I = \left\{ i \,\middle|\, i \in \mathbb{N} \wedge i \leq \left\lceil \frac{X}{X_p} \right\rceil \right\} \quad \text{and} \quad J = \left\{ j \,\middle|\, j \in \mathbb{N} \wedge j \leq \left\lceil \frac{X}{X_p} \right\rceil \right\} \tag{3.4}$$

where $\lceil\cdot\rceil$ is the rounding operation to the nearest integer number which is greater than or equal to its input. The heating values of the heat sources are added and assigned to the

---

[2] The **heat generation** or **heating** inside a volume is set equal to the **power dissipation** inside the same volume. Hereby, the terms might be interchangeably used throughout the text.

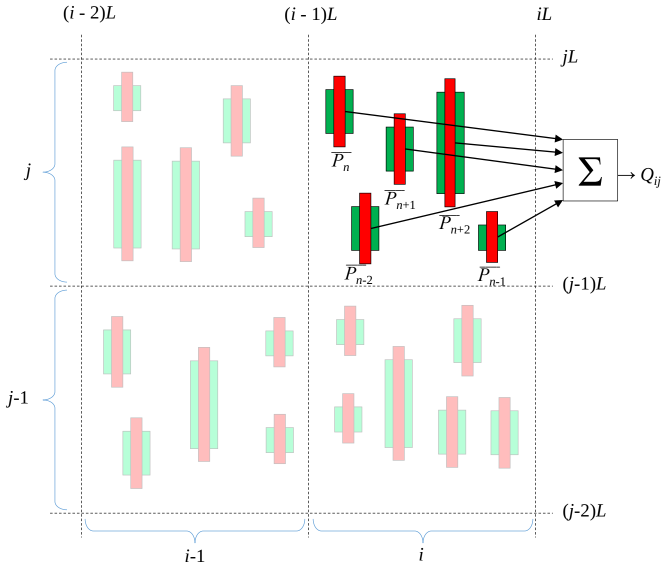


Figure 3.2 – Summation of the independent heat sources located in the same square.

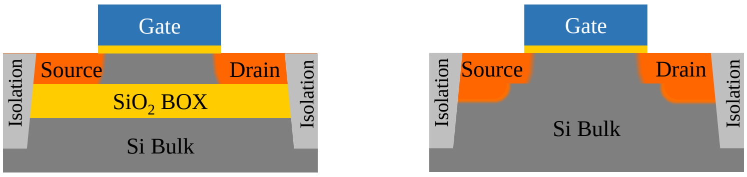
corresponding squares in the grid according to

$$\forall i \in I \wedge \forall j \in J \quad Q_{ij} = \frac{1}{X_p^2} \sum_{n=1}^{N} \overline{P_n} \quad \text{if} \quad i-1 \le \frac{x_n}{X_p} < i \wedge j-1 \le \frac{y_n}{X_p} < j \tag{3.5}$$

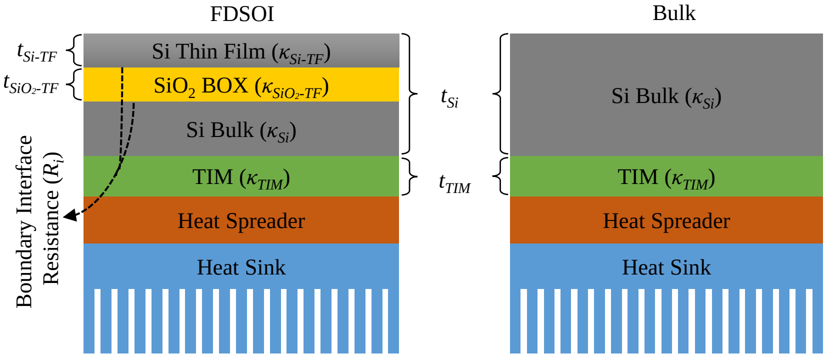
which is illustrated on Figure 3.2. In (3.5), Q_{ij} is the heat density [W/cm^2] inside the square which is between $(i-1)L$ and iL horizontally and between $(j-1)L$ and jL vertically. It is expected that the maximum heating, hence the maximum temperature value would decrease by choosing a larger pitch value. For observing that, different heat maps are created for the implemented adders by selecting different pitch values (X_p) from 0.5 µm to a sufficiently large value that is even greater than the X and Y dimensions of the block. Further details of the extracted heat maps will be given in Section 3.4.1

3.2.2 Thermal Model

For the thermal simulations, thermal models of bulk and FDSOI technologies are used to make a comparison between each other. For the 40 nm case, the thermal model of a commercially available bulk technology is used. On the other hand, for the 28 nm case, the thermal model of a commercially available FDSOI technology is used. To perform a fair comparison, the thermal and geometric parameters of the selected 28 nm FDSOI technology are scaled for 40 nm FDSOI case. The 28 nm bulk thermal model parameters are also calculated in the same way. The extracted thermal models for different layers of the chip are illustrated on Figure 3.3. The values of those parameters are shown on Table 3.1 for 28 nm and 40 nm bulk and FDSOI cases. The bulk silicon thickness is set to 150 µm for all cases. The thermal conductance of bulk Si is 148 Wm^{-1}K^{-1}, which can be considered as a good value in terms of heat removal. According



(a) Materials with different thermal properties in FDSOI and bulk devices.



(b) Thermal models of different layers for FDSOI and bulk.

Figure 3.3 – Thermal models of FDSOI (left) and bulk (right).

Table 3.1 – Thermal properties of FDSOI and bulk in 40 nm and 28 nm technologies.

Technology Parameter	**28-40 nm** **Bulk**	**40 nm** **FDSOI**	**28 nm** **FDSOI**	**Unit**
t_{Si-TF}	-	10	7	nm
t_{SiO_2-TF}	-	50	25	nm
t_{Si}	150	150	150	µm
t_{TIM}	20	20	20	µm
κ_{Si-TF}	-	12.5	10	W/m-K
κ_{SiO_2-TF}	-	0.8	0.6	W/m-K
κ_{Si}	148	148	148	W/m-K
κ_{TIM}	4	4	4	W/m-K
R_i	-	2×10^{-8}	2×10^{-8}	m^2-K/W

to the selected 28 nm FDSOI technology, the Si thin film and the BOX thicknesses are 7 nm and 25 nm respectively. These values are scaled to 10 nm and 50 nm for the modelled 40 nm FDSOI case. The thermal conductance of the Si and SiO_2 thin film structures are less than their bulk thermal conductance values due to phonon boundary scattering [86]. Therefore, these two layers are independently modelled in FDSOI. The thermal conductance of the SiO_2 film with 25 nm and 50 nm thicknesses are approximately equal to 0.6 W/m-K and 0.8 W/m-K [87, 88]. These values are more than two orders of magnitude less than the thermal conductance of bulk Si. The thermal conductance of 10 nm and 7 nm-thick Si thin films are calculated as 12.5 Wm^{-1}K^{-1} and 10.0 Wm^{-1}K^{-1} according to (2.17). The details about the correlation between the film thickness and the thermal conductivity is given in Chapter 2. The calculated thermal conductance values are more than an order of magnitude less than the thermal conductance of bulk Si. The boundary between Si and SiO_2 creates a temperature jump, hence a finite interface thermal conductance [35, 36]. The value of interface thermal resistance (R_i) is set to 2×10^{-8} m^2KW^{-1} [87, 88]. The resulting interface thermal resistance is equivalent to the thermal resistance of a 20 nm thick SiO_2 film. All of the mentioned modifications increases the thermal resistance at the bottom surface significantly in FDSOI. The carrier transport in FDSOI transistors occurs inside the Si Thin Film layer; therefore, the heat sources are placed in this layer according to their coordinate information extracted from the layout. For the bulk geometry simulations, the heat sources are placed on the top surface of the Si Bulk layer. It can be seen from Figure 3.3 that the metal routing layers and the dielectric isolation layer is ignored. Extracting the thermal model of this layer is very difficult since the distribution of the routing layers are quite complex and not homogeneous. However, the average thermal conductance of the metal-dielectric layer is quite low compared to the bottom heat conduction path (through heat spreader and heat sink). Moreover, the heat conduction of the metal-dielectric layer is even less in deep sub-micron technologies due to the low thermal conductance of the low-κ materials [46]. Therefore, the upper part of the chip is assumed to be thermally insulator. Finally, the SiO_2 trenches (shown as isolation on Figure 3.3a) are not included in the models. The exclusion of the trenches would not change the temperature profile in bulk significantly since the bottom heat removal path provides a high thermal conductance thanks to bulk Si. However, it is expected that in FDSOI one would observe larger temperature gradients and peak temperature values with the trench isolations due to the complete confinement of the device by SiO_2.

3.2.3 Creation of Temperature Maps

During the thermal simulations in HotSpot, the adders are inserted in the center of a square with an area of 64×64 μm^2. It is assumed that the blocks are surrounded by digital gates with similar power consumption. Therefore, the empty space in the thermal simulation area is filled with dummy heat generators with the same heat density of the adders. The simulations are performed with a thermal simulation resolution of 2048×2048, which should not be confused with the heat map resolution. Finally, the adder part of the temperature output obtained from the thermal simulation is extracted in Matlab and illustrated by colour maps. The extracted temperature maps are shown in Section 3.4.2.

3.3 Design of 64-bit Parallel Prefix Adders for Thermal Analysis

As it has been already explained, to perform a self-heating analysis, two adders are implemented in 40 nm and 28 nm technologies. During the implementation of the 64-bit adders, the main design goal is set as to obtain the highest clock frequency so that the extreme heating cases can be experimentally seen. For that, parallel prefix adders are preferred since they can provide the smallest critical path delay.

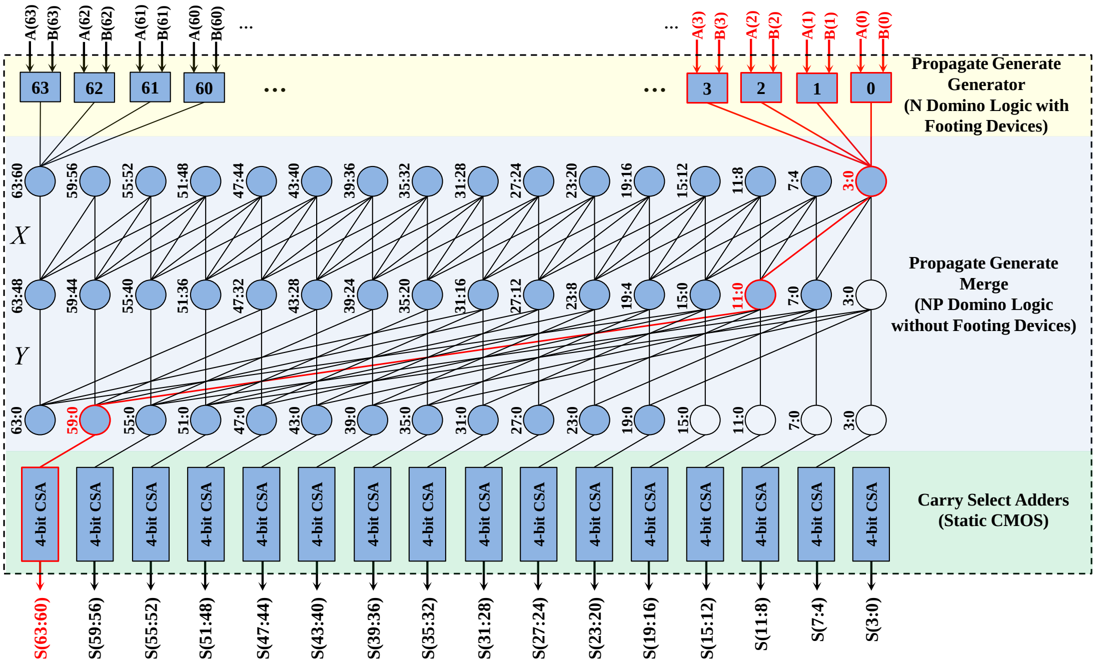


Figure 3.4 – The architectural block diagram of the implemented 64-bit Kogge-Stone parallel prefix adders with the radix-4 and sparsity-4 options. The input/output nodes, signal names and the hardware blocks on the critical path are indicated with red colour.

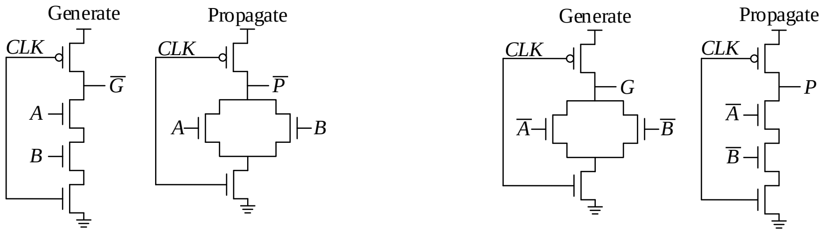
3.3.1 Architecture

The adders in both 40 nm and 28 nm technologies are implemented with Kogge-Stone technique [89] mostly in NP-domino logic. After examining and simulating different radix and sparsity options, it is observed that radix-4 [90, 91, 92] and sparsity-4 [93] options can reach to lowest critical path delay values with a reasonable area. Hence, these options are selected in the implementations. The detailed block diagram of the implemented 64-bit Kogge-Stone parallel prefix adders with radix-4 and sparsity-4 options can be seen on Figure 3.4. The interconnect lines, signal names and the blocks on the critical path are indicated by red colour. The block consists of three main bigger blocks which are *Propagate-Generate Signal Generator*, *Propagate-Generate Signal Merge* and *4-bit Carry Select Adders (CSA)*. The detailed explanation of the architecture of these blocks are given in the following sections.

Propagate-Generate Signal Generator (PG-Generator)

In parallel prefix adders, the reduction in the delay time is provided by merging the *Propagate* (P) and the *Generate* (G) signals in a parallel fashion to obtain the values of the carry-out signals. For that reason, the P and G signals are generated before the merging step according the Boolean expressions (top rectangle blocks on Figure 3.4) given by

$$P_i = \overline{\overline{A_i + B_i}} \tag{3.6}$$



(a) 40 nm

(b) 28 nm

Figure 3.5 – Schematics of N-domino logic PG-Generator gates in (a) 40 nm and (b) 28 nm. The devices at the bottom are the clocked footing devices [1].

$$G_i = \overline{\overline{A_i \cdot B_i}} \tag{3.7}$$

where the subscript i is the bit number of the A and B inputs. In (3.6) and (3.7) the upper bar (i.e. negation) is implemented by cascading an inverter at the output of a NAND or NOR gate. This is a drawback since it brings an additional gate in the data path and might possibly increase the critical path delay. For that reason, it is preferred to use negated signals either at the input ($\overline{A}$ and $\overline{B}$ instead of A and B) or at the output ($\overline{P}$ and $\overline{G}$ instead of P and G) [94]. In the 40 nm implementation, $\overline{P}$ and $\overline{G}$ are generated and the logic tree[3] is implemented with nMOS devices (Figure 3.5a). On the other hand, for the 28 nm implementation, $\overline{A}$ and $\overline{B}$ are provided at the input and P and G signals are generated at the output (Figure 3.5b). The $\overline{A}$ and $\overline{B}$ signals are generated in the previous cloud of the pipeline so that the speed can be further increased. The reason for having $\overline{A}$ and $\overline{B}$ instead of A and B in 28 nm implementation will be clarified in the following section.

Propagate-Generate Signal Merge (PG-Merge)

This block is the heart of the 64-bit adder implementation since the evaluation time of this part has an important influence on the speed of the overall implementation. In this block, P and G signals are merged to get the *Carry Out* (C_o) information of different stages in the addition. On Figure 3.4, the blue coloured circles indicate a logic gate which performs the merging operation of four P and four G signals, where the radix option is set to four for further decreasing the critical path delay by decreasing the logic depth [90, 95]. All radix-4 PG-Merge gates in the entire PG-Merge block are implemented with NP-domino logic gates. Since the logic gates in the PG-Generator blocks are implemented with N-type logic tree, the implementation is proceeded with P-type domino gates to prevent the irreversible discharge problem during evaluation. No clocked footing device is used since there is no possible current path on the evaluation trees of the logic gates [96]. The fact that the clocked footing devices are removed increases the rise and fall times of the gates and decreases the area of the gates significantly.

---

[3] In NP-domino logic, **logic tree** contains the devices that are responsible of implementing a specified Boolean function and it excludes the devices such as pre-charge and footing devices. It is also called as **evaluation tree** throughout the text.

The Boolean expression of these functions are shown by

$$P_{i:i-3} = P_i \cdot P_{i-1} \cdot P_{i-2} \cdot P_{i-3} \tag{3.8}$$

$$G_{i:i-3} = (G_i + G_{i-1} \cdot P_i) + (G_{i-2} + G_{i-3} \cdot P_{i-2}) \cdot P_i \cdot P_{i-1} \tag{3.9}$$

where the subscript $i : i-3$ indicates that the output signals are the merged P and G signals from the bits i to $i-3$. Radix-4 option provides the advantage of merging four signals with a single logic gate. However, (3.8) and (3.9) shows that the implemented CMOS logic gate will be quite complex and it will contain four transistors in series. This fact will in turn decrease the speed of the logic gate especially for the advanced technology nodes where the power supply voltages are equal to or below 1 V. Another possibility to implement a radix-4 PG-Merge gate is to cascade two radix-2 PG-Merge gates [97]. For that, (3.8) and (3.9) can be written as

$$\begin{aligned} P_{i:i-3} &= (P_i \cdot P_{i-1}) \cdot (P_{i-2} \cdot P_{i-3}) \\ &= P_{i:i-1} \cdot P_{i-2:i-3} \end{aligned} \tag{3.10}$$

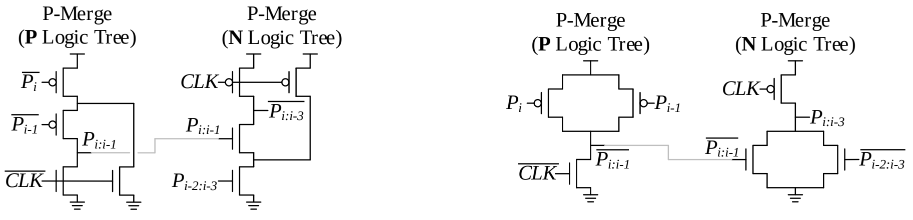
$$\begin{aligned} G_{i:i-3} &= [G_i + G_{i-1} \cdot P_i] + [(G_{i-2} + G_{i-3} \cdot P_{i-2}) \cdot (P_i \cdot P_{i-1})] \\ &= G_{i:i-1} + G_{i-2:i-3} \cdot P_{i:i-1} \end{aligned} \tag{3.11}$$

where the cascading of two radix-2 PG-Merge gates can be explicitly seen. The cascading approach has the disadvantage of having a logic depth of two compared to the approach shown by (3.8) and (3.9); however, the series resistance in each gate is quite relaxed. At this point, the delay performance can be questioned since both approaches have their own advantages and disadvantages. To observe the faster solution, the 64-bit adder is designed and simulated with both architectures. It is observed that the cascading approach is unequivocally faster than the single gate approach. As indicated in Section 3.3.1, only the $\overline{P_i}$ and $\overline{G_i}$ signals are available at the inputs of the *Propagate-Generate Signal Merge* block for the 40 nm implementation. Consequently, (3.8) and (3.9) are modified so that both the inputs and the outputs are negated according to

$$\begin{aligned} \overline{P_{i:i-3}} &= \overline{\overline{\left(\overline{P_i} + \overline{P_{i-1}}\right)} \cdot \overline{\left(\overline{P_{i-2}} + \overline{P_{i-3}}\right)}} \\ &= \overline{P_{i:i-1} \cdot P_{i-2:i-3}} \end{aligned} \tag{3.12}$$

$$\begin{aligned} \overline{G_{i:i-3}} &= \overline{\overline{\overline{G_i} \cdot \left(\overline{G_{i-1}} + \overline{P_i}\right)} + \left[\overline{\overline{G_{i-2}} \cdot \left(\overline{G_{i-3}} + \overline{P_{i-2}}\right)} \cdot \overline{\left(\overline{P_i} + \overline{P_{i-1}}\right)}\right]} \\ &= \overline{G_{i:i-1} + G_{i-2:i-3} \cdot P_{i:i-1}}. \end{aligned} \tag{3.13}$$

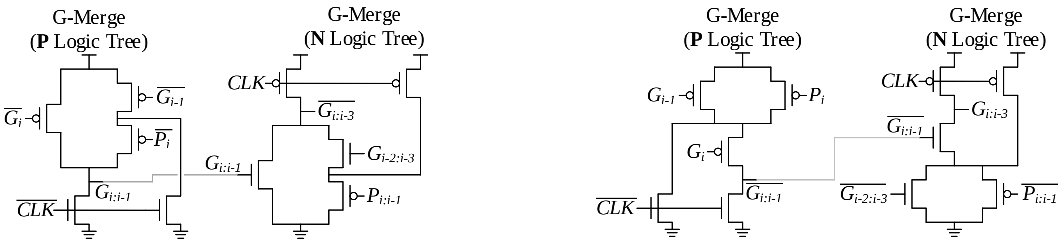
The domino logic implementation of the Boolean functions (3.12) and (3.13) can be seen on Figure 3.6a and Figure 3.7a respectively. The entire PG-Merge block is implemented by using these architectures in 40 nm. The number and type of the transistors on the critical path



(a) 40 nm

(b) 28 nm

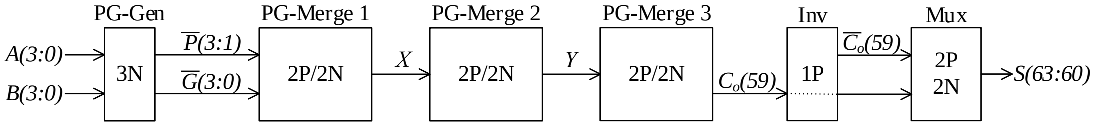
Figure 3.6 – Schematics of NP-domino logic P-Merge gates used in (a) 40 nm and (b) 28 nm.



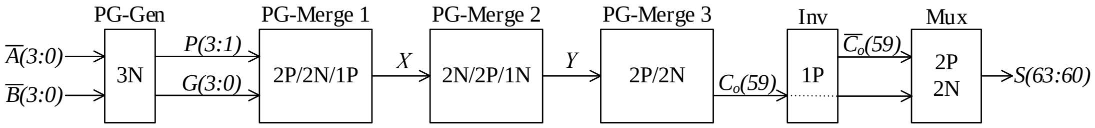
(a) 40 nm

(b) 28 nm

Figure 3.7 – Schematics of NP-domino logic G-Merge gates used in (a) 40 nm and (b) 28 nm.



(a) 40 nm



(b) 28 nm

Figure 3.8 – The number and type of the devices on the critical path of (a) the 40 nm and (b) the 28 nm adder.

from input bits to carry-out signals are summarised on Figure 3.8a. The node with the largest capacitive load on the data path is indicated by X and Y. The reason for having a very large capacitive load is because the routing on X and Y are horizontally very long and its fan-out is four. This situation can be also seen on Figure 3.4. For having a stronger drive at these points, the gates which drive X and Y nodes are implemented with N logic tree, where the output is pre-charged with pMOS devices. However, it is observed that there is a large parasitic coupling between the closely adjacent routing lines. Unless they are evaluated, all the outputs are floating during evaluate phase. Therefore, any evaluation at the adjacent interconnect lines disturb the voltage levels of the floating outputs that stay pre-charged. In 28 nm, it is observed that the disturbance is more significant. One solution for this is to increase the distance between the interconnect lines. However, this brings a complication in the routing since there are many parallel lines especially for Y. Another solution is to use inverters for the drive of the output nodes that are strongly coupled to each other. The pre-charge device of these inverters are driven by the output of the previous gate rather than the clock [98, 99] to provide continuous drive; hence they are static. In 28 nm implementation, additional static inverters are placed for the outputs at X and Y. This modification increases the logic depth by two (Figure 3.8b) in addition to larger power consumption. Moreover, either X or Y has to be evaluated by a P logic tree. Since Y contains a significantly larger capacitive load than X, it is chosen to be evaluated by an N logic tree and X is evaluated by a P network, which can be seen on Figure 3.8b. Another improvement in 28 nm is made by simplifying the implementation of the P-Merge gates. It is done by using Boolean functions of (3.14) and (3.15) rather than (3.12) and (3.13).

$$\begin{aligned} P_{i:i-3} &= \overline{\overline{P_i \cdot P_{i-1}} + \overline{P_{i-2} \cdot P_{i-3}}} \\ &= \overline{\overline{P_{i:i-1}} + \overline{P_{i-2:i-3}}} \end{aligned} \tag{3.14}$$

$$\begin{aligned} G_{i:i-3} =& \overline{\overline{G_i + G_{i-1} \cdot P_i} \cdot \left(\overline{G_{i-2} + G_{i-3} \cdot P_{i-2}} + \overline{P_i \cdot P_{i-1}}\right)} \\ =& \overline{G_{i:i-1} \cdot (G_{i-2:i-3} + P_{i:i-1})} \end{aligned} \tag{3.15}$$

That way, it is possible to implement the evaluation tree of the P-Merge gates with two parallel devices rather than two devices in series (i.e. no stacking, (Figure 3.6b) and remove the pre-charge device for the intermediate node. Although this decreases the area and complexity, it does not increase the speed directly since the G-Merge implementation contains two series devices in the evaluation tree (Figure 3.6b). This is the reason why $\overline{A}$ and $\overline{B}$ signals are mandatory at the input of the PG-Generator block.

Carry Select Adders (CSA)

To obtain the 64-bit sum value, CSAs are needed since the PG-Merge block generates only the 64-bit C_o signal. Moreover, each carry select adder has to be a 4-bit implementation (Figure 3.9) because the PG-Merge block is implemented with sparsity-4 option where only every forth C_o signal is generated [92]. The 4-bit CSA is implemented with two 4-bit ripple carry adders (RCA). In the 40 nm implementation, P and G are used as the inputs of 1-bit full adders instead of A and B input bits. This way, it is possible to decrease the number of the series transistors from three to two in the 1-bit full adder gates; however, the loading of the PG-Generator gates increase as a trade-off. For that, in 28 nm implementation A and B are used at the input of the CSA, which increases the size of CSA significantly. The implementation of the *Carry Select Adder* block is performed by static CMOS since it is not time critical.

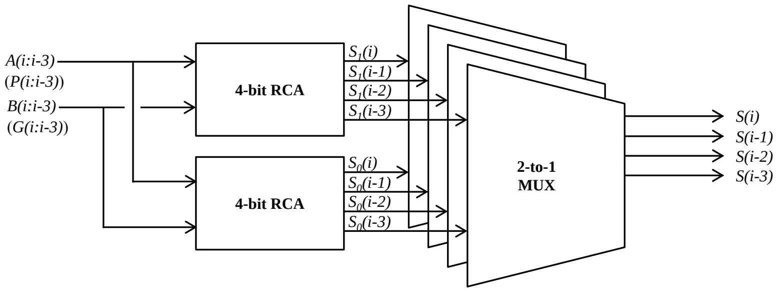
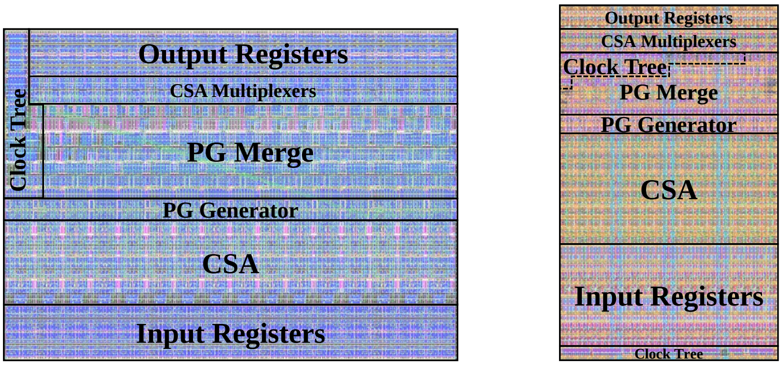


Figure 3.9 – The block diagram of the 4-bit CSA. The $S_1(i)$ and $S_0(i)$ signals at the outputs of the two 4-bit ripple carry adders are the sum signals which are generated for the two cases where C_i is equal to 1 and 0 respectively.



(a) 40 nm

(b) 28 nm

Figure 3.10 – The layouts and floorplans of the implemented 64-bit adders in (a) 40 nm and (b) 28 nm. Thermal simulations are performed on these areas.

Table 3.2 – Comparison of the implemented adders in 40 nm and 28 nm technologies.

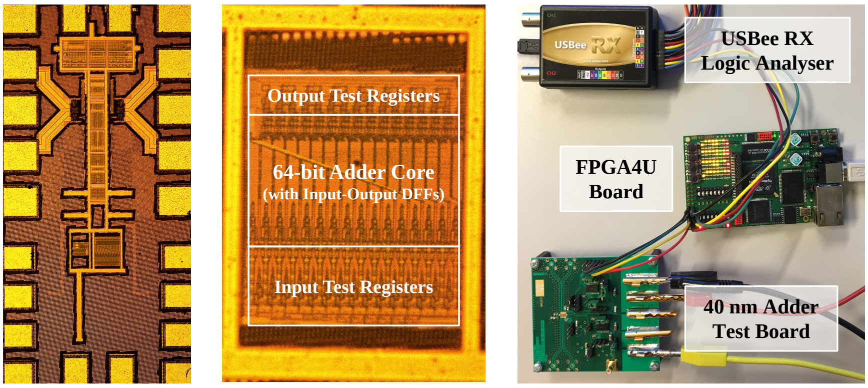
Parameters	**40 nm**	**28 nm**	**Unit**
X-dimension	54.72	26.34	µm
Y-dimension	39.75	43.12	µm
Area	2175	1136	μm^2
Device Count	10922	10989	-
Logic Depth	9	11	-
# Devices on Crit. Path	18 (9N/9P)	20 (10N/10P)	-
Clock Frequency	3.3	4.55	GHz
Evaluation Time	151	110	ps
Supply Voltage	0.9	1.0	mW
Average Power Consumption	15.92	14.57	mW
Average Power Density	732	1283	W/cm^2
Energy per 64-bit Addition	4.82	3.21	pJ

3.3.2 Performance

The entire 64-bit Kogge-Stone adder block is designed with a full-custom design approach in both technology nodes. The blocks are primarily optimized to obtain the lowest possible

Table 3.3 – Comparison of the published 64-bit adders. †The adder in [3] is a 32-bit implementation.

Work	Year	Tech [nm]	V_{DD} [V]	Delay [ps]	Power [mW]	PDP [pJ]	Area [µm^2]	X [µm]	Y [µm]
This	2017	28	1.0	95	14.57	1.38	1’136	26.3	43.1
This	2015	40	0.9	148	15.92	2.36	2’175	54.7	39.8
[100]	2012	65	1.0	148	135	19.98	10’800	120	90
[101]	2012	90	1.0	247	96.8	23.86	9’660	46	210
[102, 103]	2011	90	1.0	181	840	152.04	148’750	875	170
[104]	2009	90	1.0	240	260	62.40	31’275	417	75
[105]	2009	90	1.0	417	42	17.51	NA	NA	NA
[106]	2007	65	1.2	154	54	8.32	12’236	322	38
[107]	2005	90	1.3	250	300	75.00	72’800	280	260
[108]	2005	180	2.5	668	NA	NA	129’536	176	736
[3]†	2002	130	0.95	200	95	19.00	28’224	336	84



(a) Micrograph (b) Adder in 40 nm (c) Test setup

Figure 3.11 – (a) The die micrograph, (b) zoomed micrograph and (c) the test setup of the implemented 64-bit adder in 40 nm. Thermal simulations are performed on the 64-bit adder core block.

critical path delay while having the lowest possible power consumption and area. For that purpose, the devices are iteratively sized by simulating the blocks with extracted interconnect parasitics. Figure 3.10 shows the layouts of the core part and Table 3.2 summarizes the performance of the implemented adders in 40 nm and in 28 nm technologies and Table 3.3 shows the performance of other published parallel prefix adder implementations

The functionality of the 40 nm adder is verified with electrical measurements (Figure 3.11c). The communication with the adder is provided by FPGA4U [109] board which contains an Altera Cyclone II EP2C20 [110] FPGA. The analog and digital waveforms are also monitored with USBee RX [111] mixed signal oscilloscope and logic analyser. Some thermal images are taken on the die with FLIR Systems SC655 thermal camera. The infrared image on Figure 3.12b

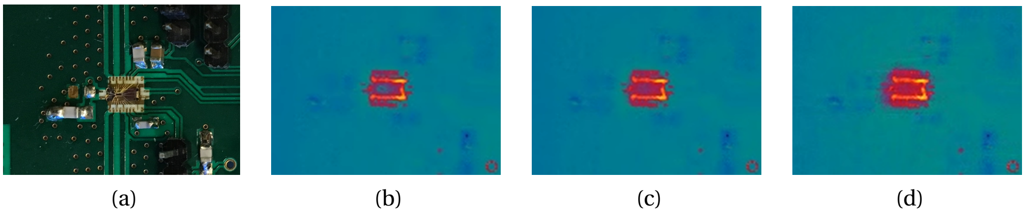


Figure 3.12 – Thermal images of the 40 nm adder taken during tests. (a) the visible image of the board where the thermal images are taken, (b) thermal image just before the addition begins, (c) thermal image just after the addition begins and (d) 20 seconds after the addition begins.

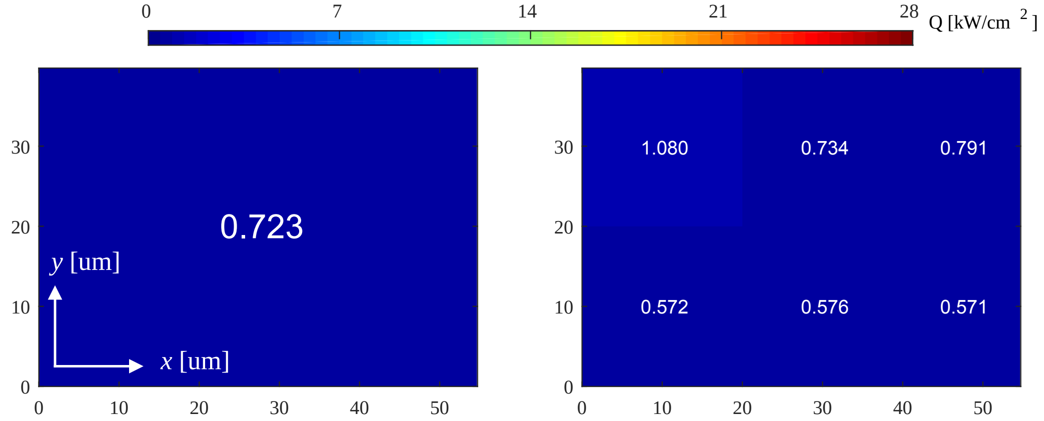
was taken just before the addition begins. The die can be detected due to its larger temperature compared to its surroundings (PCB and other components). The larger temperature is because of the control circuit and the clock generator. The image on Figure 3.12c was taken just after the addition begins. It can be seen that the die temperature is larger but the surroundings are still at the same temperature. After 20 seconds (Figure 3.12c) the die temperature is even larger due to continuous operation. In addition to this, the board is also hotter compared to Figure 3.12b and Figure 3.12c.

3.4 Comparison of Bulk and FDSOI in 40 nm and 28 nm Nodes

The circuit level design of adders are followed by their thermal analysis. In this section, heat and temperature maps of the adders are studied in detail, from top level down to nanometer scale where hot-spots are observed.

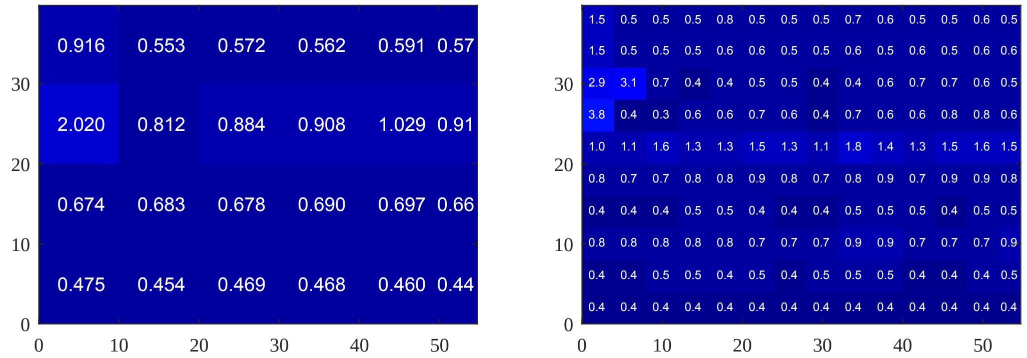
3.4.1 Heating

For the two adder implementations in 40 nm and 28 nm technologies, heat maps are created with different pitch values to observe the effect of grid size on (a) maximum and minimum heat density values, (b) maximum and minimum temperature values, (c) thermal simulation time and it is aimed to find the optimum pitch value for FDSOI and bulk geometry thermal simulations. The smallest pitch value is chosen as 0.5 µm while the largest pitch is larger than the size of the adders so that it covers the entire area. The heat maps can be observed on Figure 3.13 and Figure 3.14 for the 40 nm and the 28 nm implementations respectively. In the 40 nm implementation, for the lowest resolution case (Figure 3.13a, X_p = 60 µm), the maximum and the minimum heat density is equal to each other since there is a single square in the grid array. Therefore, the heat density of the entire area is 723.4 W/cm^2 and it is equal to the average power density of the entire block. The same observation can be made on Figure 3.14a where the heat density of the 28 nm implementation for the largest pitch value (50 µm) is equal to the average power density of the entire block, which is 1283 W/cm^2. This comparison also shows that the heat density of the adder in 28 nm is 77% larger than the one in 40 nm. This is primarily due to the increased clock frequency and the decreased area of the implementation.



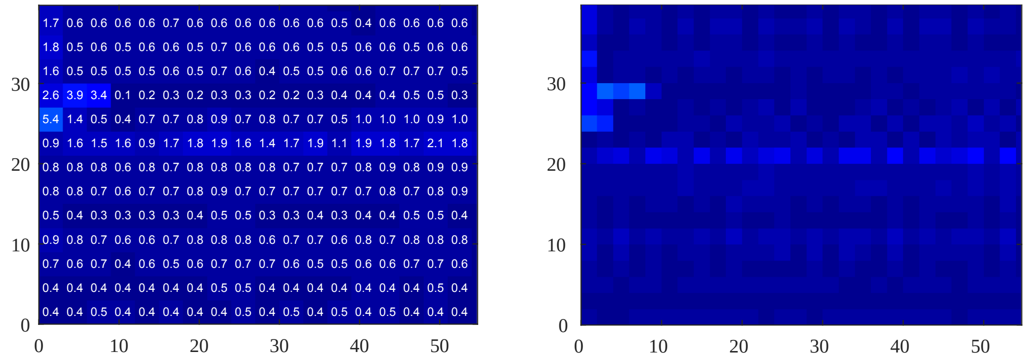
(a) $X_p = 60$ µm

(b) $X_p = 20$ µm



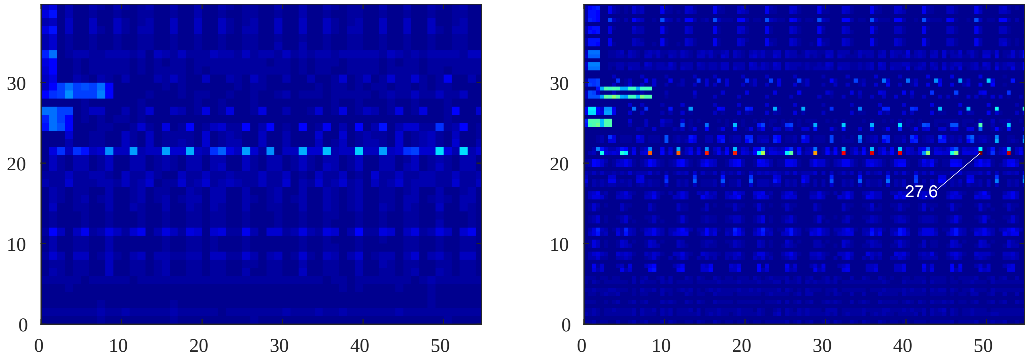
(c) $X_p = 10$ µm

(d) $X_p = 4$ µm



(e) $X_p = 3$ µm

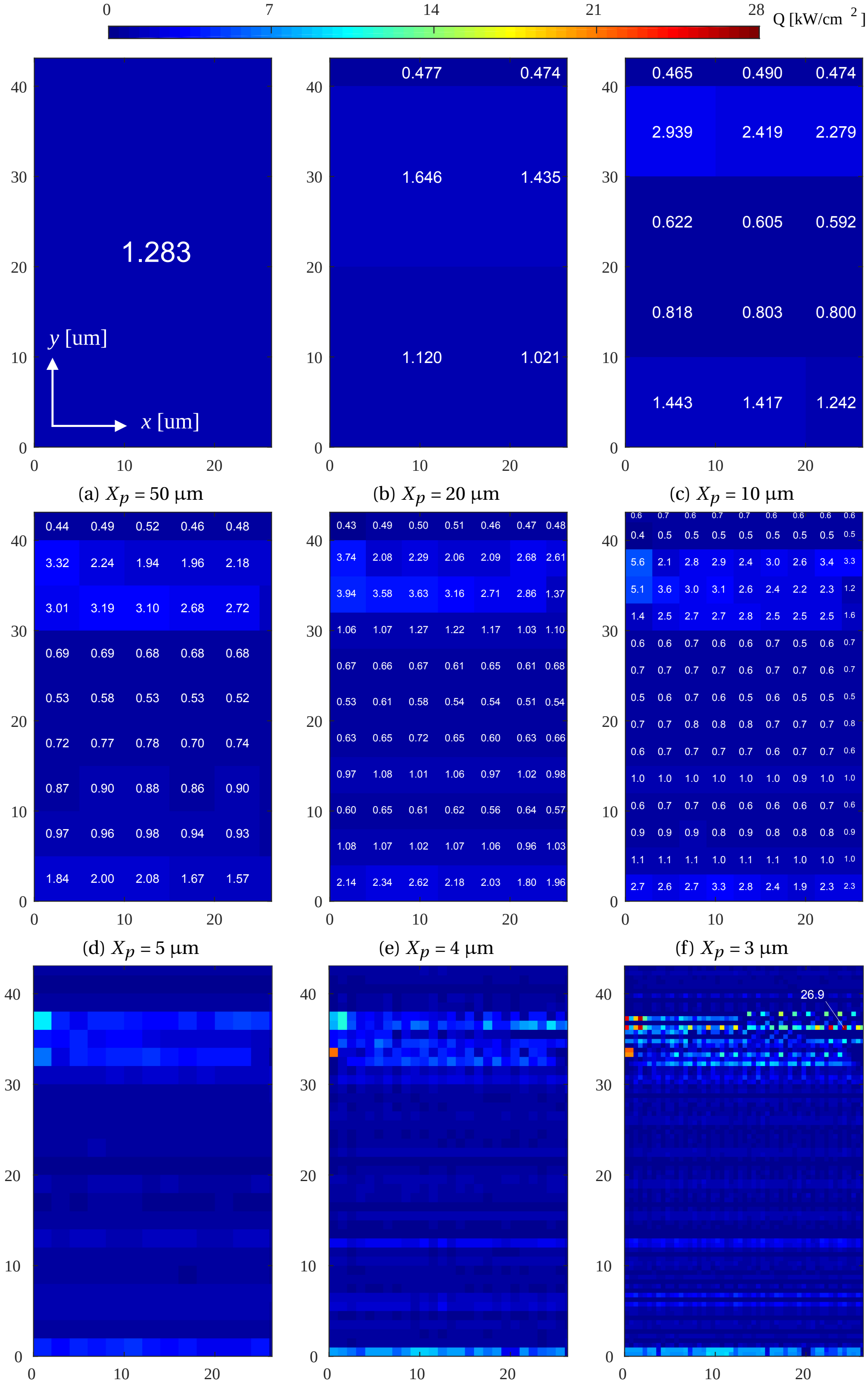
(f) $X_p = 2$ µm



(g) $X_p = 1$ µm

(h) $X_p = 0.5$ µm

Figure 3.13 – Heat maps of the 40 nm adder for different pitch values. Heat density unit is kW/cm^2.



(a) $X_p = 50$ μm

(b) $X_p = 20$ μm

(c) $X_p = 10$ μm

(d) $X_p = 5$ μm

(e) $X_p = 4$ μm

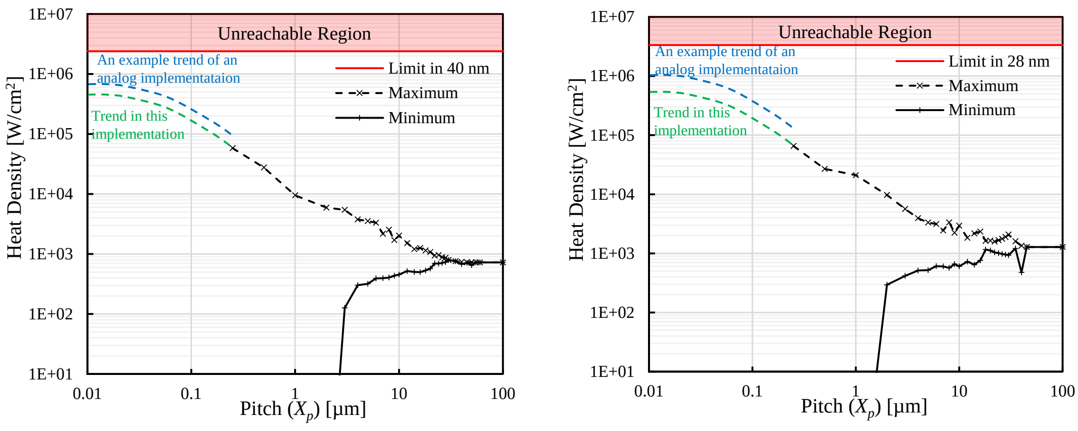
(f) $X_p = 3$ μm

(g) $X_p = 2$ μm

(h) $X_p = 1$ μm

(i) $X_p = 0.5$ μm

Figure 3.14 – Heat maps of the 28 nm adder for different pitch values. Heat density unit is kW/cm^2.



(a) 40 nm

(b) 28 nm

Figure 3.15 – Maximum and minimum heat density curves for heat maps with different resolutions. Maximum heat density saturates as the pitch value becomes comparable to the minimum length of the specified technology, whereas the minimum heat density drops to 0 W/cm^2 as resolution increases.

For smaller pitch values, different heat density values at different grid squares are observed. Maximum heat density in the overall area increases by decreasing the pitch value in both 40 nm and 28 nm implementations. Figure 3.15 shows this trend, which is quite similar for both technology nodes. For the smallest pitch value ($X_p = 0.5$ μm) on Figure 3.13h and 3.14i, the largest heat density is 27.6 kW/cm^2 and 26.9 kW/cm^2 respectively for 40 nm and 28 nm. These values are more than 20 times larger than the heat density values of the entire blocks. Moreover, by increasing the resolution, the existence of some other spots can be observed where the local heating is much less than the average heating. Besides, some of the grid points have 0 W/cm^2 heat density, mainly due to the inactive devices or absence of any devices. The larger differences between the maximum and minimum heat densities on the grid might translate into larger temperature gradients depending on the heat diffusion paths and the thermal conductance of the structure.

In deep sub-micron technologies, it is possible to place more than 10 transistors inside a square of 0.5 μm× 0.5 μm. Hence, it is anticipated that the largest heat density would further increase by decreasing the granularity of the simulation grid. In fact, choosing $X_p = 0.25$ μm results in a maximum heat density of 58.2 kW/cm^2 for 40 nm and 65.7 kW/cm^2 for 28 nm (the corresponding heat map is not shown on Figure 3.13 and Figure 3.14 due to the small sizes of the grid squares), which are already more than two times larger than the case where $X_p =$ 0.5 μm. However, the increasing trend in maximum heat density stops once the pitch value is less than half of the minimum gate length value of the specified technology ($L < \lambda$). In other words, it has to be guaranteed that the device with the highest heat density has to cover at least one grid square completely. Consequently, the largest heat density observed on a grid square saturates at a constant value for smaller pitch values less than 20 nm for 40 nm technology and 14 nm for 28 nm technology. This trend is shown with the green dashed fictitious curves on Figure 3.15. For the specific adders implemented in this study, the largest heat density is measured as 406.8 kW/cm^2 for 40 nm 554.3 kW/cm^2 for 28 nm as the pitch value is set smaller than half gate length. It can be seen that these values are roughly an order of magnitude greater than the ones obtained for $X_p = 0.25$ μm. Nevertheless, they are not necessarily the largest achievable heat density of the analysed technology nodes. The largest achievable value of the heat density can be obtained once a minimum gate length nMOS device is operating in its full-conducting state (i.e. $V_{GS} = V_{DS} = V_{DD}$). This value is measured as 2.4 MW/cm^2 for 40 nm and 3.35 MW/cm^2 for 28 nm technologies, which are almost another order of magnitude greater than the largest heat density of the implemented adders. Nevertheless, such high heat density values are almost never encountered in digital circuits since the power consumption is dynamic in CMOS logic gates. However, the local heat density might get closer to this value in analog circuits where the blocks are biased by constant current sources. Therefore, it might be necessary to decrease the pitch of the heat map down to the transistor gate length in the thermal analysis of analog circuits. At this point, one should be careful while assigning heating values to their corresponding coordinates by applying (3.5). For the pitch values lower than the minimum gate length, a single transistor occupies more than a single square in the grid. Therefore, the transistor should be partitioned and its total heating should be distributed among the grid squares that is occupied by the device.

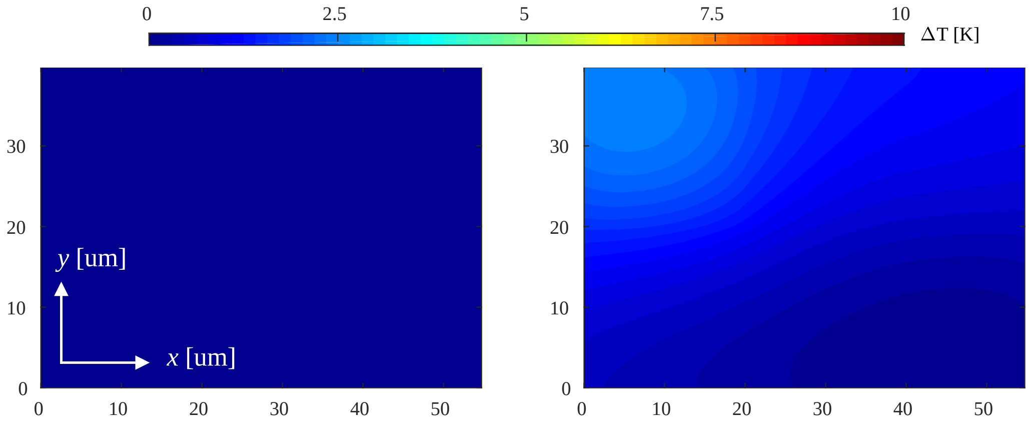
This analysis clearly indicates that performing a heating analysis (or creating heat maps) of a chip with low grid resolution filters out the spots where the heat density is either much higher or much lower than the other spots. According to Fourier's law of thermal conduction (2.2), it is expected that the hot-spot will occur close to one of the heat generators with the highest value. Therefore, it is very likely that the hot-spot is mislocated when the grid resolution is lower. However, the key question is if higher grid resolution necessarily translates into very high temperature gradients and large temperature differences on the chip. To understand this, it is necessary to perform thermal simulations with a correct thermal model of the chip.

3.4.2 Temperature

Thermal simulations are performed by providing the extracted heat maps (Section 3.4.1, Figure 3.13 and Figure 3.14) for the listed four cases:

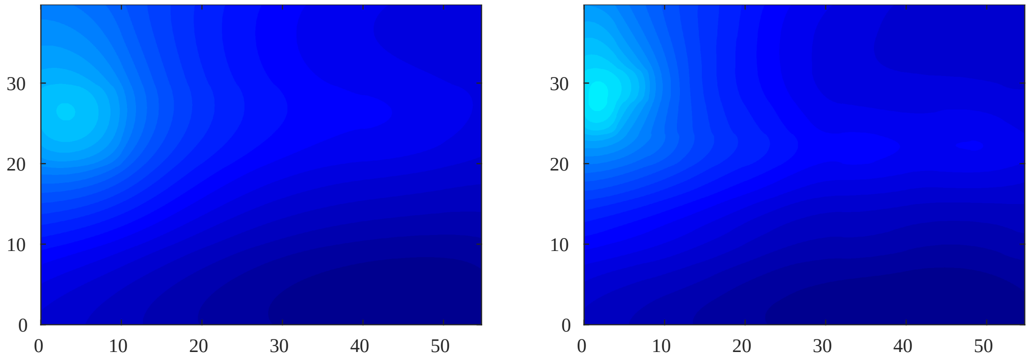
- 40 nm Bulk
- 40 nm FDSOI
- 28 nm Bulk
- 28 nm FDSOI

to compare the effect of self-heating in bulk and FDSOI and to observe the effect of technology scaling in temperature profiles and peak temperature values in a design. During the thermal simulations, the proposed thermal models for the bulk and the FDSOI cases are used (Section 3.2.2, Figure 3.3).



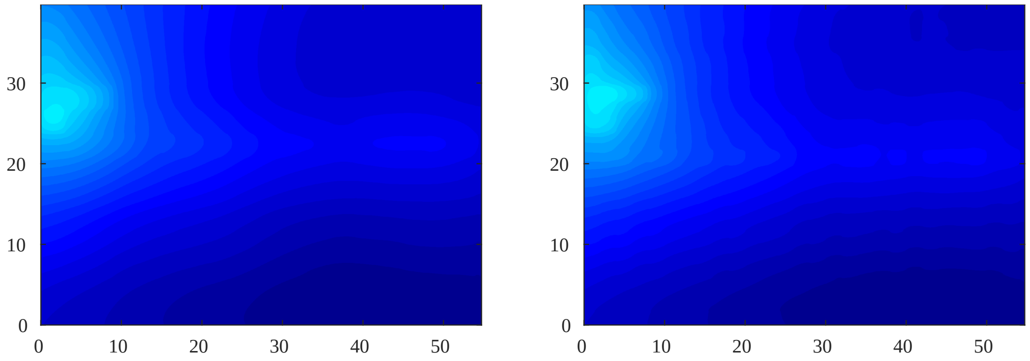
(a) $X_p = 60$ µm

(b) $X_p = 20$ µm



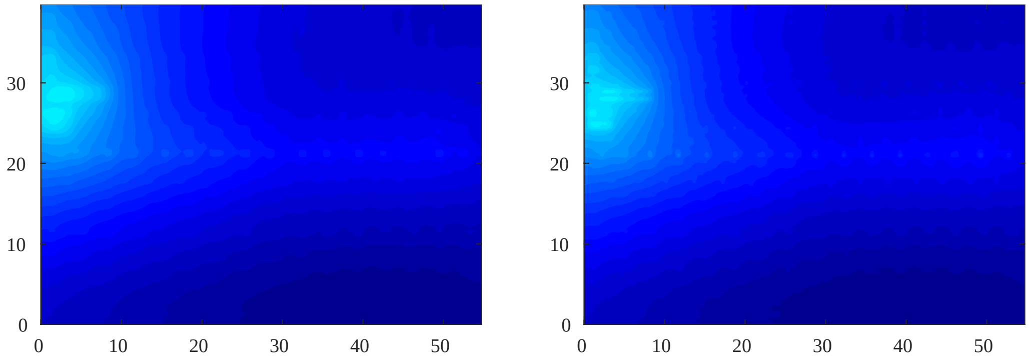
(c) $X_p = 10$ µm

(d) $X_p = 4$ µm



(e) $X_p = 3$ µm

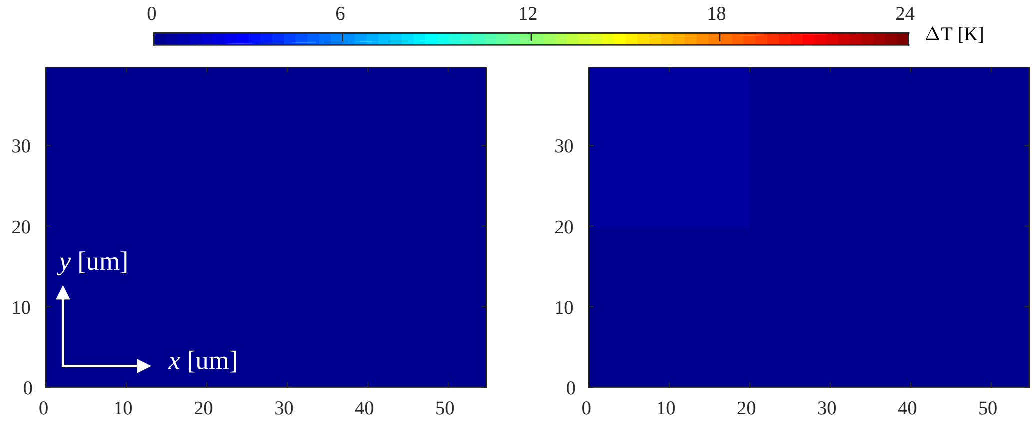
(f) $X_p = 2$ µm



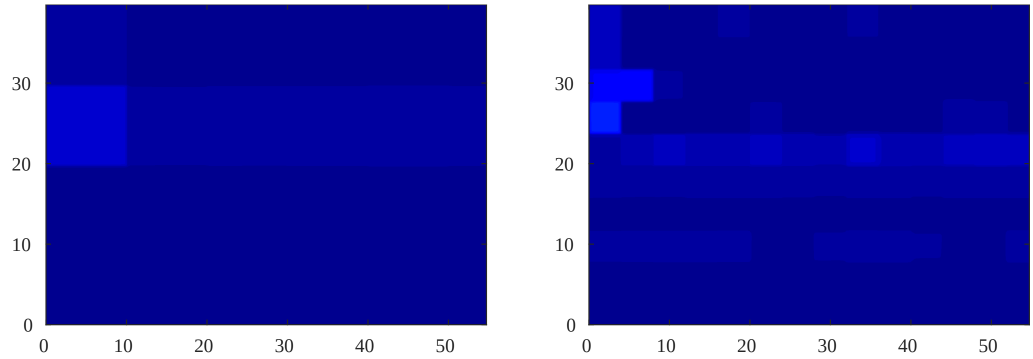
(g) $X_p = 1$ µm

(h) $X_p = 0.5$ µm

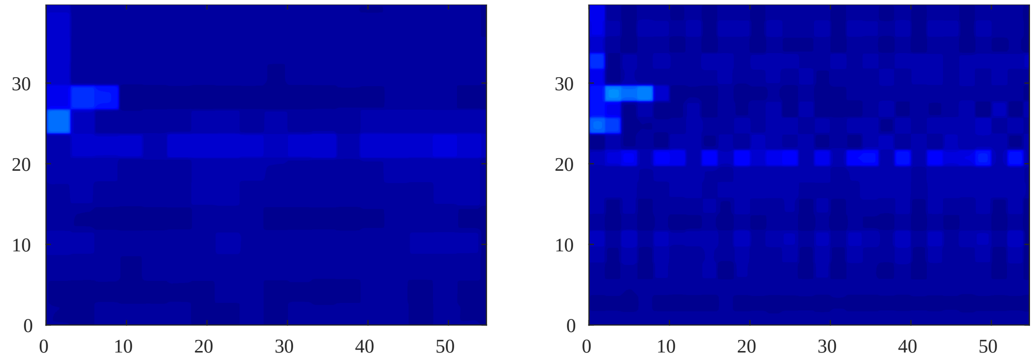
Figure 3.16 – Temperature maps of the 40 nm adder in bulk geometry for different pitch values.



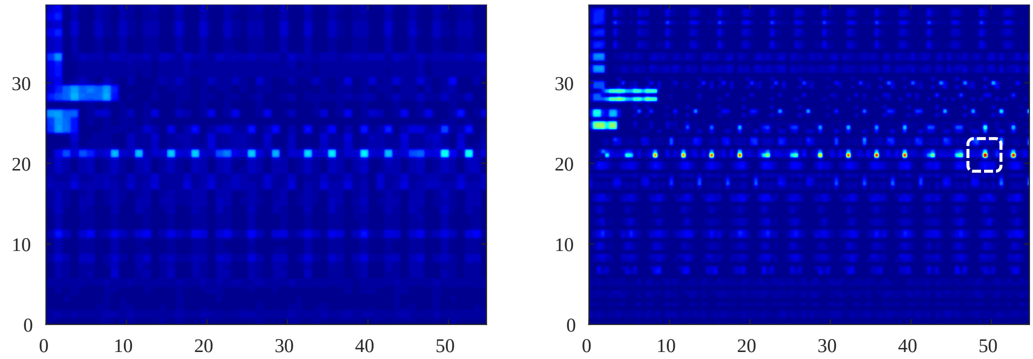
(a) X_p = 60 µm (b) X_p = 20 µm



(c) X_p = 10 µm (d) X_p = 4 µm

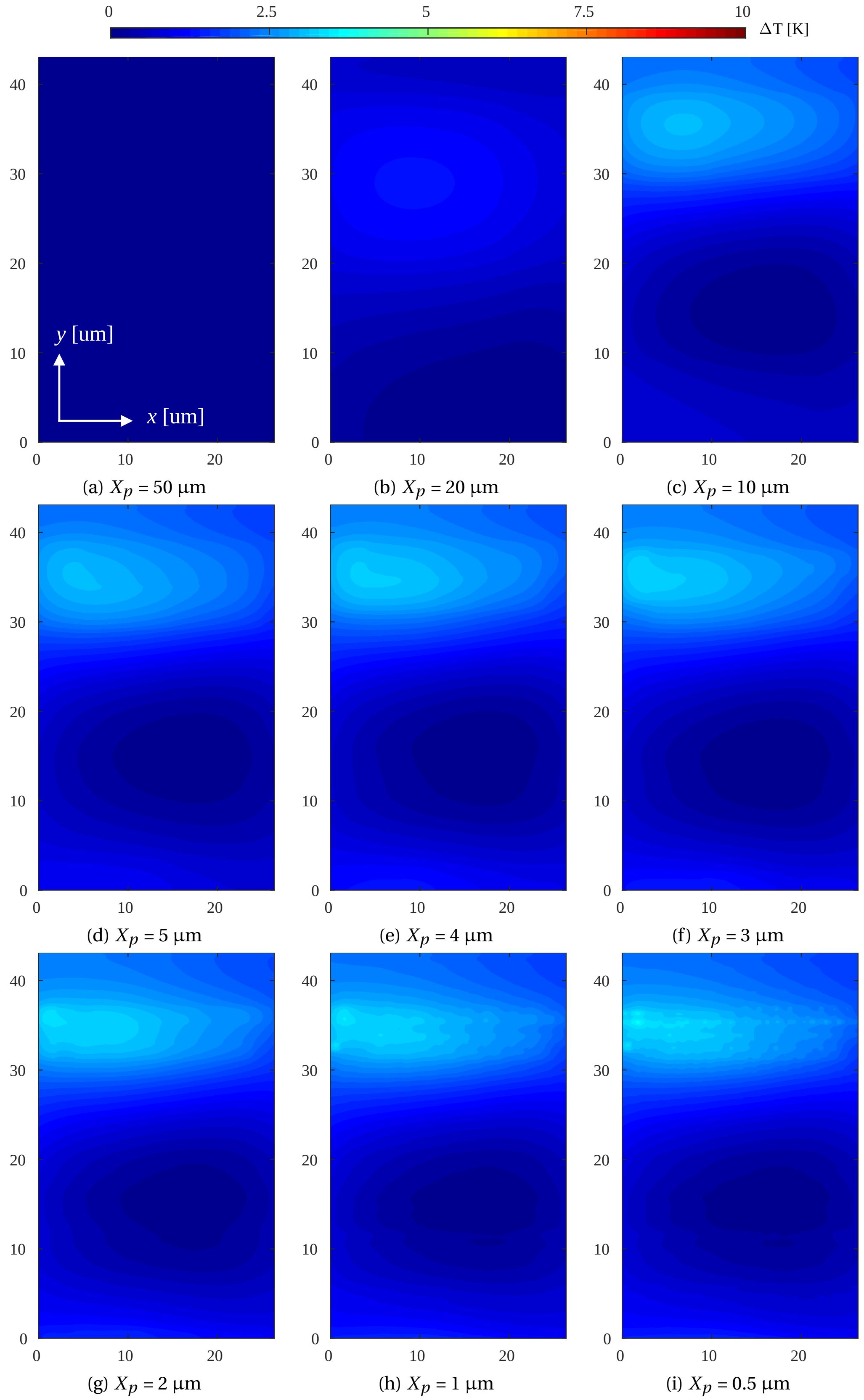


(e) X_p = 3 µm (f) X_p = 2 µm



(g) X_p = 1 µm (h) X_p = 0.5 µm

Figure 3.17 – Temperature maps of the 40 nm adder in FDSOI geometry for different pitch values.



(a) $X_p = 50$ µm

(b) $X_p = 20$ µm

(c) $X_p = 10$ µm

(d) $X_p = 5$ µm

(e) $X_p = 4$ µm

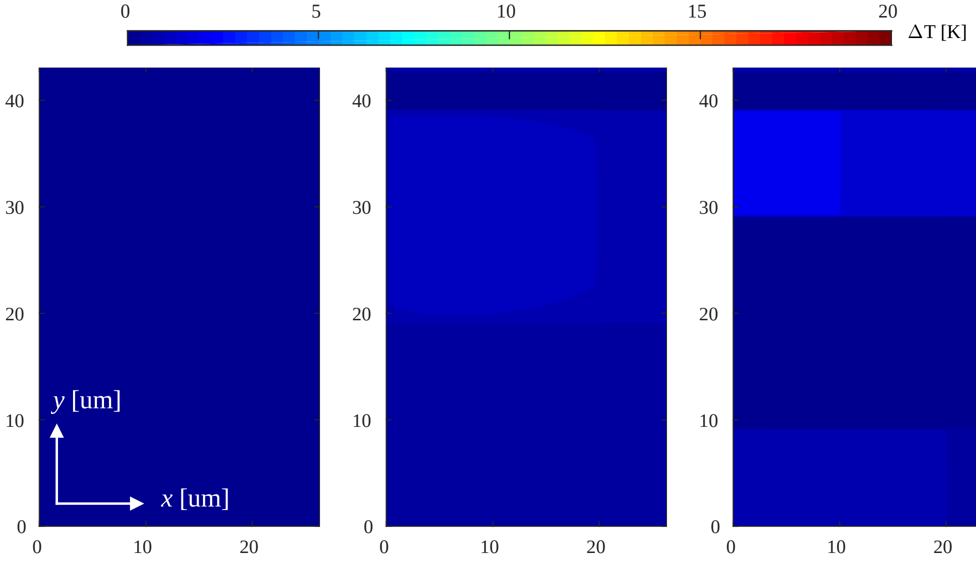
(f) $X_p = 3$ µm

(g) $X_p = 2$ µm

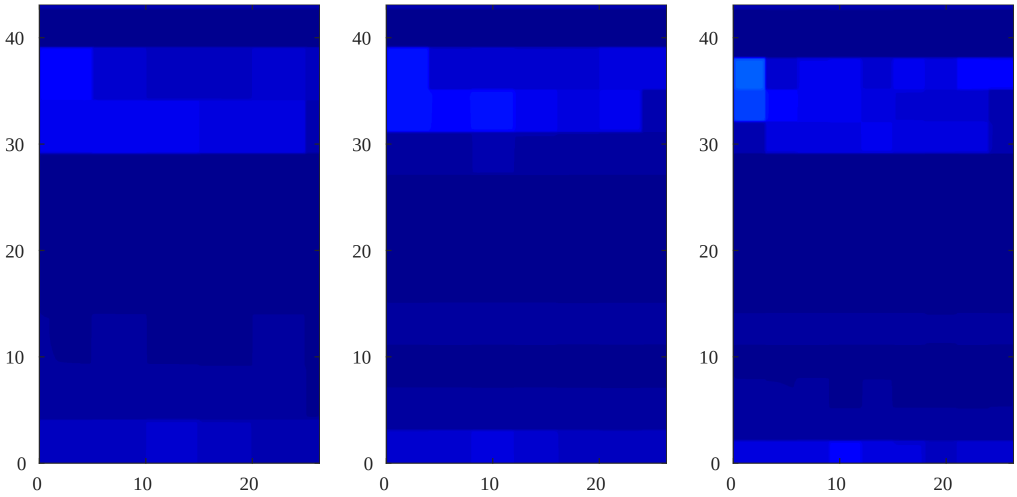
(h) $X_p = 1$ µm

(i) $X_p = 0.5$ µm

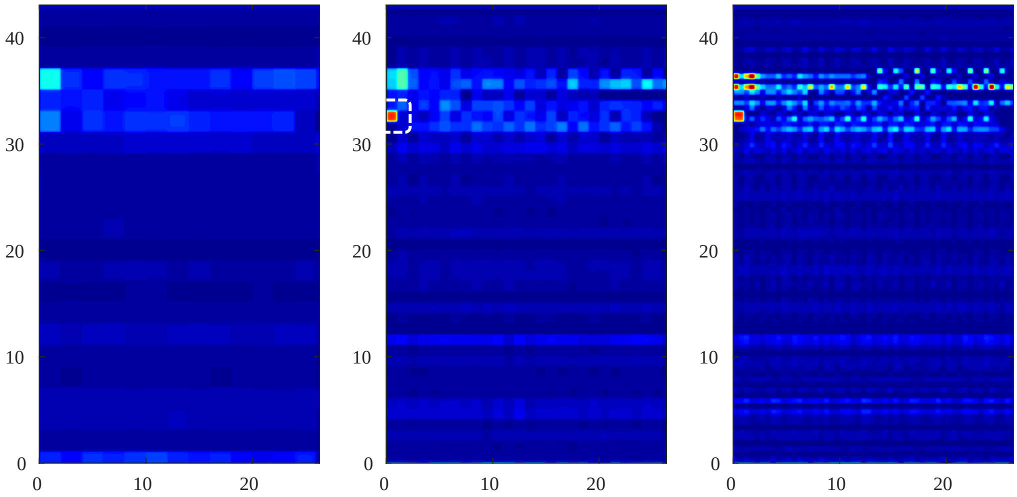
Figure 3.18 – Temperature maps of the 28 nm adder in bulk geometry for different pitch values.



(a) $X_p = 50$ µm (b) $X_p = 20$ µm (c) $X_p = 10$ µm

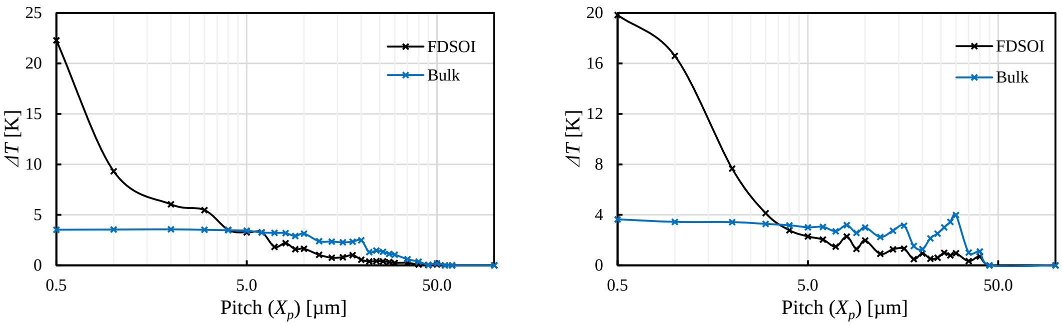


(d) $X_p = 5$ µm (e) $X_p = 4$ µm (f) $X_p = 3$ µm



(g) $X_p = 2$ µm (h) $X_p = 1$ µm (i) $X_p = 0.5$ µm

Figure 3.19 – Temperature maps of the 28 nm adder in FDSOI geometry for different pitch values.



(a) 40 nm

(b) 28 nm

Figure 3.20 – Maximum and minimum temperature values for all temperature maps with heating inputs of different resolutions.

The resulting temperature maps obtained from the thermal simulations can be observed on Figure 3.16, 3.17, 3.18 and 3.19 for the previously listed four cases. It should be noted that these temperature maps are normalized with respect to the lowest temperature for each case rather than providing the absolute temperature, in order to provide information about how high the temperature of the hot-spot is compared to different locations in the design. Hence, the temperature maps provide the temperature rise

$$\Delta T(x, y) = T(x, y) - T_{min} \tag{3.16}$$

with respect to the minimum temperature in the entire area (T_{min}), where $T(x, y)$ is the absolute temperature at the specified x and y coordinate.

Temperature Profile of the Bulk Geometry

Figure 3.16 and Figure 3.17 show that the temperature maps for the bulk and the FDSOI cases are quite different in 40 nm node. The same interpretation can be made also for the 28 nm node by observing Figure 3.18 and 3.19. In all four cases, the largest pitch value clearly does not give any temperature gradient since there is a single grid square input in the thermal simulation. Hence the highest temperature is 0 K larger than the minimum temperature of the entire area (i.e. ΔT = 0 K). For the bulk thermal geometry in both technology nodes, X_p = 20 µm (Figure 3.16b and Figure 3.18b) gives a hot-spot that can be observed at the point where the heat density is the largest (Figure 3.13b and Figure 3.14b). ΔT at these hot-spots is 2.48 K for 40 nm and 1.28 K for 28 nm. As we decrease the pitch of the heat map (from X_p = 20 µm to X_p = 10 µm), we observe that the shape and the location of the hot-spot changes slightly in addition to the little increase in ΔT value, which is measured as 3.14 K for 40 nm and 3 K for 28 nm. However, further decreasing the pitch does not change the location, shape and the ΔT of the hot-spot significantly in bulk geometry (Figure 3.16c-3.16h for 40 nm and Figure 3.18c-3.18i for 28 nm). For the lowest pitch value, the value of ΔT at the hot-spot location is 3.53 K for 40 nm and 3.68 K for 28 nm, which are not even 1 K more than the case of X_p = 10 µm. This trend can also be observed on Figure 3.20, where the largest value of ΔT is quite

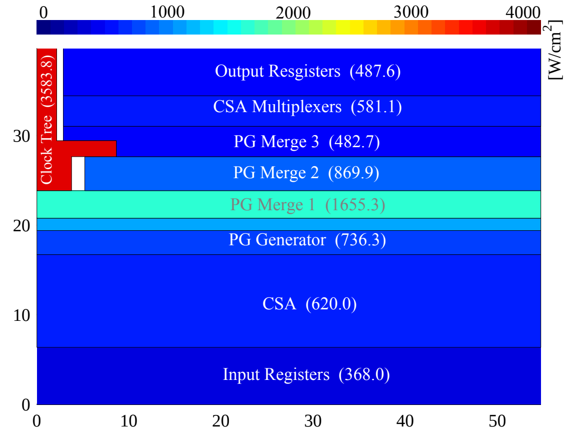


Figure 3.21 – Heat map of different blocks in 40 nm with their individual heating (x and y in µm).

constant as X_p is increased up to 10 µm and starts to decrease with some fluctuations as it is further increased.

By analysing the heat generation of different blocks and the floor-plan of the adders, we have found that the hot-spot in the **bulk** geometry is due to the large power dissipation and close placement of the clock distribution buffers. In the 40 nm case, the clock distribution buffers are located on the upper left corner of the adder (Figure 3.21); whereas, in 28 nm case, they are located horizontally between $y = 32$ µm and $y = 36$ µm, relatively on the left side. Since they are closely placed and their activity rate is high, they can be easily detected even by observing a heat map with relatively large pitch values (up to 5 µm). However, the grid square with the largest heating on the heat map with $X_p = 0.5$ µm ($x_0 = 21$ µm, $y_0 = 49$ µm for 40 nm and $x_0 = 24$ µm, $y_0 = 36$ µm for 28 nm) is not located on a point where clock distribution buffers are placed. It is found that in both technology nodes, the grid square with the largest heat density contain mainly pre-charge devices. Contrary to the clock distribution buffers, these devices are not necessarily closely placed since they should lie close to their corresponding logic gates. Hence they can only be detected on heat maps with pitch values lower than 1 µm. Consequently, although they are responsible of the largest local heating of the entire design, they do not create large area hot-spots such as the clock distribution buffers do due to their close placement. As a result of this, we can conclude that the hot-spots in bulk geometry is due to the clock tree although the individual heating of their constitutive transistors is not the largest in the entire design.

This analysis clearly shows that very high-resolution heat maps are not necessary for bulk thermal simulations since different heating inputs with a large range of resolution (from X_p = 0.5 µm to X_p = 10 µm) result in very similar temperature profiles and hot-spots with similar temperature values. This is due to the fact that the thermal conductance of bulk silicon (148 Wm^{-1}K^{-1}) is large enough for spreading the generated heat along horizontal and vertical distances on the order of 10 µm despite large gradients on the heating in short distances (0.5 - 2 µm). Therefore, even very highly localized heat density peaks do not give rise to hot-spots in bulk technologies.

Temperature Profile of the FDSOI Geometry

In the bulk case, it has been already shown that the temperature maps show quite similar profiles for pitch values less than 10 μm. On the other hand, the temperature maps of the FDSOI geometry are not only quite different than the bulk geometry, but also show quite different profiles when they are compared in between each other for heating inputs with different resolutions (Figure 3.17a to Figure 3.17h for 40 nm and Figure 3.19a to Figure 3.19i for 28 nm). Moreover, the maps do not tend to show similar profiles with the decreasing pitch values. Even for the smallest pitch values, the temperature profiles seem quite different from each other. This can be seen by comparing the differences between Figure 3.17g and Figure 3.17h for 40 nm, and Figure 3.19h and Figure 3.19i for 28 nm. The non-converging behaviour of the temperature profiles show that even $X_p = 0.5$ μm is not small enough to get a realistic temperature map in FDSOI geometry.

Regarding the temperature values of the hot-spots, the peak temperature shows an increasing trend with the decreased pitch, which can be observed on Figure 3.20 for both technologies. For the smallest pitch value ($X_p = 0.5$ μm), this value reaches up to 22.3 K for 40 nm and 19.8 K for 28 nm. Similar to the previously explained fact that the temperature maps do not converge to a unique profile with the decreasing pitch down to 0.5 μm, the temperature of the hot-spot does not saturate neither. On the contrary, it continues to increase similar to what is observed on the heat maps of Figure 3.15.

Another observation can be made regarding the temperature gradients. Significant temperature differences (~20 K) are observed in very low distances in the order of ~200 nm. Some of the highest gradients can be observed on the encircled regions on Figure 3.17h for 40 nm and on Figure 3.17h for the 28 nm. By comparing these points with the temperature maps of the bulk case, it can be seen that the temperature gradients are significantly larger than the ones in the bulk geometry.

A comparison between the heat maps and temperature maps of FDSOI shows that their distributions are very similar. The grid squares (i.e. pixels) of the heating input can be clearly observed on the temperature maps. Even for the pitch value of 0.5 μm the distinction between the pixels can be observed, which is not the case in bulk even for the largest pitch values. This is another point that shows us the necessity to perform a higher resolution thermal simulation to have a more natural temperature profile in FDSOI.

All of the previously mentioned points show that the generated heat cannot be readily dispersed with the help of a diffusion path and it directly translates into elevated temperature along the distances down to 0.5 μm. This is mainly because of the high thermal resistance geometry around the individual devices due to the Si thin film, SiO_2 and the boundaries between these layers, the thermal properties of which can be found on Table 3.1. From these observations, it is evident that one should further increase the resolution of the heating input for FDSOI thermal simulations to have a better temperature map that has a more realistic temperature distribution with smoother transitions. In that case, it would be possible to observe higher temperature gradients and to pinpoint the devices that operate with the highest temperature.

Finally, the floor-plan analysis show that the hot-spots in **FDSOI** do not necessarily appear on the clock tree distribution network contrary to the bulk case. They find themselves rather on the nanometer scale regions where the individual devices with the largest heat density values are located. In these adder designs, we found that the hot-spot regions contain pre-charge devices since their heat density values are observed to be the largest (The different group of devices and their heating values will be explained in Section 3.5.2 in more detail). However, this does not necessarily mean that the pre-charge devices will always be the main responsible of the hot-spot. In fact, if a group of device with another function had larger heat density, a hot-spot would appear on where they are placed. Consequently, we can conclude that the hot-spots in FDSOI are created by the individual devices with the highest heat density.

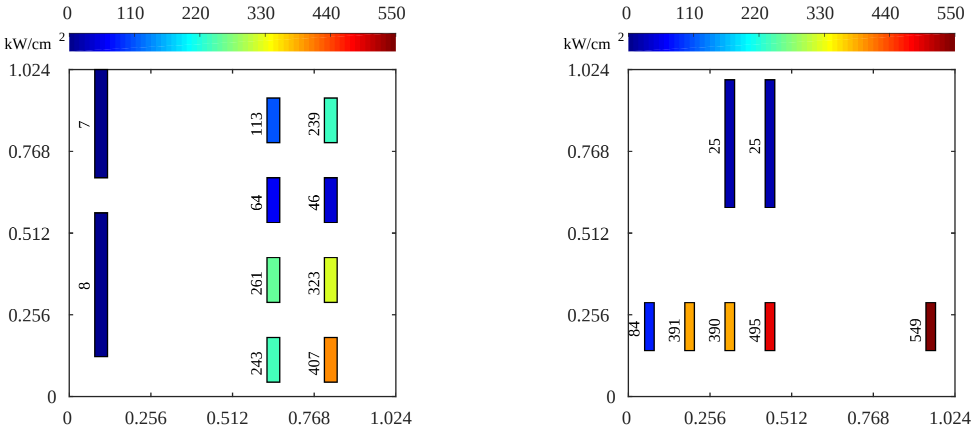
Focus on Hot-Spots

More detailed thermal simulations are performed by providing higher resolution heat maps. This time, rather than simulating the entire block, we focus on the regions where the largest heat density values are observed on Figure 3.13h for 40 nm and on Figure 3.14i for 28 nm. During the high-resolution simulations, the heat generators are introduced as individual devices (Figure 3.22) rather than a grid array as it is done by applying (3.5). Like this, it becomes possible to see the peak temperature values on individual FDSOI devices, which is not possible with $X_p = 0.5$ µm (on Figure 3.17h and Figure 3.19i). The heat density of each device is calculated as

$$Q_n = \frac{\overline{P_n}}{W_n L_n} \tag{3.17}$$

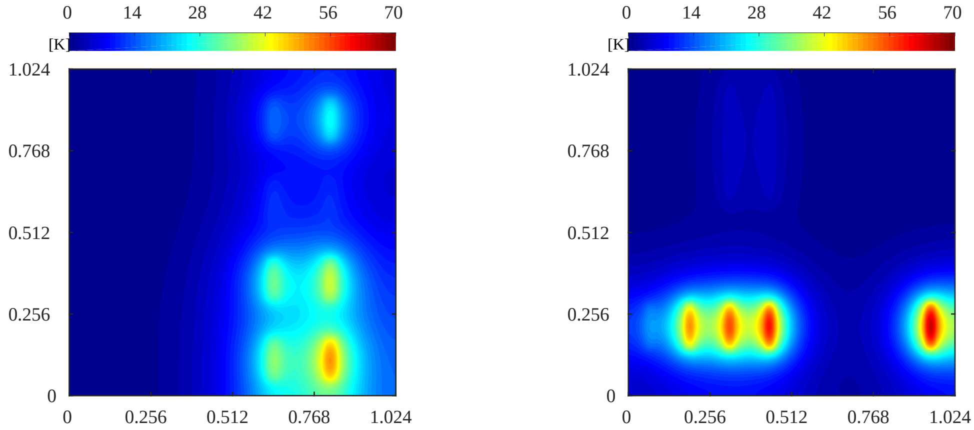
where W_n and L_n are the width and the length of device n.

For the 40 nm adder, the transistor with the largest heat density can be seen on Figure 3.22a with a power dissipation of 22.8 µW, in addition to some other devices with smaller heat density values. Considering the size of the device, this power dissipation value corresponds to a heat density of 407 kW/cm^2. This value is more than an order of magnitude larger than the heat density of the pixel with the largest heating observed on Figure 3.13h. The resulting temperature map for 40 nm FDSOI can be seen on Figure 3.23a. It can be seen that the temperature profile in the FDSOI case starts assuming a more natural appearance with the increased resolution as expected. The highest temperature is observed precisely on the device with the largest heat density (right bottom on Figure 3.23a). The temperature at this point is 50.5 K higher than the minimum temperature of the design. Although this device is surrounded by some other high heat density devices, it can also be seen that the temperature value on this device rapidly decreases on the right side of the device. This points that the temperature rise of this device is mainly due to its self-heating.



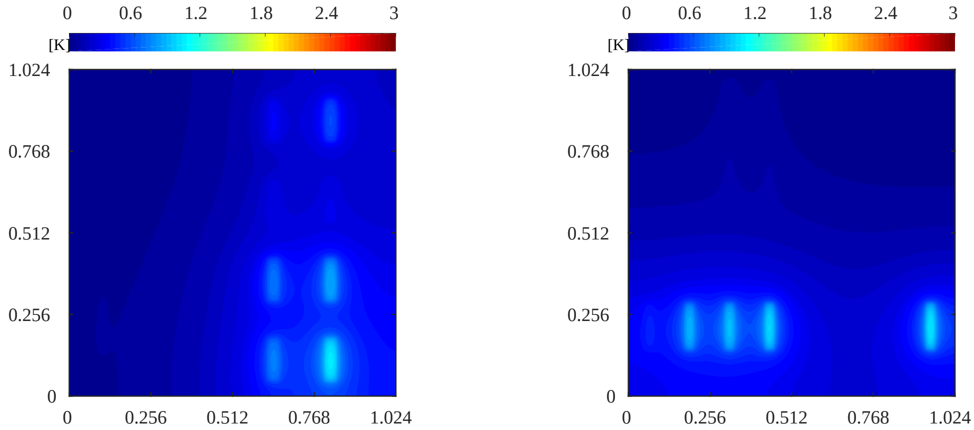
(a) 40 nm (b) 28 nm

Figure 3.22 – Heat maps of (a) 40 nm and (b) 28 nm technologies focused on the hot-spots.



(a) 40 nm (b) 28 nm

Figure 3.23 – Temperature maps of (a) 40 nm and (b) 28 nm FDSOI technologies focused on the hot-spots.



(a) 40 nm (b) 28 nm

Figure 3.24 – Temperature maps of (a) 40 nm and (b) 28 nm bulk technologies focused on the hot-spots.

The heat map of the same experiment for the 28 nm adder can be seen on Figure 3.22b. The device with the largest heat density in the focused area is located on the right bottom corner of the square with a heat density of 549 kW/cm^2 and an average power dissipation of 24.7 µW. The heat density of this device is only slightly less than the heat density of the device which has the largest value (554.3 kW/cm^2) in the entire design. Figure 3.23b shows that the peak temperature of the focused region is 64.2 K and it is located on the device with the largest heat density. This is quite similar to 40 nm FDSOI case. However, the peak temperature of the 28 nm adder is more than 15 K larger than the 40 nm one. This is primarily due to the larger maximum heat density of the 28 nm technology since the devices have smaller area. In addition to this, the device on Figure 3.22b is not surrounded by any other high heat density device that would increase its temperature by heat transfer, which is not the case in 40 nm on Figure 3.23a. This means that the effect of self-heating on the peak temperature becomes even more prominent with technology scaling.

The temperature maps are also generated for the bulk geometry and can be seen on Figure 3.24. It can be seen that the devices with very high heat density can still be detected in bulk due to their larger temperature values. Nevertheless, these values are only 1.06 K in 40 nm and 1.03 K for the 28 nm even for the devices with the largest heat density. These values are significantly smaller than the ones in FDSOI for both technology nodes. In fact, these values are also much smaller than the largest temperature values observed on the clock distribution block. This proves one more time that in bulk geometry it is sufficient to consider the heating values of the blocks of sizes of 5×5 µm^2 to 10×10 µm^2 rather than considering each device in order to get a realistic temperature profile and detect the hot-spots.

These observations show that the hot-spot in FDSOI is located on the device that has the largest heat density primarily due to its self-heating. On the other hand, this is not the case in bulk and one finds the hot-spot on the block with the largest power density. Moreover, the peak temperature of a small region in FDSOI can reach to much larger values (more than 20 times) than bulk. Additionally, the temperature of the hot-spot becomes even larger in smaller technology nodes in FDSOI since self-heating effects become more prominent due to device scaling; whereas, the difference in the device temperature due to its self-heating is not so significant in bulk. The increased temperature of the devices due to self-heating also changes other parameters of the devices. For example in the FDSOI case, the temperatures of different devices in the vicinity of 500 nm distance differ by more than 45 K. With this temperature difference, the drain currents may differ by 5% to 10%, which can cause serious performance issues, especially in circuit blocks such as read/sense amplifiers where symmetry is important.

Heating Effects of Adjacent Devices in FDSOI

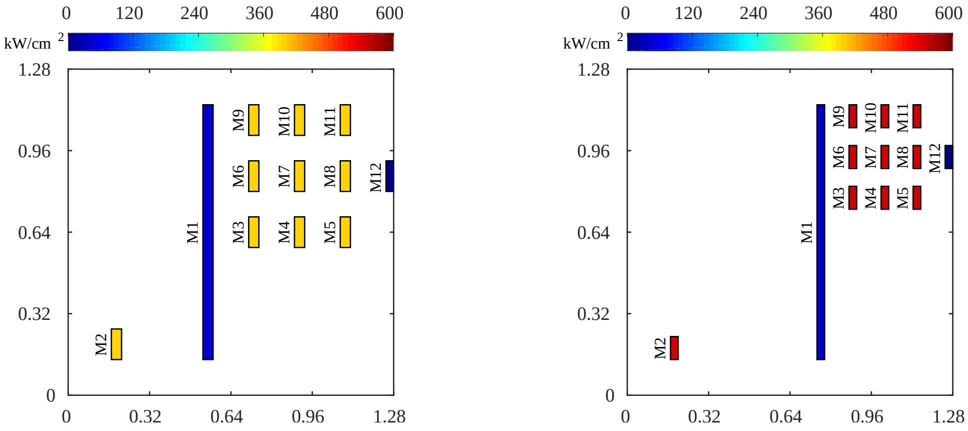
Up to this point, we have mainly focused on the effects of self-heating. However, especially in FDSOI, closely placed devices have significant effect on each other's temperature. To understand the effect of adjacent devices on each other, a heating scenario is created for both technologies. Figure 3.25 shows the locations of the devices with their names indicated

Table 3.4 – Parameters of the devices on Figure 3.25

Technology	**40 nm**			**28 nm**		
Device	**M1**	**M2 - M11**	**M12**	**M1**	**M2 - M11**	**M12**
Heat Density [kW/cm^2]	48.0	400.0	0	49.5	550.0	0
Power Dissipation [μW]	19.2	19.2	0	14.85	14.85	0
Width [μm]	1.00	0.12	0.12	1.00	0.09	0.09
Length [μm]	0.04	0.04	0.04	0.03	0.03	0.03

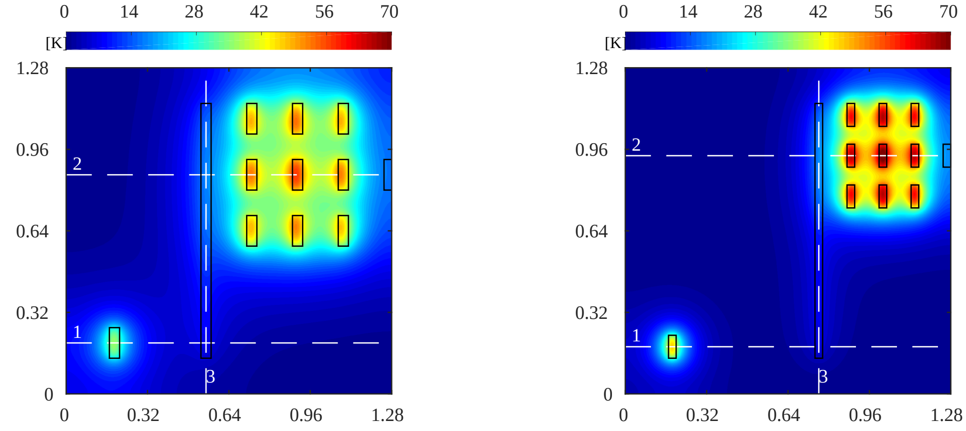
on the left. The colours of the devices are determined according to the colour-bar on the top, which represents the heat density. For the test case of Figure 3.25, the horizontal and the vertical distances of the devices are set to minimum according to the design rules of the utilized technologies. The highest heat density values are chosen according to the heat density distribution of the adders, which are shown on Figure 3.29. The average power dissipation of devices M2-M11 is roughly equal to the average power dissipation of the device that has the largest heat density in the entire adder circuit according to heating analysis. The size of these devices is set to the minimum size for each technology while their aspect ratio is kept equal to three. According to the previous analyses in FDSOI, one of these devices should be responsible of the hot-spot of a digital block in larger scale. Device M12 has the same size of M2-M11 with no power dissipation. Finally, the power dissipation of M1 is equal to the power dissipation of M2-M11 while its heat density is much lower than the others due to its larger size. The rest of the parameters for each device are summarized on Table 3.4.

The resulting temperature profile of the listed heat generators are shown on Figure 3.26. The white dashed lines on Figure 3.26 are the cut-lines that pass on the centre of the devices. The temperature waveforms occurring on these cut-lines are shown on Figure 3.27. The peak temperature on cut-line 1 shows that the device with the highest heat density (M2) can increase the temperature itself alone by 36 K in 40 nm and 46 K in 28 nm. The peak temperature on cut-line 2, which is the peak temperature of the entire area, is observed on M7 with a temperature increase of more than 57 K in 40 nm and 71 K in 28 nm compared to the coolest point. Additionally, the temperature of M7 is 21 K and 25 K higher than the temperature of M2 in 40 nm and 28 nm respectively although they have the same heat density. This proves that the influence of the adjacent devices (M3-M6 and M8-M11 in the example) can increase the peak temperature of the overall circuit significantly. Another observation that can be made on cut-line 2 is although M12 does not dissipate any power, its temperature is roughly 17 K higher than the lowest temperature in the area for both technologies. This is primarily due to the self-heating of M3-M11, which are located very close to M12. The temperature of M1 is more than 15 K lower than the temperature of M2 although their power dissipation values are equal. This result proves that the most substantial factor of creating a nanometer scale hot-spot in FDSOI is heat density rather than power dissipation alone.



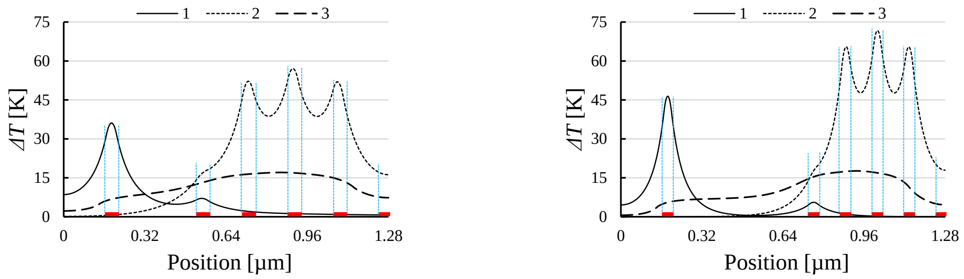
(a) 40 nm (b) 28 nm

Figure 3.25 – Heat maps of (a) 40 nm and (b) 28 nm for observing extreme case temperature profiles.



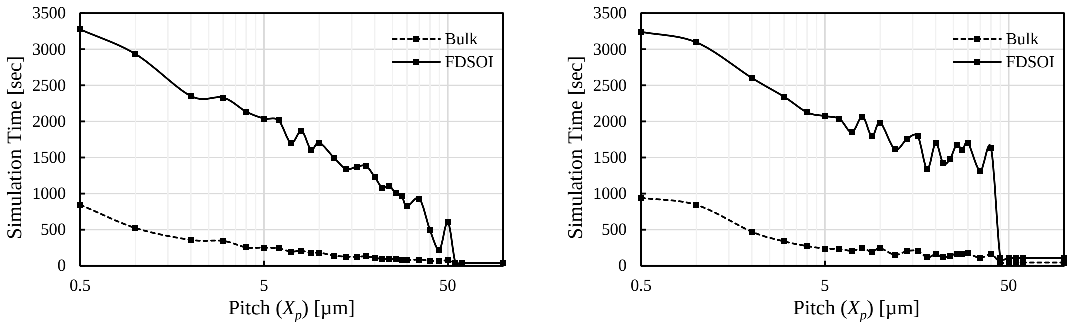
(a) 40 nm (b) 28 nm

Figure 3.26 – Temperature profiles of (a) 40 nm and (b) 28 nm due to the heating inputs of Figure 3.25.



(a) 40 nm (b) 28 nm

Figure 3.27 – Temperature profiles along the cut-lines (Figure 3.26) taken at the centre of the devices. The red thick lines show the location of the devices on Figure 3.26.



(a) 40 nm

(b) 28 nm

Figure 3.28 – Simulation time of bulk and FDSOI with respect to pitch. In this analysis the resolution of the heating input is swept while the resolution of the thermal simulation is fixed at 2048 × 2048.

Finally, the temperature of M1 shows more than 10 K variation along its width (cut-line 3 on Figure 3.27). The gradient is due to the adjacent devices (M3, M6 and M9) which are very strong heaters. As a result of this variation, the threshold voltage and mobility of M1 would show different behaviours along its width. While this effects only the speed in digital circuits, it might create important mismatch problems in analog circuits.

3.4.3 Simulation Time

In Section 3.4.2, it was shown that one can get a accurate temperature profile by having a pitch value of 5-10 µm for the bulk thermal geometry; however, this value should go down to 10-20 nm for FDSOI. Although increasing the resolution of the simulation provides more detailed temperature profiles in FDSOI, it is also necessary to consider the utilized computational resources and the simulation time.

Figure 3.28 shows the simulation times for each simulation case of Figure 3.16 to Figure 3.19 and some other additional points. It can be seen that the simulation time is quite small for large pitch values and it increases with the increasing resolution. For X_p = 0.5 µm, bulk geometry simulations take around 700-900 sec, which is much less than FDSOI. This is due to the fact that the geometry of the bulk technology is simple and the thermal conductance is relatively high. Fortunately, one does not need such high-resolutions in bulk since the simulation results converge to a sufficiently similar profile for small X_p values. In fact, X_p = 5 µm already gives reliable results and the simulation takes less than 250 sec. On the other hand, FDSOI simulations with X_p = 0.5 µm take around 3300 sec. Moreover, it was previously shown that the resolution should be even higher to get a more reliable temperature profile. This is due to the fact that the FDSOI geometry is complex and it contains different layers with different thermal conductance. This means that, the simulation time will be even much larger than 3300 sec to get a reliable result. Consequently, to perform a reliable thermal simulation in FDSOI, one needs a much longer simulation with higher resource utilization than a bulk thermal simulation.

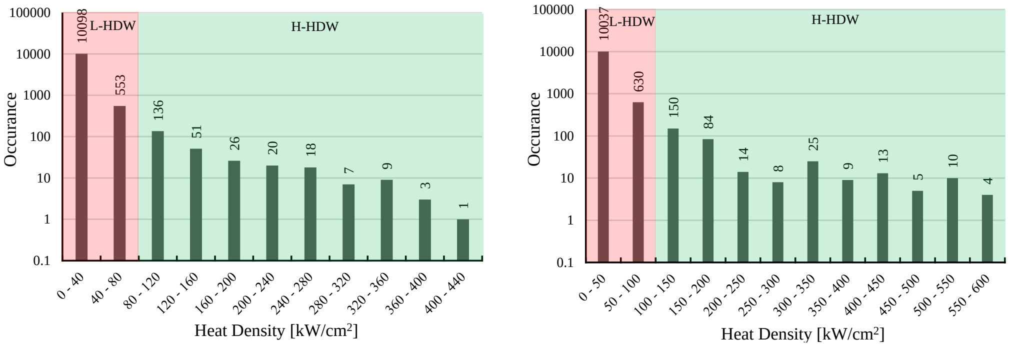
3.5 Circuit Level Analysis of Hot-Spots

It has been shown that the heating of both adders is quite non-uniform. Thereby, the temperature profile in FDSOI is also non-uniform. In this section, all nMOS and pMOS devices in the adders are analysed in terms of their heat density values. To understand the level and reasons of the non-uniformity, heat density histograms are generated. The devices are grouped according to their function and each group is examined in terms of their average and maximum heat density values. The groups with the largest heat density are studied, the reasons of such large heating is investigated and solutions are proposed for increasing the temperature uniformity and decreasing the hot-spot temperature.

3.5.1 Heat Density of Individual Devices

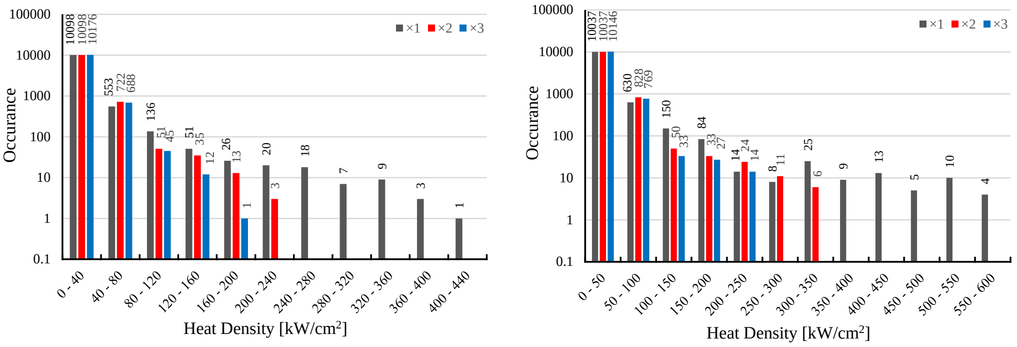
The hot-spot in FDSOI is most probably created by the self-heating of individual devices with the highest heat density. Moreover, their cumulative effect can increase the peak temperature significantly. Figure 3.29 shows the distribution of the transistors with respect to their heat density for the 40 nm and the 28 nm adders. It can be seen that the heat density range is limited by [0 kW/cm^2, 440 kW/cm^2] window in 40 nm and by [0 kW/cm^2, 600 kW/cm^2] window in 28 nm. However, both histograms show an exponential decaying characteristic and most of the devices are cumulated in the region with very low heat density. At this point, we can define two windows on the histograms of Figure 3.29 and call them *Low Heat Density Window* (L-HDW) and *High Heat Density Window* (H-HDW). The border (i.e. the threshold value) between each other is quite arbitrary and defined as 80 kW/cm^2 for 40 nm and 100 kW/cm^2 for 28 nm in this study. What is important is (a) the ratio of the largest heat density in the design to the threshold value and (b) the total number of devices in each window. For the 40 nm node, there are 10651 devices in L-HDW, whereas there are only 271 devices in H-HDW. This means that almost 97.5% of all the devices has a heat density less than 80 kW/cm^2 and only approximately 2.5% of the devices are generating more than one fifth of the largest heat density observed in the entire circuit (80 kW/cm^2/406.8 kW/cm^2 $\approx$ 20%). Similar situation can be also observed in 28 nm adder. Figure 3.29b shows that L-HDW contains 10667 devices, while there are only 313 devices in H-HDW. In terms of ratios, the L-HDW contains approximately 97% of all the devices. On the other hand, only 3% of the devices are in H-HDW, which is slightly higher compared to the 40 nm case in addition to the higher selected threshold value. The described situation can also be observed on Figure 3.13 and Figure 3.14 where there are only very few red or yellow squares (due to the devices in H-HDW) and almost the entire area of the adders is covered with dark blue (due to the devices in L-HDW). Therefore, if the devices in the small portion (H-HDW) of the heat density histograms are detected and modified in a way, the peak temperature of the block can be significantly decreased.

The modification of the most critical devices with high heat density can be performed by increasing their width with a trade-off of a slightly increased area and parasitic capacitances. Figure 3.30 shows the histograms of the three cases where the width of the devices in H-HDW



(a) 40 nm (b) 28 nm

Figure 3.29 – Heat density distribution of the transistors in the 64-bit adder circuits.



(a) 40 nm (b) 28 nm

Figure 3.30 – The heat density distribution after increasing the width of the high heat density devices by a factor of 2 and 3.

are kept constant (×1), doubled (×2) and tripled (×3) (Figure 3.31). For the 40 nm node, it can be seen on Figure 3.30a that the heat density distribution can be squeezed into a smaller window only by increasing the width of the devices with heat density of more than 80 kW/cm^2. The new values of maximum heat density are less than 225 kW/cm^2 and 165 kW/cm^2 for the ×2 and ×3 cases respectively. Similar results are obtained also in 28 nm node by applying the same solution. Figure 3.30a shows that the device with the maximum heat density generates less than 305 kW/cm^2 and 225 kW/cm^2 for the ×2 and ×3 cases respectively.

Although the maximum heat density can be decreased significantly by performing the described analysis and modifications, it is quite cumbersome to perform an extensive power dissipation analysis by considering each device. Moreover, to have accurate results from the described analysis, one should perform the necessary simulations by considering the interconnect parasitics. Consequently, the modifications on the critical devices in H-HDW have to be performed on top of a layout-ready design, which would increase the design time significantly. For that reason, a correlation between the devices in H-HDW and their functional properties is needed to be sought, so that the common properties of these devices and the

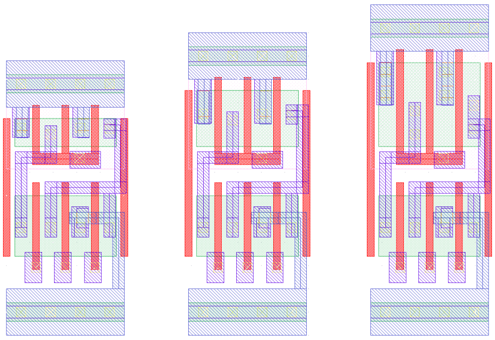

(a) ×1 (b) ×2 (c) ×3

Figure 3.31 – Standard cell of a *Generate-Merge* gate with different size pre-charge devices.

reason of their large heat density can be understood. This way, the necessary thermal design precautions can be taken during the early design stages rather than performing a very long analysis on an already finalized design.

3.5.2 Self-Heating of Devices with Different Functions

To observe the group of devices with different average heat density values, the devices are categorized according to their functions. Considering the different functions in the circuit, there are nine different groups. Each group is explained by means of Figure 3.32, which shows the schematic of a simple logic function implemented in dynamic logic.

1. **Static (ST):** The transistors of the static logic gates into which no clock signal enters (ST_1 and ST_2 on Figure 3.32).
2. **Logic Tree (LT):** The transistors used for implementing the Boolean functions of the dynamic logic gates (LT_1, LT_2 and LT_3 on Figure 3.32).
3. **Footing Device (FD):** The footing device, which is used for preventing any leakage via the logic tree path during the pre-charge phase (FD on Figure 3.32).
4. **Pre-charge Output Node (PO):** The load device, which is used to charge or discharge the output node during the pre-charge phase (PO on Figure 3.32).
5. **Pre-charge Intermediate Node (PI):** The devices used for pulling up or down the intermediate nodes in the logic tree of dynamic gates for preventing charge sharing problem (PI on Figure 3.32).
6. **Clock Tree (CT):** The devices of the static inverters that distribute the clock signal **only** to the dynamic logic gates (CT_1 and CT_2 on Figure 3.32).
7. **Pass Transistor Logic (PT):** The devices used to implement transmission gate multiplexers (PT_1 and PT_2 on Figure 3.32).

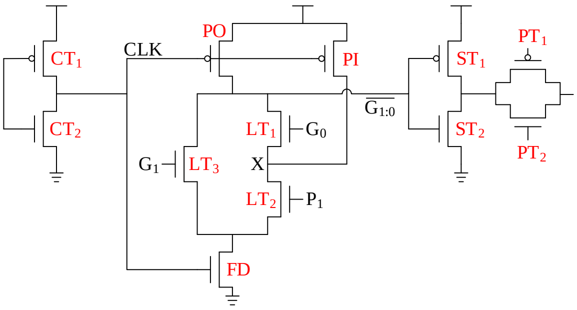


Figure 3.32 – The schematic that illustrates the devices belonging to different groups. The logic gate in the centre corresponds to the domino logic implementation of Generate-Merge function with N-type logic tree.

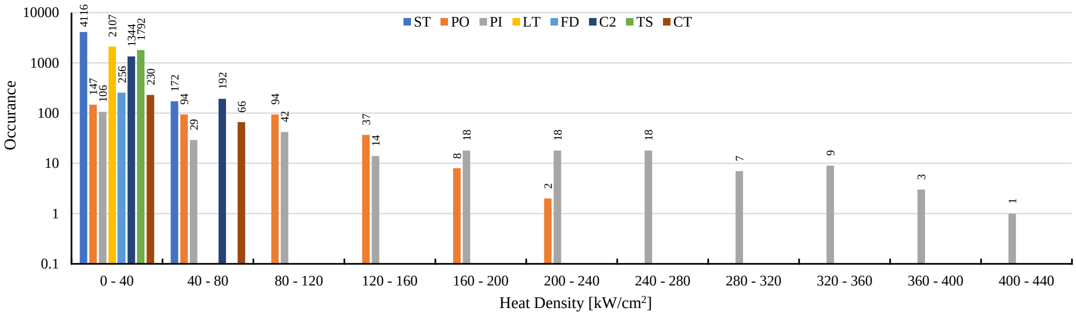
8. **True Single Phase Clock D Flip Flop (TSPC-DFF) (TS):** The constitutive transistors of the TSPC-DFFs (Not shown on Figure 3.32).
9. **C^{2}MOS D Flip Flop (C2):** The constitutive transistors of the C^{2}MOS-DFFs (Not shown on Figure 3.32).

The heat density distribution of the 64-bit adder circuits (Figure 3.29) are elaborated on Figure 3.33 by providing individual distribution of each device group. It can be seen that the heat density distributions of individual groups in both technology nodes shows significant differences in terms of their mean values and population range.

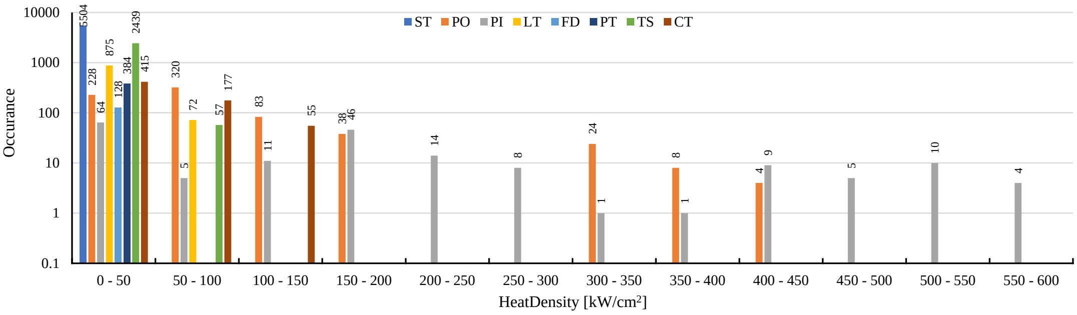
For the 40 nm node, in the lowest heat density range [0 kW/cm^2, 40 kW/cm^2], it is possible to find any group of devices. In fact, a large portion of all type of devices are populated in this window. However, in the high heat density range [80 kW/cm^2, 440 kW/cm^2], only two device groups out of eight exists. These are the PO and PI groups. According to Table 3.5, the heat density value of any group other than PO and PI is less than 72.9 kW/cm^2; while the devices in the PO and PI groups can reach to heat density values more than 400 kW/cm^2. Moreover, the average heat density value of the PO and PI devices is quite high compared to the other groups.

The situation in 28 nm is not so different. The lowest heat density window [0 kW/cm^2, 50 kW/cm^2] contains all types of devices. However, the clock distribution devices have relatively larger heat density, where their maximum value reaches to 141.4 kW/cm^2. Nevertheless, the device groups with the largest heat density are still PO and PI with a maximum heat density of more than 550 kW/cm^2.

The results obtained from implementations in both technology nodes indicate that PO and PI are the most critical groups in terms of creating a hot-spot in a circuit in FDSOI. As a circuit level solution, instead of performing an extensive heating analysis, one can focus on the sizing of these devices in the early stage of a design so that the local peak temperature values in a circuit can be decreased significantly.



(a) 40 nm



(b) 28 nm

Figure 3.33 – Distribution of heat density for different group of devices in the 64-bit adders implemented in (a) 40 nm and (b) 28 nm.

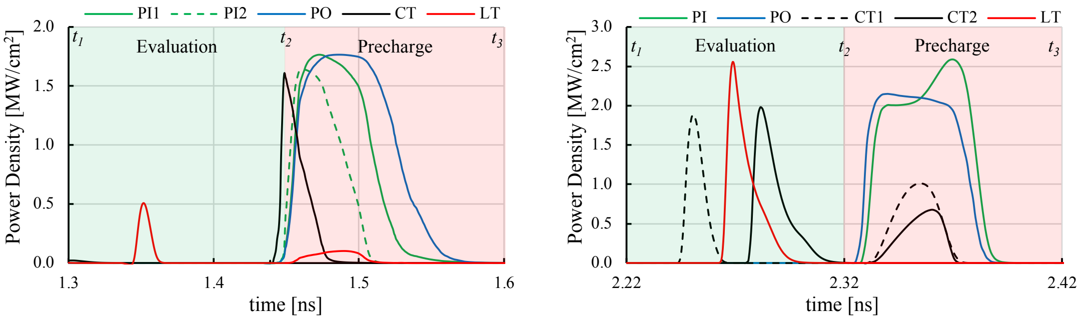
Table 3.5 – Maximum and mean heat density values and the number of devices from different device groups in 40 nm.

Group	Max [kW/cm^2]	Mean [kW/cm^2]	Number
PI	406.8	108.0	265
PO	238.9	62.6	382
ST	72.9	11.6	4288
CT	70.3	34.1	296
C2	69.3	14.4	1536
LT	34.8	5.1	2107
TS	20.3	6.4	1792
FD	16.1	8.1	256

The large heat density of the devices in the PO and PI groups can be understood by comparing the time domain power density behaviours of these devices with the devices from the other groups. Figure 3.34 shows the time domain power density waveforms of the devices from the PO, PI, CT and LT groups for both technology nodes. The evaluation (from t_1 to t_2) and the pre-charge (from t_2 to t_3) phases can be seen on Figure 3.34. The devices that are located on any path between the input and the output are mainly active in the evaluation phase. Some of

Table 3.6 – Maximum and mean heat density values and the number of devices from different device groups in 28 nm.

Group	Max [kW/cm^2]	Mean [kW/cm^2]	Number
PI	554.3	179.3	178
PO	421.3	82.4	705
CT	141.4	53.0	647
LT	94.3	17.7	947
TS	65.4	15.8	2496
ST	48.8	10.4	5504
PT	35.7	5.7	384
FD	26.2	13.2	128



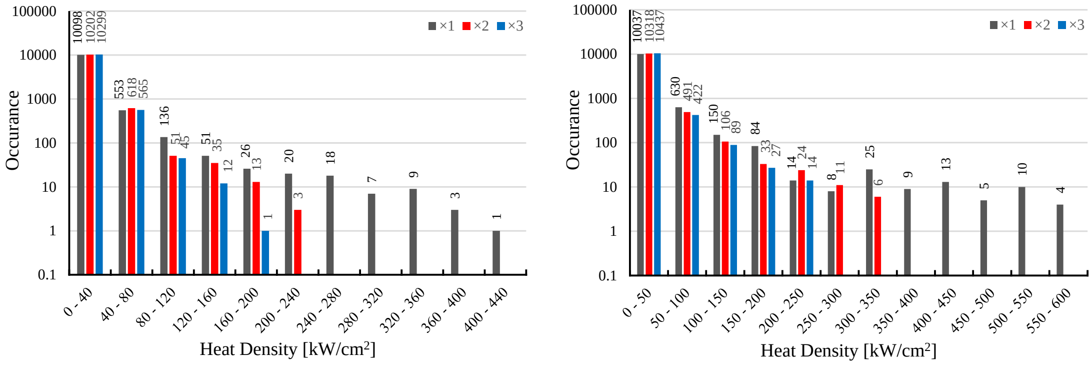
(a) 40 nm

(b) 28 nm

Figure 3.34 – Time domain power dissipation density waveforms of devices from different groups.

the examples of these devices are LT, ST and FD. Since these devices are on the data path, for a design with a short critical path delay, they should be sized in a way that they are conducting (i.e. ON) for a short period compared to the entire evaluation period. Consequently, these devices conduct current, hence generating heat for a relatively short time. This can be seen on Figure 3.34 by observing the power density waveform of the LT device which is conducting for a very small time interval compared to the other devices. The PO and PI devices are operating only during the pre-charge phase. However, they do not have to charge or discharge their corresponding nodes as fast as the devices on the data path. As long as they pre-charge the capacitance seen at their drain terminal before the beginning of the next evaluation phase, the circuit can safely work. Therefore, they are sized as small as possible for having a smaller area and less parasitic capacitance. The consequence is having a longer conduction time. Since the PO and PI devices conduct for longer durations compared to the other groups, they have larger average heat density values according to (3.17) and they become the most critical devices in terms of creating a hot-spot.

The activity rate is another important parameter of heat density in addition to conduction



(a) 40 nm

(b) 28 nm

Figure 3.35 – The heat density distribution after increasing the width of the PI and PO devices by a factor of 2 and 3.

duration of a device. The CT devices have the highest activity rate among all the devices. The probability of conduction of a CT device is one in each clock period. However, for all the rest of the devices this probability is less than one and it depends on the input data, logic function and the location of the device in the circuit architecture. Nevertheless, the clock tree devices have less power density than the PO and PI devices. This is due to our previous reasoning that they have to provide a fast transition since they affect the critical path delay, hence the speed of the circuit. A fast transition brings less average power density (Figure 3.34). Finally, it can be seen from Table 3.5 and Table 3.6 that the maximum and the mean power density values of PI are significantly larger than the ones of PO although they both conduct during the pre-charge phase and they have the same task. This is mainly due to the probability of discharging or charging an intermediate node during the evaluation phase that is higher than the probability of making an evaluation at the output node in most of the cases. This probability increases as the number of series devices decrease and the parallel branches increase between the corresponding node and the power supply in a logic tree (Figure 3.32). Consequently, the temperature of the pre-charge devices become higher.

As it has been already mentioned, for having better thermal robustness the maximum local temperature values should be decreased by decreasing the heat density values of the critical devices (PO and PI). This can be done by increasing the width of these devices with a trade-off of an increased area and a slightly decreased speed similar to what has been done in Section 3.5.1. Figure 3.35 shows that the heat density distribution can be squeezed into a smaller window only by increasing the width of the PO and PI devices by a factor of 2 and 3. It can be seen that the histograms of Figure 3.30 and Figure 3.35 are the same when only H-HDW is considered and the maximum heat density values are significantly decreased. Moreover, by knowing which group of devices are critical, the proper sizing can be done much more easily and the maximum heat density value can be further decreased only considering a few devices in the entire circuit. In addition to proper sizing, the devices can be separated from each other to relax the heat flow [63]. Finally, thermal vias can be added to the drain ends of these devices

Table 3.7 – Change of thermal parameters in scaling from 40 nm to 28 nm.

Technology	40 nm	28 nm	Unit	Increase
Average heat density of the adder	723.4	1283	W/cm^2	77%
Maximum heat density of the adder	406.8	554.3	kW/cm^2	36%
Maximum achievable heat density	2.4	3.35	MW/cm^2	40%
ΔT_{max} of the adder in FDSOI	64.2	50.5	K	24%
ΔT_{max} of the adder in bulk	3.53	3.68	K	4%

similar to what is shown in [112] and [113] to provide better heat diffusion paths, which will in turn increase the delay time.

3.6 Conclusion

In this work, we have primarily compared thermal behaviours of digital circuits in bulk and FDSOI geometries. This comparison was performed on two custom designed 64-bit Kogge-Stone adders in 40 nm and 28 nm technologies.

It was shown that the hot-spots in FDSOI have significantly larger temperatures than the ones on bulk. The hot spots in FDSOI are mainly created by nanometer scale heat generators such as individual devices due to their confinement by low thermal conductance materials. The sizes of the hot-spots are small and this results in very high temperature gradients. On the other hand, the hot-spots have larger sizes in bulk due to the larger thermal conductance of bulk Si. The temperature gradients are also smoother in bulk. Due to this, the hot-spots are mostly created by building blocks rather than individual devices. For obtaining a reliable temperature map in FDSOI, one should perform a thermal simulation with a resolution down to tens of nanometers. On the other hand, in bulk, a resolution of some micrometers is sufficient. However, increasing the resolution of thermal simulations results in using more computational resources and the simulations take very long time compared to bulk.

It was also shown that the self-heating effects become more prominent with the device scaling in FDSOI. The peak temperature of the hot-spots are roughly 15 K larger in 28 nm compared to 40 nm. The rest of the parameters are compared on Table 3.7. The results point out that the self-heating effects in the future technologies will bring more critical reliability and performance issues, not only in FDSOI but also in other confined geometries such as FinFET.

The circuit level analysis of hot-spots shows that the most critical devices in terms of creating a hot-spot are the pre-charge devices in domino logic. These devices are the most critical group in terms of an electromigration breakdown. It is demonstrated that sizing the critical devices slightly larger (2 to 3 times) would decrease the maximum heat density by 60% with a trade-off of a slight decrease in the speed and an increase in the area. Like that, the electromigration probability in a drain contact can be significantly decreased.

4 Self-Heating Effects on the Noise Performance of FDSOI MOSFETs

In Chapter 2 and Chapter 3, we have shown that the local temperature of FDSOI devices can rise significantly due to self-heating effects. Moreover, this situation is even more serious for devices with constant bias, which are widely used in analog design. In this chapter, self-heating effects on the analog performance of FDSOI MOSFETs are investigated mainly focusing on the thermal and the flicker noise.

4.1 Introduction

When it comes to the precise estimation of the temperature of a device, analog design is by far more critical compared to digital. A digital implementation would demonstrate its correct functionality in a wide temperature range, even if serious precautions are not taken during the design; certainly with a different performance such as lower speed and higher power dissipation. On the other hand, analog design contains much wider range of constraints like gain, matching, noise, linearity etc., which are very sensitive to temperature. A slight shift in one of these parameters might result in malfunction of the entire IC. For example, a mismatch due to the different temperature values of a differential pair in a comparator might cause incorrect estimation of the data in a decision feedback equalizer (DFE), which might result in error propagation and failure of the communication.

One of the main drawbacks of elevated temperature is the deterioration of the thermal and flicker noise performance. Thermal noise current has a direct linear dependence on temperature. For this, any increase in temperature manifests itself as increased thermal noise. On the other hand, there is no direct relationship between temperature and flicker noise. Nevertheless, due to the temperature dependence of different device parameters, flicker noise also worsens at higher temperature values. To understand the phenomena better and quantify these effects, this chapter is dedicated to the analysis of thermal and flicker noise in FDSOI MOSFETs considering self-heating effects and demonstration of the self-heating influence on the thermal and flicker noise performance. To quantify the self-heating influence, a self-heating thermal model is used. The thermal model for the self-heating is included in the

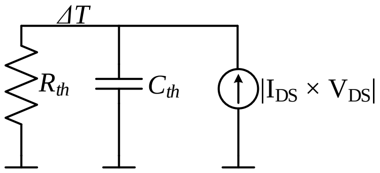


Figure 4.1 – The thermal model that provides the temperature rise of each device due to their self-heating [2].

device model by the used design kit, which is a commercially available 28nm FDSOI technology. The temperature rise of each device due to their individual self-heating is calculated by considering the thermal model of Figure 4.1. According to the model, R_{th} is the thermal resistance and C_{th} is the heat capacitance. The values of these parameters are calculated for each device according to their geometries (width, length, number of fingers etc.). The power dissipation of the device is modelled by a current source. The temperature value on the upper common node gives the temperature increase ΔT of the device due to its self-heating. The complete model is included in the device by the design kit provider and can be switched on or off for a specific simulation. This way, with the provided self-heating model, it is possible to consider the self-heating effects for all bias conditions.

The influence of self-heating on the thermal noise is examined by activating and deactivating the self-heating thermal model and comparing the results. It is observed that self-heating effects increase the noise significantly especially for short channel devices. The influence is observable even in medium bias conditions, which demonstrates that considering self-heating effects and the temperature values of each device during the design is necessary in order to estimate correct noise levels and tune the parameters properly.

This chapter is organized as follows. In Section 4.2, the theory of thermal noise in MOSFET devices is given in addition to different short channel thermal noise models proposed by different authors. At the end of this section, thermal noise current and input referred thermal noise voltage performance of FDSOI MOSFET devices are compared for the cases where self-heating effects are taken into account and ignored. In Section 4.3 the influence self-heating effects on the flicker noise performance of FDSOI devices are investigated with the theory and numerical examples. Finally, in Section 4.4, the summary of the work and the conclusions are provided.

4.2 Thermal Noise

One of the main sources of noise in MOSFET devices is the *thermal noise*. It covers the entire spectrum and it has a flat power density for all frequencies. For this reason, it is also referred as *white noise*. The source of thermal noise is the thermal agitation of the charge carriers inside a conductor at equilibrium. Hence, it has a linear dependency on the temperature.

4.2.1 Theoretical Background and Related Work

The power spectral density of the thermal noise current in the channel of a MOSFET device is given by [114]

$$S_{Int} = 4kT\gamma g_{d0} \tag{4.1}$$

where k is the Boltzmann constant, T is the absolute temperature, g_{d0} is the zero-bias drain conductance of the MOSFET device (i.e. the drain conductance of the device when $V_{DS} = 0$ V is applied) and γ is the excess noise coefficient. For long channel devices in strong inversion, γ gradually decreases from 1 to 2/3 while moving from linear region to saturation and it becomes equal to 2/3 in saturation. However, for devices with a short channel, the value of γ tends to be larger due do various factors. *Channel length modulation* is one of the effects that increase γ for short channel devices. Increasing V_{DS} creates a pinch-off region on the drain end and decreases the effective length of the device where the inversion layer charge is present in the channel. This increases the thermal noise current as the device moves deeper into saturation [37]. *Velocity saturation* is another factor that affect the charge density and consequently the thermal noise current of the device. In [115, 116, 117], the authors included velocity saturation in the derivation of the thermal noise current and observations made on an increasing trend of γ under smaller channel lengths and higher velocity saturation.

Most of the authors explain the large γ by *hot electron effect* [115, 117, 118, 119]. This effect is mainly observed in short channels under large V_{DS} voltage. Large lateral electric field in the channel creates hot electrons, which have higher temperature (i.e. larger average kinetic energy) than the lattice. In this case, the average temperature of these hot electrons has to be considered rather than the lattice temperature. The temperature of the hot electrons can be approximated by [120]

$$T_e = T_L\left(1 + \frac{E}{E_c}\right)^n \tag{4.2}$$

where T_e is the hot electron temperature, T_L is the lattice temperature, E_c is the critical electric field and E is the lateral electric field in the channel with $n = 2$, 1, or 0. While $n = 0$ corresponds to no hot carriers case, $n = 2$ gives the other extreme. In [117], (4.2) is modified as $T_e = T_L + \delta T_L\left(\frac{E}{E_c}\right)^2$ where δ is used as a fitting parameter to adjust the simulations to the measurement. With the hot carrier model, the higher temperature electrons show more random fluctuations and this can be modelled by providing a larger γ in (4.1).

Another explanation of larger γ under very large V_{DS} is the *avalanche multiplication* [118, 121]. When V_{DS} is larger than the band-gap voltage, high energy carriers are scattered by valence band electrons and additional electron-hole pairs are created. The thermal fluctuations of these additional carriers contribute to the overall noise and this increases γ.

The listed modifications on the thermal noise current provide good approximations for the short channel devices; however, they don't account for the self-heating of individual devices *for a large range of bias conditions* or they assume the increase in γ due to self-heating is already included in the model by fitting between measurements and simulations. To obtain more realistic values for the thermal noise current power spectral density under different bias conditions, one has to estimate the individual temperature of a device accounting its self-heating effects. Most of the current simulators can find solutions under a single temperature input that is valid for all the devices in the netlist. Consequently, different devices with different lattice temperatures are assumed to have the same temperature value. This would result in some errors in the calculation of the thermal noise current and hence in the noise figure of amplifiers, especially when they are implemented in advanced technologies like FDSOI, FinFET where the devices are confined inside low thermal conductance materials. As it has been already mentioned, very large temperature gradients and different temperature values in short distances can be observed in these technologies, mainly due to low thermal conductance and high power densities [19, 26, 33, 34]. Therefore, the noise analysis should be performed simultaneously by considering the self-heating effects.

4.2.2 Thermal Noise Considering Self-Heating

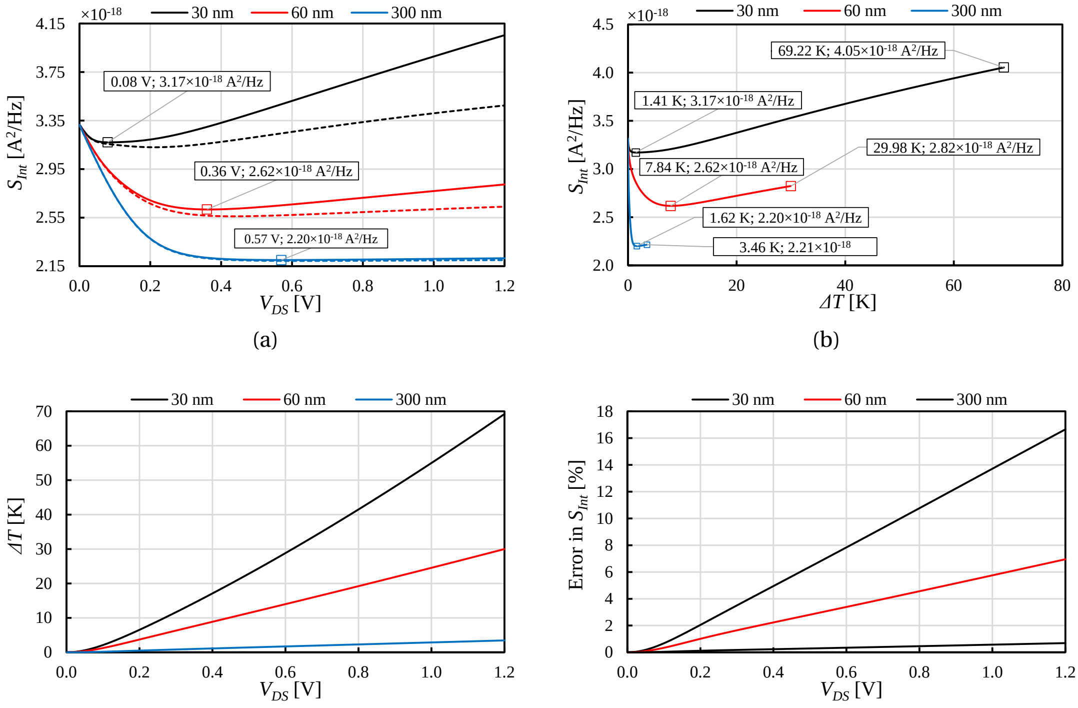
Thermal Noise Current

For observing the effect of self-heating on the thermal noise current, nMOS devices with different gate lengths are simulated by activating and deactivating the self-heating thermal model. The gate lengths of the devices are varied from 30 nm to 300 nm. The gate voltage is kept constant at $V_{GS} = 0.6$ V and V_{DS} is set as the sweep parameter. All devices are implemented with 20 fingers and their finger width is adjusted so that their zero bias drain conductance (g_{d0}) is set approximately equal to 20 mA/V. Like this, all devices have equal thermal noise current under no power dissipation and no self-heating. Among all devices, the ones with the gate lengths of 30 nm, 60 nm and 300 nm are studied in more detail to observe the short, medium and the long channel behaviours. Some of the important parameters of these devices can be found on Table 4.1.

Figure 4.2a shows the thermal noise current. It can be seen that the behaviour of the device with the gate length of 300 nm shows a perfect long channel behaviour. The relatively large thermal noise current in the 0 - 300 mV window is due to linear region operation. The thermal noise current in saturation region drops to two third of the value that is observed at $V_{DS} = 0$ V in linear region. Moreover, it is quite flat in saturation. On the other hand, the behaviour of the 30 nm device is quite far from the long channel behaviour. The increase in the thermal noise current in the saturation region as V_{DS} increases can be explained with velocity saturation and hot electrons. In addition to these observations, it can be seen that for the short channel device (L = 30 nm), thermal noise current shows quite different trends for the cases where self-heating is activated and deactivated. For low V_{DS} values, the two curves are almost overlapping. On

Table 4.1 – Parameters for the devices with short, medium and long channels.

Parameter	L	W	Area	$g_m\|_{V_{DS}=1V}$	$P\|_{V_{DS}=1V}$	$Q\|_{V_{DS}=1V}$
Unit	nm	nm	μm^2	mA/V	mW	kW/cm^2
Short Channel	30	20 × 545	0.33	15.45	4.286	1311
Medium Channel	60	20 × 805	0.97	17.49	3.204	331.7
Long Channel	300	20 × 2515	15.1	18.65	2.732	18.11

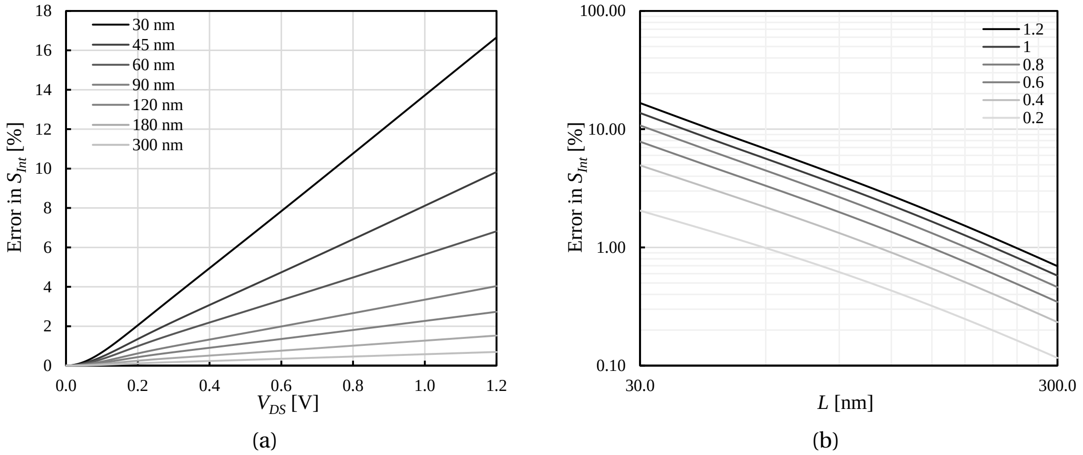


(a) (b)

(c) (d)

Figure 4.2 – Thermal noise current and temperature of 30 nm, 60 nm and 300 nm nMOS devices as V_{DS} is changed. The continuous curves correspond to the case where self-heating effects are considered and the dashed curves correspond to the case where they are ignored.

the other hand, as V_{DS} increases, the curves start to split. This is because of the higher power dissipation under larger V_{DS} and the consequent higher operating temperature due to self-heating. The same observation can also be made for the long channel device (L = 300 nm); however, the difference is much less significant than the short channel device. This is not due to the lower power dissipation of the long channel device compared to the short channel one. In fact, their power dissipation values are more or less at the same order (Table 4.1); however, the short channel device has a much smaller area and its resulting power density is much larger. Therefore, the temperature of the short channel device rises to a much higher value due to its self-heating and this results in a large thermal noise current.



(a) (b)

Figure 4.3 – Error in thermal noise current (a) with respect to V_{DS} for devices with different gate lengths from 30 nm to 300 nm and (b) with respect to gate length for V_{DS} = 0.2 V, 0.4 V, 0.6 V, 0.8 V, 1.0 V, 1.2 V. For V_{DS} = 0 V, there is no error.

The temperature dependence of the thermal noise current can also be observed on Figure 4.2b. It can be seen that the dependence is quite linear for each device with different gate lengths, which agrees with (4.1). The temperature rise due to the self-heating of the devices can be observed on Figure 4.2c. It can be seen that the channel length has a significant effect on the operating temperature and this results in incorrect estimation of the parameters as the gate length decreases and power density increases. For observing the amount of error in different parameters, we use (1.1). The error in thermal noise current can be observed on Figure 4.2d. It can be seen that for short channel devices, the error reaches to 17% for large V_{DS} bias even for V_{GS} = 0.6 V. The estimated temperature rise at this point is almost 70 K. These values would further increase under larger drive voltages that is permitted by the technology (up to 1 V). On the other hand, the temperature rise and the error in the long channel device is less than 1%, which is quite negligible.

From a design point of view, one might be interested in the smallest value of the thermal noise current. It can be seen on Figure 4.2a that the curves look more like a valley especially when the self-heating effects are considered. In each case, S_{Int} first decreases, reaches its smallest value and then it starts to increase as V_{DS} increases. The smallest values of S_{Int}, the corresponding V_{DS} and ΔT values are indicated on Figure 4.2a and Figure 4.2b. The increase on the left side of the minimum value of S_{Int} is because the channel has more carriers in linear region. On the other hand, the right side of the minimum value of S_{Int} is due to previously explained short channel effects and the temperature rise due to self-heating. At the end, the bias point that provides the smallest S_{Int} is obtained at the border of the linear and the saturation region. As the gate length decreases, this point moves more into the linear region due to the large heating. This is due to the rapid self-heating of the device as the drain voltage increases in saturation, which is shown by device level electro-thermal simulations in Chapter 2.

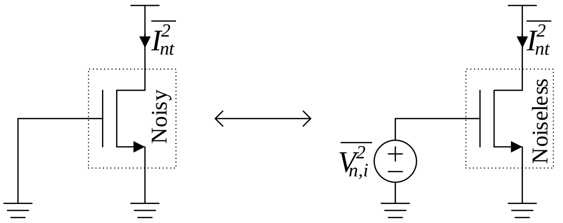


Figure 4.4 – Model of the noisy device with a fictitious input referred noise voltage source $\overline{V_{n,i}^2}$

Finally, Figure 4.3 shows the amount of error when self-heating is not taken into account. The error increases linearly with increasing V_{DS} and exponentially with decreasing gate length. Moreover, the error can go up to 17% under large V_{DS} for 30 nm gate length.

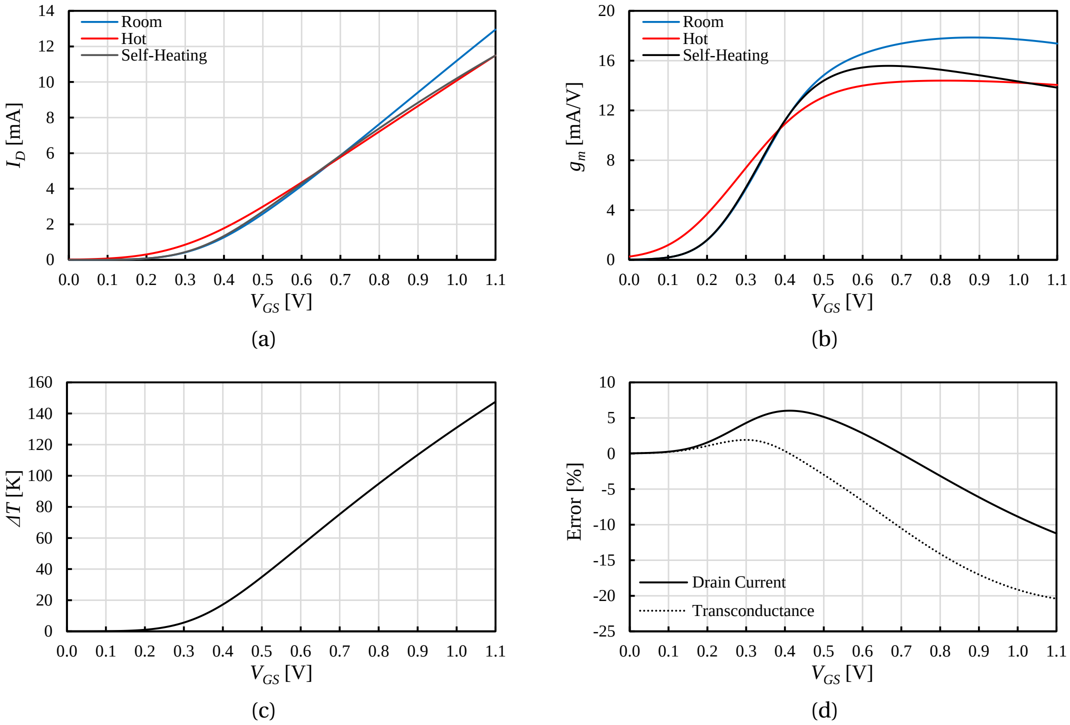
Input Referred Thermal Noise Voltage

In many applications like design of low noise amplifiers, the figure of merit for noise performance is the voltage quantity instead of the current since a voltage level is sensed at the output of the following stage rather than the current. Hence, the performance of blocks are measured by parameters like input referred noise voltage or noise figure. Therefore, one should consider the self-heating effects also on the other parameters that influence the input referred thermal noise voltage. The input referred noise voltage (Figure 4.4) of a device can be expressed as

$$S_{Vn,i} = \frac{S_{Int}}{g_m^2} = \frac{4kT\gamma g_{d0}}{g_m^2} \tag{4.3}$$

where g_m is the gate-to-drain transconductance of the device. In (4.3), g_{d0} is not influenced by self-heating since the parameter is set at the operating point where V_{DS} and the power dissipation is zero. However, g_m has a dependency on the temperature; therefore, it will be influenced by self-heating. Changing the temperature modifies g_m mainly due to the temperature dependency of two other parameters that are the threshold voltage (V_T) [122, 123] and the mobility (μ) [37, 124]. Since g_m is the first derivative of the drain current I_D, we can analyse its temperature dependence by observing the V_G-I_D curves for different temperatures.

Figure 4.5 shows how I_D and g_m changes with respect to V_G for the short channel device reported on Table 4.1. The drain voltage is set to 1 V. On Figure 4.5a and Figure 4.5b, three different cases can be observed. The blue curves correspond to the case where self-heating effects are ignored and the temperature of the device is set to 300 K. The black curves are obtained when the self-heating model is activated. The corresponding temperature rise due to self-heating can be observed on Figure 4.5c. It can be seen that the temperature is roughly 150 K larger than the room temperature at full drive. Finally, the red curves correspond to the case where the self-heating effects are deactivated and the temperature of the device is set to the maximum temperature observed by activating the self-heating effects. On Figure 4.5a, it can be seen that the device for the hot case turns on before the room and the self-heating case.

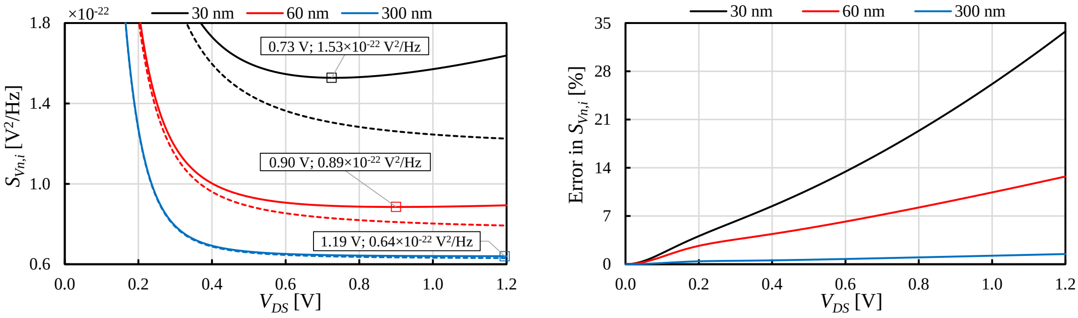


(a) (b)

(c) (d)

Figure 4.5 – (a) drain current, (b) transconductance, (c) temperature rise due to self-heating, (d) the calculation error in drain current and transconductance when self-heating effects are not considered.

This is because at higher temperatures, V_T decreases. For the room and the self-heating case, the devices turn on with the same V_G bias. With a smaller V_T, the device turns on at a lower V_G. For the self-heating case, the temperature is still not so large around $V_G \approx V_T$ since the device is not capable of heating itself in moderate inversion. Therefore, for lower V_G, the room and the self-heating curves overlap since the operating temperatures are not so different, which can be observed on Figure 4.5c. For the transconductance, the situation is the same where g_m of the hot device is larger than the other two. On the other hand, at larger temperatures mobility is smaller. Therefore, once the device turns on and enters in strong inversion, the rate of change of I_D is lower at larger temperatures. This can be seen on Figure 4.5a where I_D increases more rapidly in room case compared to the hot and the self-heating case. For the self-heating case, the behaviour at lower V_G is similar to the room temperature. However, at larger V_G levels, the device heats itself up more and the curve splits from the blue curve and converges to the red curve. Because of this, in reality (i.e. in self-heating case) g_m decreases with increasing V_G, hence the increasing temperature. Considering these two opposite trends, with the increasing temperature, g_m would be larger for lower V_G values and it starts to be smaller for higher V_G, which can be observed on Figure 4.5b. Since the temperature increase due to device self-heating is more prominent for large V_G (strong inversion and saturation) and the devices are mostly biased in strong inversion for high speed, we focus more on the region where self-heating decreases the transconductance. Consequently, input referred thermal noise voltage increases according to (4.3) which worsens the noise figure of an amplifier.



(a)

(b)

Figure 4.6 – (a) input referred thermal noise voltage and (b) its calculation error when self-heating effects are ignored. The devices are biased with $V_G = 0.6$ V.

Figure 4.6a shows the input referred thermal noise voltage of the devices with different channel lengths for the two cases where self-heating thermal model is active and inactive. We can make three conclusions by observing these plots. Firstly, the input referred thermal noise voltage increases as the gate length is decreased, even without considering self-heating effects. This is because the transconductance decreases with decreasing channel length due velocity saturation although each device has the same g_{d0}. Furthermore, the thermal noise current increases as the gate length decreases due to hot electrons and velocity saturation. When we consider the self-heating effects, we see that the augmentation in the input referred thermal noise voltage is even larger with the decreased channel length. This is because the thermal noise current is further increased due to the larger temperature of the device and the transconductance is further decreased due to lower mobility. Secondly, the two cases, where the self-heating effects are considered and ignored, result in different input referred thermal noise voltage characteristics. The difference between two cases become more observable as the gate length is decreased and V_{DS} is increased. For large V_{DS} values on the 30 nm device, the error can reach up to 35% under $V_G = 0.6$ V. The error would further increase as the gate voltage is increased, which is permitted by the technology until 1 V. When compared to thermal noise current, the error in input referred thermal noise voltage is twice larger (comparing Figure 4.6b and Figure 4.2d). This arises from the drop in g_m with self-heating. Our last observation is that the input referred noise voltage continuously decreases as V_{DS} increases when self-heating effects are ignored. However, in case of self-heating, it decreases in the beginning, then starts to increase as V_{DS} increases. Therefore, there is a point where input referred noise voltage is smallest, even for the long channel device. For obtaining the best noise figure, one should consider biasing the device at this point. However, ignoring the self-heating effects might result in miscalculation of the optimum bias point. It can be observed that this point shifts to left as the gate length decreases and the curvature of the curve increases, which limits the biasing of the device into a smaller region. Hence, to have the lowest possible input referred noise voltage, shorter channel devices should be biased closer to the linear region by preventing the undesired effects like excessive self-heating, hot electrons, high velocity saturation and avalanche multiplication as much as possible.

4.3 Flicker Noise

Flicker noise is the dominant noise of source MOSFET devices at low frequencies. Flicker noise is also referred as *1/f noise*, since it is almost inversely proportional to the frequency.

4.3.1 Theoretical Background and Related Work

Regarding the main source of flicker noise, there are different theories and observations over the decades. In fact, the shape and power of the noise has a strong dependency on the fabrication technology. Yet, there is still no single certain explanation for the phenomenon. The three widely approved theories are *carrier density fluctuations theory* [125] proposed by A. L. McWorther, *mobility fluctuations theory* [126] proposed by F. N. Hooge and the *unified theory*, which claims that the first two theories are both involved at the same time.

According to the theory of *carrier density fluctuations*, the carriers are randomly trapped and released by the traps located at the Si-SiO$_2$ interface [127]. This results in random fluctuations in the number of carriers. As a result of this, the channel current is modulated at random instants in time. The traps might be also located inside the oxide. The deeper the traps find themselves inside the oxide, the less likely an event of capturing a carrier becomes, which means that the deeper traps have larger average times (τ) between the captures and releases [128]. For a long channel device, by assuming a sufficiently large number of traps with a uniform distribution of τ, one gets a power spectral density which is proportional to $1/f$. However, in many cases the trap distribution is not uniform and the behaviour is not exactly in the form of $1/f$. According to the carrier fluctuations, the power spectral density of the gate referred flicker noise voltage is expressed as [129]

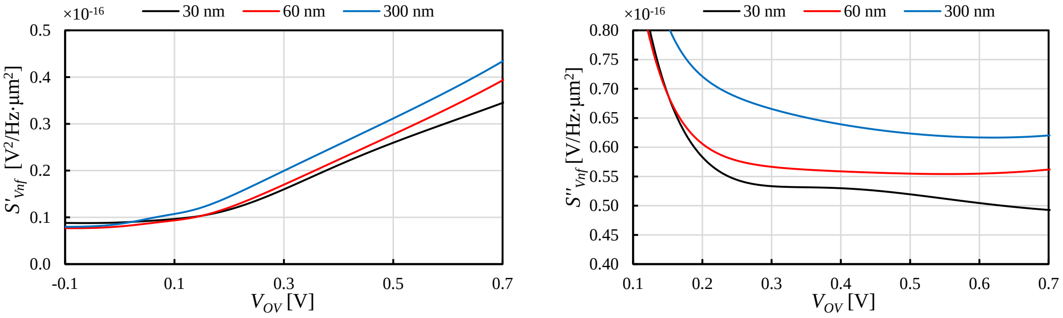
$$S_{Vnf} = \frac{K}{WLC_{ox}'^2}\frac{1}{f^\gamma} \tag{4.4}$$

where $0.7 \leq \gamma \leq 1.2$ depending on the trap distribution and C'_{ox} is the gate oxide capacitance per unit area. The resulting power spectral density of the drain current due to flicker noise is

$$S_{Inf} = \frac{K}{WLC_{ox}'^2}\frac{g_m^2}{f^\gamma}. \tag{4.5}$$

According to carrier fluctuations theory, flicker noise is independent of gate and drain voltages, which can be seen by observing (4.4) and (4.5). This will not be the case in mobility fluctuations theory.

Mobility fluctuations theory claims that flicker noise is in fact a bulk phenomenon arising from the fluctuations in the mobility due to lattice scattering. The power spectral density of the



(a) (b)

Figure 4.7 – (a) Flicker noise voltage normalized with respect to effective device area and (b) further normalized with respect to overdrive voltage.

input referred flicker noise voltage can be written as

$$S_{Vnf} = \frac{\alpha_1 q \frac{\mu_{1/f}}{\mu_{\text{eff}}}}{2WLC'_{ox}} \frac{(V_{GS} - V_T)}{f^{\gamma}} \tag{4.6}$$

where α_1 is a dimensionless flicker noise parameter constant, q is the elementary charge, $\mu_{1/f}$ is the effective $1/f$ noise mobility and μ_{eff} is the effective mobility [130]. It can be seen that contrary to carrier density fluctuations theory, the noise voltage is a function of the overdrive voltage $V_{OV} = V_{GS} - V_T$.

Finally, according to the *unified theory*, the flicker noise is arising from both carrier density fluctuations and mobility fluctuations [131, 132]. The random capture and release of the carriers affect the mobility in the channel due to Coulomb scattering and the power spectral density of the input referred noise voltage is expressed as [37]

$$S_{Vnf} = \frac{K(V_{GS})}{WLC'^{2}_{ox}} \frac{1}{f^{\gamma}} \tag{4.7}$$

where $K(V_{GS})$ depends on V_{GS} similar to (4.6) but not a direct linear function of $V_{GS} - V_T$. This model is widely used in many design kits and it provides good approximation with the experimental data.

On Figure 4.7a, flicker noise of the devices from Table 4.1 can be observed with respect the overdrive voltage. To be able to observe the curves of each device and to perform a fair comparison, flicker noise is normalized to the device area as

$$S'_{Vnf} = WL\, S_{Vnf}. \tag{4.8}$$

It can be seen that in weak inversion flicker noise is relatively constant. In strong inversion, flicker noise increases with the overdrive voltage; however, it is not perfectly linear. Therefore, the behaviour of the devices in the used 28 nm FDSOI technology is better approximated by

(4.7). To observe the influence of the overdrive voltage, the result of (4.8) is further normalized with respect to the overdrive voltage by

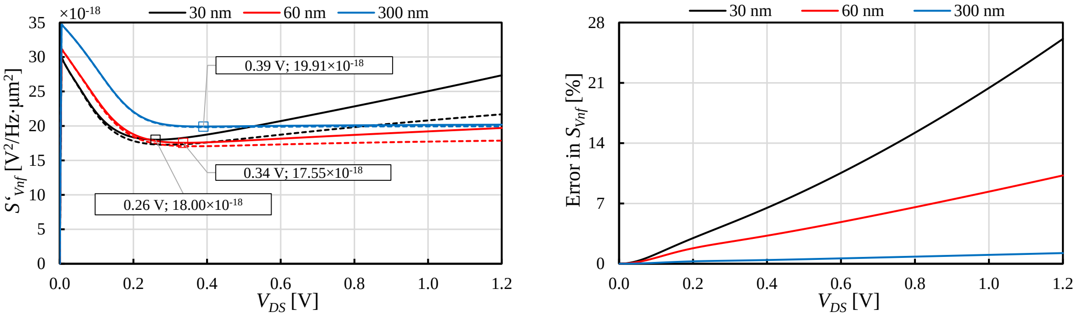
$$S''_{Vnf} = \frac{S'_{Vnf}}{V_{GS} - V_T}. \tag{4.9}$$

It can be seen on Figure 4.7b that the normalized flicker noise voltage is quite constant in strong inversion for the devices with different sizes.

Contrary to thermal noise, we do not see a direct effect of temperature on the flicker noise. Although not so many, there are couple of publications in the literature where the temperature dependence of flicker noise is reported. In fact, depending on the technology and the operating point, both positive and negative correlation is observed by different authors. In [133], it was shown that the drain voltage noise spectrum increases linearly with the temperature. This correlation is attributed to the energy dependent trap density. In [134] on the other hand, it was shown that the trap density increases with decreasing temperature. While this dependency is quite strong for the transistors of LOCOS technology, it is weak for the classical devices. In [135], measurements are performed on different devices with different gate lengths. It was observed that the flicker noise of the devices which show characteristics of carrier density fluctuations theory decrease with temperature. On the other hand, for the devices where the mobility fluctuations are dominant, an increasing flicker noise with respect to temperature is observed. In [136], it is shown that the slope of the spectra (γ) increases from 0.84 to 1.09 as the temperature is decreased from room temperature down to 30 K. In [137], it is also shown that γ has a similar temperature dependence for the temperature values lower than the room temperature and it increases as the temperature exceeds 350 K.

4.3.2 Flicker Noise Considering Self-Heating

For observing the effects of self-heating, the flicker noise of the devices that are reported on Table 4.1 is analysed. The same experiment of the thermal noise section is performed, where the gate voltage is kept constant and the drain voltage is swept. According to (4.7), the parameter with the strongest temperature dependence is the threshold voltage, which decreases with temperature. Consequently, one expects to observe a larger flicker noise voltage if the temperature increases. The flicker noise voltage can be observed on Figure 4.8a. It can be seen that the flicker noise voltage of the long channel device gradually decreases in linear region and is quite flat in saturation. The reason of gradual decrease is because as the device moves in saturation, the traps in the drain end of the device influence the device less due to fewer charges [37]. Once the device is in saturation, the charge distribution in the channel does not significantly change with drain voltage. Consequently, we observe a flat flicker noise voltage for any V_{DS} larger than 0.3 V in the long channel device. The situation is slightly different in short channel devices. For the 30 nm device, an increase in the flicker noise can be observed as V_{DS} increases in saturation. This is due to drain induced barrier lowering (DIBL). For short channel device, the threshold voltage decreases as the drain current increases [138]



(a)

(b)

Figure 4.8 – (a) flicker noise voltage at the gate at 20 MHz according to (4.8) and (b) its calculation error when self-heating effects are ignored. The devices are biased with V_G = 0.6 V.

due to the lowered barrier induced by the large potential of the drain. This effect becomes stronger as the gate length gets smaller. Consequently, the flicker noise voltage becomes larger with increasing the drain voltage in devices that have a small gate length. With the self-heating effects, flicker noise voltage is larger especially at large V_{DS} due to the larger temperature of the device. Naturally, it is even larger for the short channel devices due to their larger heat density and the calculation error reaches up to 25% under V_{DS} = 0.6 V.

4.4 Conclusion

The thermal and flicker noise performance of MOSFET devices in 28nm FDSOI technology is examined by considering their self-heating effects with a compact self-heating thermal model. It is shown that the thermal noise current (S_{Int}) of short channel devices increases significantly with self-heating of the device. For medium V_{GS} drive, the increase can be around 17%. In addition to this, the input referred thermal noise voltage ($S_{Vn,i}$) also increases due to self-heating. Besides, the increase in $S_{Vn,i}$, which can go up to 35% at medium V_{GS} bias, is even more prominent than S_{Int} mainly due to the drop in g_m at larger temperature values because of self-heating. The situation is also similar in flicker noise voltage with an increase of around 25%,

5 Self-Heating Aware Design of LNAs with Short Channel Devices

In Chapter 4, it is shown that the self-heating effects increase the thermal noise significantly in FDSOI, especially in short channel devices. This situation would decrease the SNR and resolution of noise critical blocks. Low noise amplifiers are examples of such blocks where the most critical performance parameter is the noise figure. Consequently, excluding the self-heating effects in the design of LNAs in advanced technologies would cause incorrect results. In this chapter, we show the significance of considering the self-heating effects and the amount of deviation in the noise figure of LNAs.

The chapter starts with the details of the integrated inductor design in Section 5.1 which is followed by the self-heating aware design of common gate cascode LNAs in Section 5.2. In Section 5.3 the measured noise figure values along with some of the other measurement results are provided. The chapters ends with the conclusion in Section 5.4

5.1 Integrated Inductor Design

The design of a proper integrated inductor is quite critical for having the desired performance of LNA. Ideally, the impedance of an inductor is purely reactive and it linearly increases with frequency and its self-inductance:

$$Z_{Li} = j\omega L \tag{5.1}$$

(5.1) is the desired behaviour of integrated inductors. With this behaviour, it would be possible to design an LNA that can operate at any frequency or in a very wide frequency band. However, in reality integrated inductors have some resistance due to various loss mechanisms and contain some parasitic capacitance. These undesired effects are illustrated on Figure 5.1 and summarized in the following section.

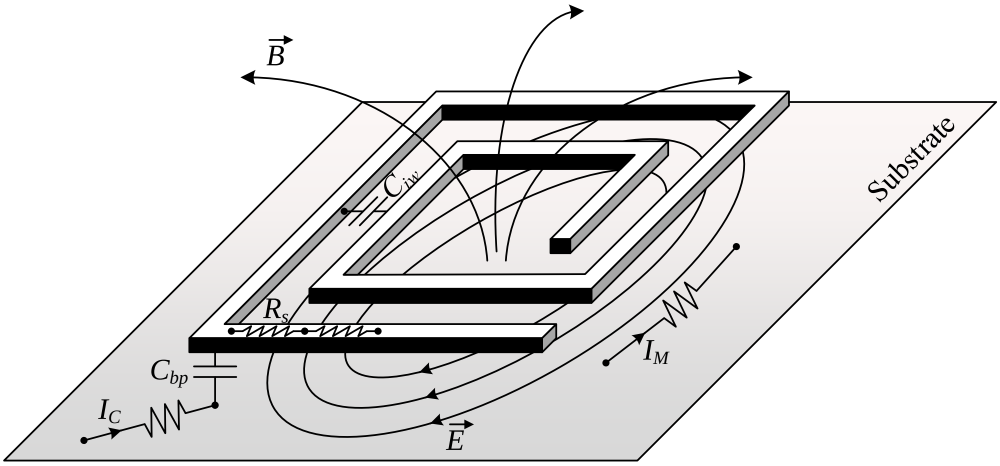


Figure 5.1 – Loss mechanisms of an integrated inductor.

5.1.1 Undesired Effects

Parasitic Capacitances

Bottom Plate Capacitance: Due to the parallel plate and the fringing fields between the winding and the closest conductive layer under the winding (which is the substrate on Figure 5.1), any integrated inductor contains a certain amount of bottom plate capacitance (C_{bp} on Figure 5.1). To decrease this capacitance, the width of the winding can be decreased and higher metal layers can be preferred for increasing the distance between the bottom plate. However, the former brings a trade-off in increased series metal resistance and the designer is limited by the number of available metal interconnect layers in the latter solution.

Inter-Winding Capacitance: Due to the parallel plate and the fringing field capacitances between the neighbouring winding metals, an integrated inductor with more than one turn contains a certain amount of inter-winding capacitance (C_{iw} on Figure 5.1). To decrease this capacitance, the spacing between the turns can be increased. However, this brings a trade-off where inductor area increases and inductance value decreases.

Loss Mechanisms

Winding Metal Resistance: The winding of integrated inductors are implemented with the existing interconnect metal layers. Due to the non-zero resistance of the interconnect metals (R_s on Figure 5.1), the inductor winding contains a certain amount of resistance which can be represented as a resistor in series with the inductor. Due to this series resistance, the inductor shows a non-zero voltage drop which can limit the voltage headroom and create mismatch for the topologies where one of the current mirror devices are degenerated with and inductor. However, the lowest value of this inductor is limited by the sheet resistance of a certain technology node. Using thicker metals can decrease the value of this series resistance.

Skin Effect: At higher frequencies, the electrons start to drift closer to the surface of the inductor. Consequently, the effective electron transport area decreases as the frequency increases, which in return increases the series resistance. For this reason, increasing the width of the winding (N) does not bring any benefit at higher frequencies and it should be set to its minimum by considering the skin effect.

Capacitive Coupling to the Bottom Plate: Due to the non-zero capacitance between the inductor and the bottom plate, C_{bp}, the voltage fluctuations on the inductor are coupled to the bottom plate of the inductor. Since the bottom plate resistance is neither zero nor infinity, the voltage fluctuations on the lower terminal of C_{bp} create a current flow in the bottom plate, which is represented by I_C on Figure 5.1. This undesired effect is pronounced more as the operating frequency increases.

Magnetic Coupling to the Bottom Plate: The changing amplitude and direction of the magnetic field created by the inductor creates an electromotive force on the bottom plate carriers. The finite resistance of the bottom plate results in eddy current, hence power dissipation and loss. Having a higher resistance bottom plate (which is the substrate) can decrease this effect.

5.1.2 Lumped Model of an Integrated Inductor

The integrated inductor structure on Figure 5.1 is a complex 3D distributed system composed of different resistive and capacitive components that were previously explained. For most realistic simulation results, one can use a field simulator that uses finite-element analysis methods. Although providing quite precise predictions for the electrical behaviour of a structure, these type of simulations are computationally expensive and take considerable time. Moreover, the generated models are not so explicit and they do not provide a simple insight about the behaviour. For this, a simple lumped model that gives acceptable results for a wide frequency range would be necessary and handy during the initial design steps. For creating such a model, it would be necessary to analyse each types of parasitic capacitance and loss mechanism separately and assign their overall effect to a lumped circuit element, if it is possible.

In our analysis, we will assume that one of the inductor terminals is connected to the AC ground so that the extracted lumped model is further simplified. We can make this assumption since each inductor in the LNAs analysed in this work have one of the terminals connected either to ground or to the supply.

Firstly, the inductance of an arbitrary structure, such as the one on Figure 5.1, can be found by applying Ampere's Law. The extracted inductance would give the simplest ideal model (Figure 5.2a) the impedance of which is modelled by (5.1). As it is stated previously, this model does not contain any parasitic effect. For this reason, we add the lumped parasitics to this inductor. Figure 5.1 shows that one of the terminals of the bottom plate capacitance is shorted to bottom conductive layer. This layer is always connected to the ground; therefore, the bottom plate capacitance can be modelled as a capacitor connected parallel to the inductor. It can be

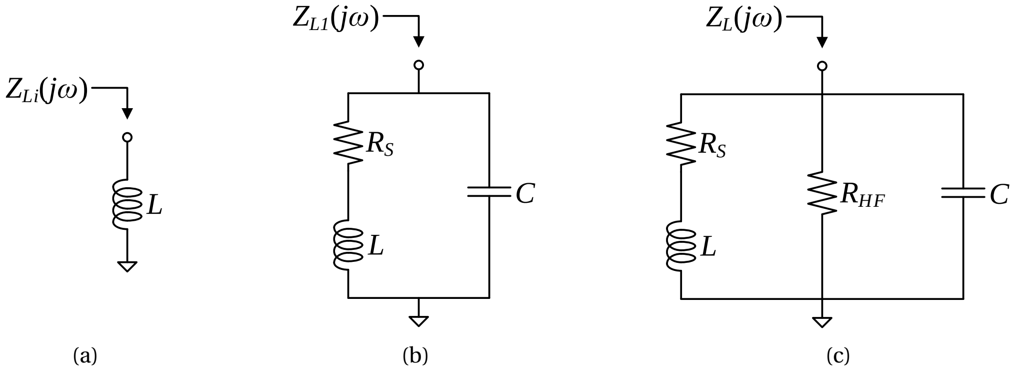


Figure 5.2 – Simple lumped models of an integrated inductor for different cases: (a) ideal, (b) first order lumped model where only the winding series resistance is taken into account, (c) more detailed model where other high frequency losses are considered.

shown that the inter-winding capacitance can also be modelled as a parallel capacitor [139]. Consequently, all the parasitic capacitors can be modelled as lumped parallel capacitor, which is shown on Figure 5.2b as C. As it is previously explained, the resistivity of the winding metal brings a series resistance to the inductor. By adding this series resistor, R_S, we can get a first order approximation for an integrated inductor shown on Figure 5.2b.

The first order model on Figure 5.2b provides a perfect estimation of the integrated inductor behaviour for lower frequencies because the rest of the losses start to manifest themselves after a certain frequency. Therefore, the model of Figure 5.2b is quite fundamental because it provides the ultimate performance that an integrated inductor can possess. In other words, an integrated inductor that has a parasitic capacitance of C and a series resistance of R_S cannot provide a closer behaviour to ideal than the one on Figure 5.2b *at any frequency*. Yet, C and R_S can be tuned properly for a more ideal behaviour at the resonance frequency ω_0. For this reason, one would target to converge to the behaviour of the first order model, while considering the high frequency losses and using a more detailed model of Figure 5.2c, which will be explained shortly.

Before providing the model of Figure 5.2c, let us analyse the first order circuit on Figure 5.2b in more detail. The quality factor of the series R_SL section of the circuit on Figure 5.2b at the resonance frequency is

$$Q_S = \frac{\omega_0 L}{R_S} = \frac{1}{\omega_0 R_S C} \tag{5.2}$$

where the subscript S in Q_S denotes that the loss in the quality factor is due to the series resistance R_S.

The first order model can be represented by a parallel RLC tank (Figure 5.3) at the resonance frequency, ω_0, by performing impedance transformations for L and R_S. By using the quality

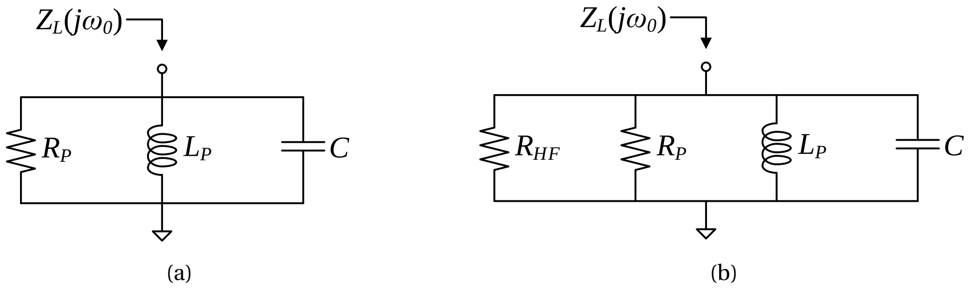


Figure 5.3 – Parallel RLC representations of the models on (a) Figure 5.2b and (b) Figure 5.2c

factor Q_S expressed in (5.2), we can write the parallel equivalent resistor and inductor as

$$R_P = R_S\left(1+Q_S^2\right) \tag{5.3}$$

and

$$L_P = L\left(1+\frac{1}{Q_S^2}\right). \tag{5.4}$$

The self-resonance frequency of the integrated inductor can be easily found by using the parallel RLC circuit on Figure 5.3a as

$$\omega_0 = \frac{1}{\sqrt{L_P C}} = \frac{1}{\sqrt{\left(1+\frac{1}{Q_S^2}\right)LC}}. \tag{5.5}$$

Finally, by using (5.2) and (5.5) the impedance of the integrated inductor on Figure 5.2b can be expressed as

$$Z_{L1}(j\omega) = \left(R_S + j\omega L\right)\Bigg\|\frac{1}{j\omega C} = \frac{\left(1+Q_S^2\right)\left(R_S+j\omega L\right)}{1+Q_S^2\left[1-\left(\frac{\omega}{\omega_0}\right)^2\right]+jQ_S\left(\frac{\omega}{\omega_0}\right)}. \tag{5.6}$$

The magnitude $Z_{L1}(j\omega)$ for the first order model is plotted for different R_S values on Figure 5.4a. It can be seen that the behaviour of the inductor impedance looks like a bell shape and it naturally becomes purely real and converges to R_P at ω_0. For $R_S = 0$, both Q_S and R_P goes to infinity, which can be also seen on Figure 5.4b. From a design point of view, it is desired to have the value of $R_P = Z_{L1}(j\omega_0)$ as large as possible. For instance, for load inductors, with a larger R_P the load resistance and the gain of the LNA can be controlled better. Similarly, for the source degeneration inductors, the input can be set equal to the characteristic impedance of the transmission line with a higher accuracy for a larger frequency band in order to decrease

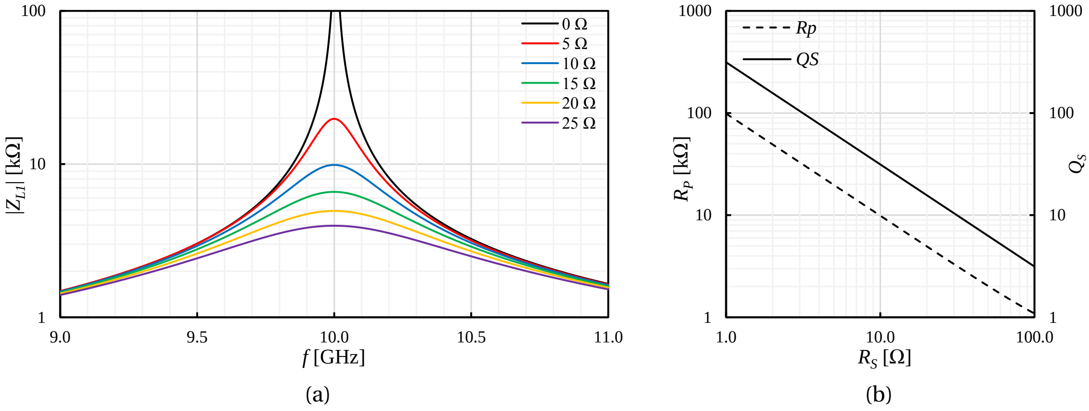


Figure 5.4 – Behaviours of $|Z_{L1}|$, R_P and Q_S with respect to the inductor series resistance R_S for $f_0 = 10$ GHz, $L = 5$ nH and $C = 50.7$ fF.

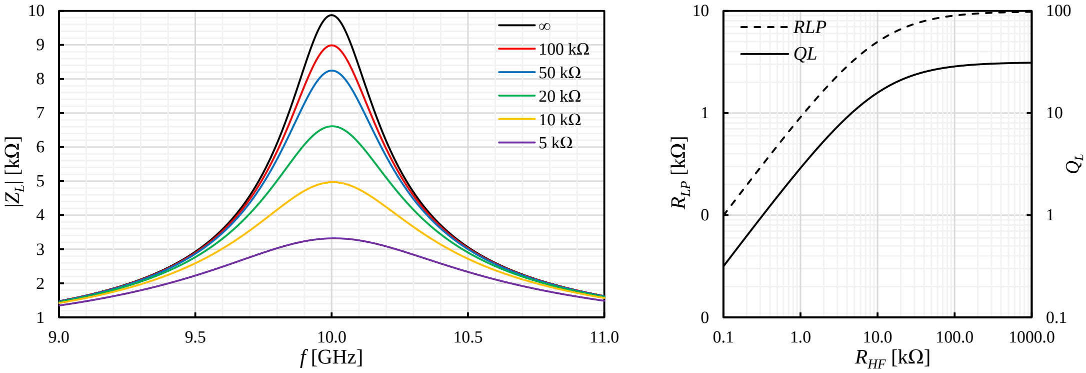
the reflections at the input. Therefore, R_S should be as low as possible so that the quality factor and the parallel resistance is higher. Nevertheless, even if R_S vanishes, due to skin effect at high frequencies, capacitive and magnetic coupling to the bottom plate, an additional parallel resistance effect is observed in the vicinity of the resonance frequency. For this, the model shown on Figure 5.2c is essential.

It can be seen that all of the loss mechanisms other than winding metal resistance is included in a single parallel resistance R_{HF}. However, such a simple model is questionable since the loss mechanisms arise in a complex 3D distributed system and they have strong dependency on both geometry and frequency. For example, very simply, the skin effect increases the value of R_S as the frequency increases, which means that it can be modelled better by having a frequency dependent R_S rather than a parallel resistor [140]. However, such a sophisticated model is more essential for a parametric inductor (such as a parametric cell in a design kit) where a large range of inductor geometry and frequency band is needed to be covered. In our case, we are more interested in the impedance behaviour of a fixed geometry inductor around the resonance frequency. A simple parallel resistor R_{HF} would give a quite close behaviour to the real integrated inductor since it can cover all high frequency losses in the vicinity of ω_0. For lower frequencies, it would be negligible due to the low impedance of L. Similarly, for higher frequencies, the same situation is observed due to the low impedance of C. The validity of this model will be further investigated in Section 5.1.3.

By using (5.6) the impedance of the inductor model on Figure 5.2c can be expressed as

$$Z_L(j\omega) = \big(R_S + j\omega L\big) \bigg\| \frac{1}{j\omega C} \bigg\| R_{HF} = Z_{L1}(j\omega) \big\| R_{HF} \,. \tag{5.7}$$

From (5.7) and Figure 5.2c, it can be seen that the effect of R_{HF} is a decrease in the impedance at any frequency since it creates a parallel conductance path. Therefore, the parallel equivalent



(a) (b)

Figure 5.5 – Behaviours of $|Z_L|$, R_{LP} and Q_L with respect to parallel high frequency resistance R_{HF} for $f_0 = 10$ GHz, $L = 5$ nH, $R_S = 10\ \Omega$ and $C = 50.7$ fF.

resistance at ω_0, let us call R_{LP} also decreases according to

$$R_{LP} = Z_L(j\omega_0) = R_P \parallel R_{HF} = \frac{R_P R_{HF}}{R_P + R_{HF}}. \tag{5.8}$$

Similarly, the overall quality factor of the integrated inductor at ω_0, Q_L also decreases according to

$$Q_L = \left(Q_S^{-1} + Q_{HF}^{-1}\right)^{-1} = \frac{Q_S Q_{HF}}{Q_S + Q_{HF}}. \tag{5.9}$$

(5.8) and (5.9) can be observed on Figure 5.5b. It can been that R_{LP} saturates to R_P as R_{HF} increases; which means that the maximum parallel equivalent resistance, R_{LP}, that can be obtained is limited by the series resistance, R_S. Consequently, one needs to have a sufficiently large R_{HF}, meaning that the high frequency losses of an integrated inductor should be as low as possible, at least not significant compared to the losses due to R_S. For examining the quality of an integrated inductor in this regard, we can define a figure of merit parameter, **P**, which shows the performance of an integrated inductor under the influence of high frequency losses. We define **P** as

$$\mathbf{P} = \frac{\text{low frequency losses}}{\text{all losses}} = \frac{R_P^{-1}}{(R_P \parallel R_{HF})^{-1}} = \frac{R_{HF}}{R_P + R_{HF}}. \tag{5.10}$$

(5.10) indicates that the contribution of the high frequency losses are not as significant as **P** approaches to one. On the other hand, as the high frequency losses become more prominent, **P** goes down to zero. While examining the performance of an integrated inductor with respect to **P**, one needs to be careful and compare two inductors that have the same resonance frequency since for different resonance frequencies, the influence of high frequency losses would be different.

5.1.3 Parametric Cell (pCell) for Integrated Inductors

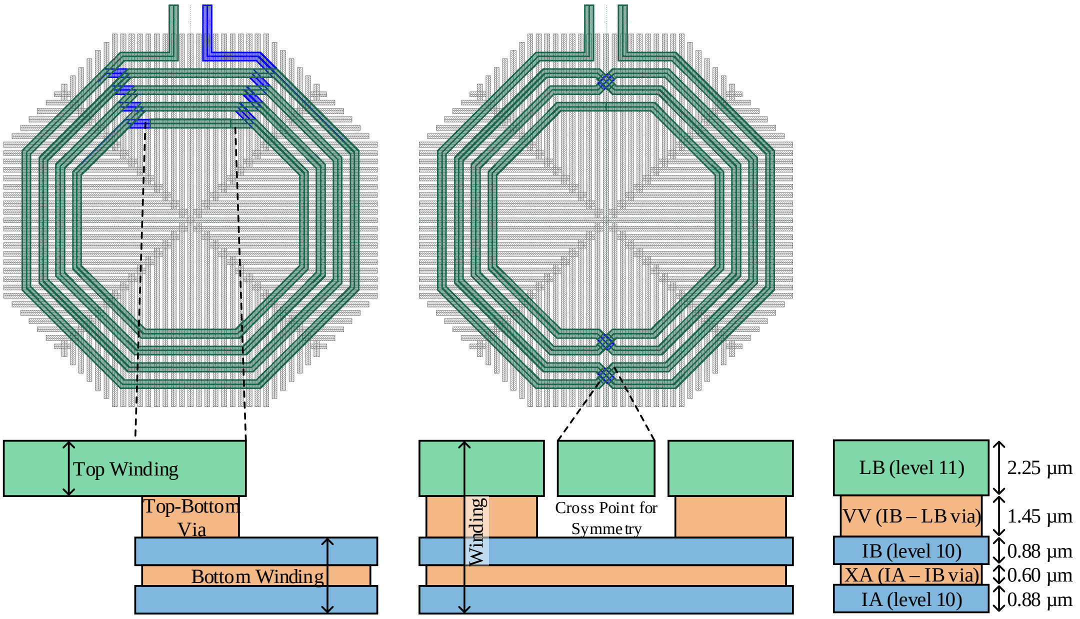
Drawing the layout of an integrated inductor that provides the DRC criteria is a very time consuming task although the spiral looks like a simple metal routing. The reasons can be sorted as the filling metals for minimum density rules, the shielding grid for decreasing the magnetic coupling losses, parallel vias between the windings, multi-stacked structures implemented with different metal layers, symmetric corners of an octagonal winding etc. Moreover, as it has been already indicated, simulating the layout of an inductor is computationally heavy and takes quite a lot of time. Therefore, most of the design kit providers provide a parametric cell (pCell) for different types of inductors which are already optimized according to the technology parameters. However, in some cases, even the provided inductor pCells might not provide the desired performance in a design. At this point, one needs to design and optimize a custom designed inductor.

In our case, the area of the implemented LNAs were quite restricted and the inductor pCells provided by the design kit supplier were found to be quite large. Therefore, we were obliged to design an in-house inductor by the following steps:

- Drawing the layout of the inductor
- Performing a finite element RF simulation for obtaining s-parameters
- Generating a circuit model for circuit level simulations
- Performing circuit level simulations
- Extracting the parameters for fitting to the model on Figure 5.2c

To get the best performance, firstly, combinations of different geometrics (square, octagonal, stacked, with/out shield, symmetric etc.) were simulated and characterized. The best performance was obtained by having either a two level **stacked** or a **symmetric** geometry with a shielding at the bottom plate (Figure 5.6). After this point two different pCells were created for these two geometries by writing scripts using Cadence SKILL so that a specific geometry can be tuned for the desired design specifications like area, ω_0, minimum R_{LP}, minimum Q_L etc. The adjustable parameters of the designed pCells are summarized on Table 5.8.

For the two-stack inductor, a thick top metal layer (LB) is used for the top winding. For the bottom winding, two metal layers (IA and IB) connected with a continuous via (XA) are used. Since the thickness of LB, 2.25 µm, is more than two times larger than the thickness of IA and IB, 0.88 µm, two parallel metal layers are necessary to have similar resistance both for the top and the bottom windings. Nevertheless, the resistance of the top winding is still larger than the resistance of the bottom winding. This means that the inductor is asymmetric with respect to its terminals. As a result of this, the terminal connected to the bottom winding would give a lower quality factor due to larger RC time constant [141]. For this, it is preferred to connect the bottom terminal to an AC ground. For having a better electrical symmetry, the



(a) Two-stack (b) Symmetric (c) Routing layers

Figure 5.6 – Layout screen-shots and cross-section views of the (a) two-stacked and (b) symmetric inductor pCells synthesized in 28 nm FDSOI technology.

symmetric geometry can be used, which is shown on Figure 5.6b. In this implementation, the three highest metal layers (LB, IB, IA) and the continuous vias (VV, XA) in between them are used for implementing a single winding. Like this, the series resistance is decreased compared to the two-stack implementation. However, the inductance of the symmetric implementation is lower than the two-stack one due to the smaller number of turns.

With the pCells of two different geometries, the validity of the lumped circuit model on Figure 5.2c was tested by synthesizing and simulating symmetric and two-stack inductors. The default parameters shown on Table 5.1 were used except for the two-stack inductor's winding which is set to eight. The s-parameters of the synthesized inductor layouts were generated with Momentum tool of Keysight Technologies. By simulating the generated circuit models of the RF simulator, the main parameters of the lumped circuit, L, C, R_S, and R_{HF} were extracted. The behaviours of the two models, which are RF simulation and lumped, are compared on Figure 5.7. The parameters of the two inductors are summarized on Table 5.2.

It can be seen that the two curves, which are the impedance of the integrated inductors obtained by RF simulations and the impedance of the model of Figure 5.2c, match quite well. The error between these two curves can be seen on the bottom graphs of Figure 5.7. Especially at the resonance frequency, there is almost no error. For the higher and lower frequencies the error is less than 10% and it increases as the frequency increases because R_{HF} is a frequency independent resistor. Its value is set according to the amount of loss at f_0 and the losses increase with the frequency. This can also be seen from Figure 5.7b where the

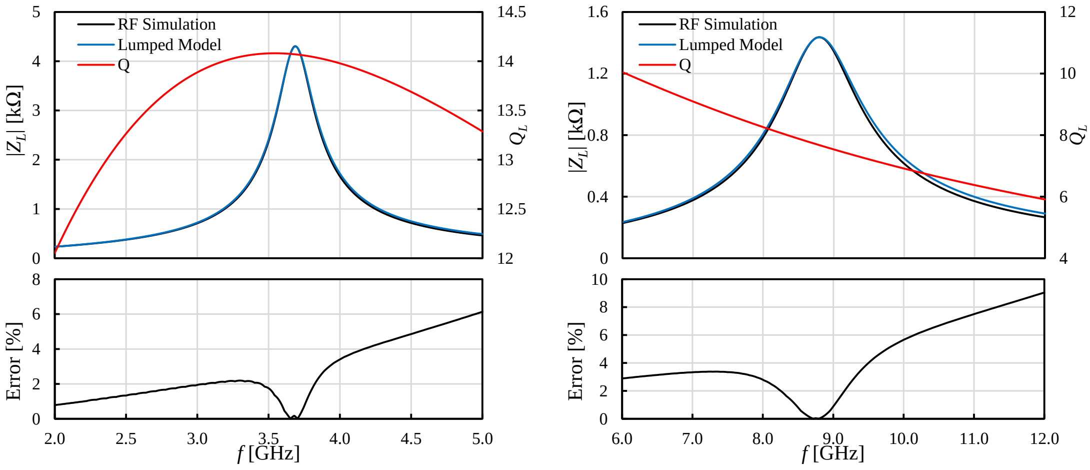
Table 5.1 – Adjustable parameters and their explanations of the designed inductor pCells that are shown on Figure 5.8. Important note: all dimensions are shrinked by 0.9 in fabrication.

Param.	Default	Explanation
N	4	number of windings
R	200 µm	outer radius of the inductor
W	5 µm	width of the winding
S	5 µm	spacing between two neighbour windings
S_T	20 µm	spacing between the P and N terminals
P_{sh}	5 µm	distance between the centres of two neighbouring shield metals (i.e. shield pitch)
W_{sh}	3 µm	width of the shield metal
E_{sh}	11 µm	extension of the shield from the edge of the outermost winding
L_{fill}	0.88 µm	edge length of the square shaped fill metals
P_{fill}	0.45 µm	empty distance between the centres of two neighbouring fill metals (i.e. fill pitch)
W_{fo}	5 µm	width of the outer fill region
D_{fo}	10 µm	empty distance between the outer fill region and the outer edge of the outermost winding
D_{fi}	5 µm	empty distance between the inner fill region and the inner edge of the innermost winding

Table 5.2 – Extracted lumped circuit parameters of the two-stack and symmetric integrated inductors.

Parameter	Two-stack	Symmetric
f_0	3.69 GHz	8.79 GHz
L	13.2 nH	3.39 nH
C	141 fF	96.8 fF
R_S	10.4 Ω	2.54 Ω
R_{HF}	8.26 kΩ	1.60 kΩ
Q_L	14.1	7.68
P	0.48	0.10

impedance of the RF simulation model is smaller than the lumped one. However, this error is not critical for the analysis of LNAs since the gain is significantly low at those frequencies. The resonance frequency of the two-stack inductor is lower than the one of the symmetric one. This is because of the large inductance of the two-stack inductor due to its larger number of turns. The quality factor and **P** of the two-stack inductor is larger than the symmetric one. This can be explained by the lower resonance frequency and the resulting lower high frequency losses of the two-stack inductor compared to the symmetric one when both have the same area. Therefore, for lower frequencies, two-stack inductor is a better candidate. Finally, the



(a) Two-stack

(b) Symmetric

Figure 5.7 – Comparison of the behaviours of the RF simulation results and lumped model of (a) a two-stack and (b) a symmetric integrated inductor.

series resistance, R_S, of the symmetric inductor is smaller than the one of the two-stack one. Therefore, it can be preferred in the applications where voltage drop on the inductor is critical. However, this might bring an area trade-off.

Patterned Ground Shield

One possible way to decrease the losses due to the capacitive coupling to the substrate is to implement a highly conductive metal shielding between the substrate and the inductor. However, having a continuous highly conductive plane just below the inductor would create a low resistance path for the eddy currents and the shield would behave like a second inductor in the presence of a magnetic field created by the inductor. In this case, the inductor losses would increase significantly especially due to the magnetic coupling to the bottom plane. For this reason, it is preferred to implement a patterned shield where the resistance of the eddy current paths are increased by removing some portions of the shield perpendicular to the eddy current paths [142] as shown on Figure 5.8. Yet, decreasing the density of the shield would increase the losses due to higher capacitive coupling to the substrate.

For finding the best shield geometry, two experiments were performed. In the first experiment, the shield pitch of the inductor was swept from 0.5 µm to 50 µm while shield-width/shield-pitch ratio was kept constant at W_{sh}/P_{sh} = 0.6. The results of this experiment can be seen on Figure 5.9a. It can be seen that the performance of the inductor makes a peak around a pitch value of 5.5 µm. Further increasing P_{sh} decreases the quality factor abruptly since having wider shields creates larger areas for eddy currents. On the other hand, by decreasing the pitch, the distance between two neighbour shield metals also decreases. Like this, the capacitance between two shield lines becomes larger. Consequently, due to larger capacitive

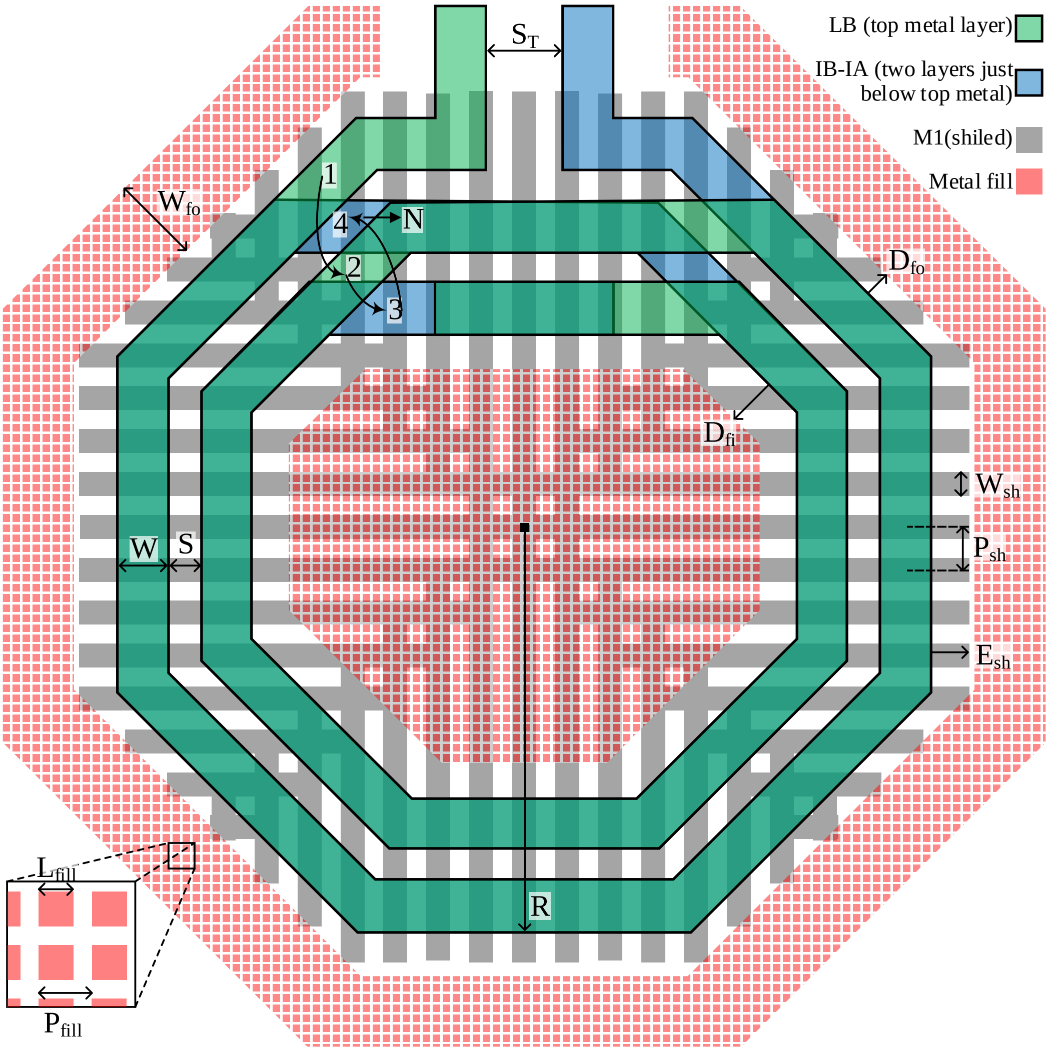
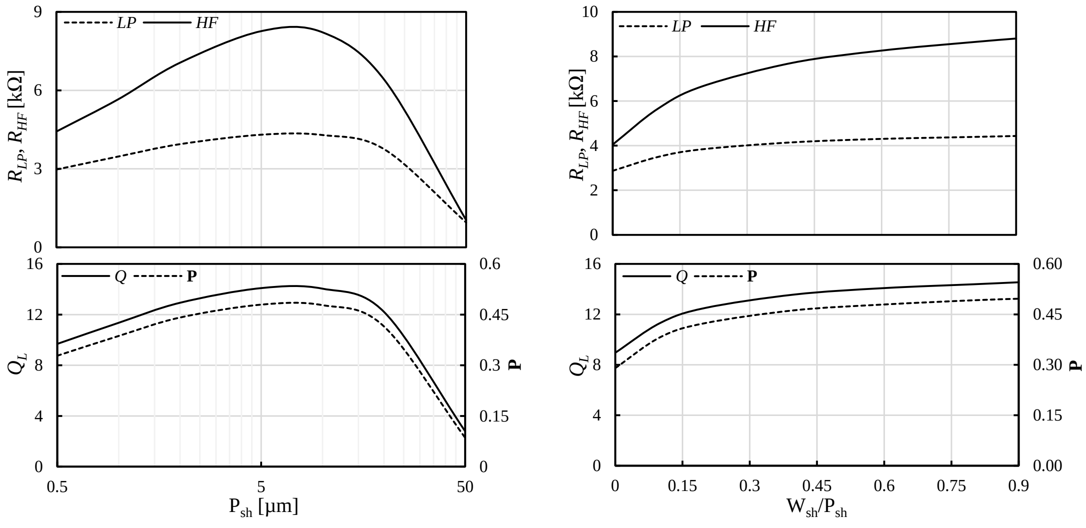


Figure 5.8 – Example of the integrated stacked inductor drawing and its adjustable parameters.

coupling between the shield metals, the patterns lose their high resistance function at high frequencies and the patterned shield behaves more like a continuous metal plane. As a result of this, magnetic coupling increases and the quality factor drops. Finally, it can be seen that the quality factor drops quite rapidly by increasing the pitch whereas its behaviour is relatively flat as the pitch is decreased. Therefore, it might be safe to choose a pitch value that is slightly smaller than the peak of the curve on Figure 5.9a so that an abrupt drop in the quality factor due to process variations is prevented.

After finding the optimum shield pitch, which is 5 µm, in the first experiment, a second experiment was performed where the shield width is swept from 0 µm to 4.5 µm while shield pitch was kept constant at P_{sh} = 5 µm. In this case, it can be seen that the quality factor increases as the W_{sh}/P_{sh} ratio is increased. This is because by having a larger W_{sh}/P_{sh}, the overall effect of the shielding increases and capacitive coupling to the substrate becomes less



(a)

(b)

Figure 5.9 – Behaviours of the inductor parameters as (a) the shield pitch of the inductor was changed while shield-width/shield-pitch ratio was kept constant at $W_{sh}/P_{sh} = 0.6$, (b) shield width was changed while shield pitch was kept constant at $P_{sh} = 5\mu m$.

effective. However, having $W_{sh}/P_{sh} = 1$ creates a continuous shield and this is not desired due to the reasons that are previously explained.

5.2 Common Gate Cascode LNA

The main noise source of the common gate cascode LNA topology (Figure 5.10) is the M1 device since the noise contribution of the cascode device is insignificant. For this reason, this topology is quite interesting in terms of observing the self-heating effects on the noise figure since the heat density and the thermal noise of M1 can be easily modulated by the cascode bias voltage, V_{B2}, without significantly affecting the other design parameters. Therefore, we focus on the design of the common gate cascode LNAs by providing a detailed design guideline while considering the self-heating effects.

5.2.1 Design Parameters

The main design parameters of the common gate cascode amplifier are the input impedance, output impedance, voltage gain, transconductance and the noise figure. Since most of the parameters are limited by the matching criteria, the design is not so flexible. The following explains the details of tuning each parameter and the performance limitations in short and long channel implementations.

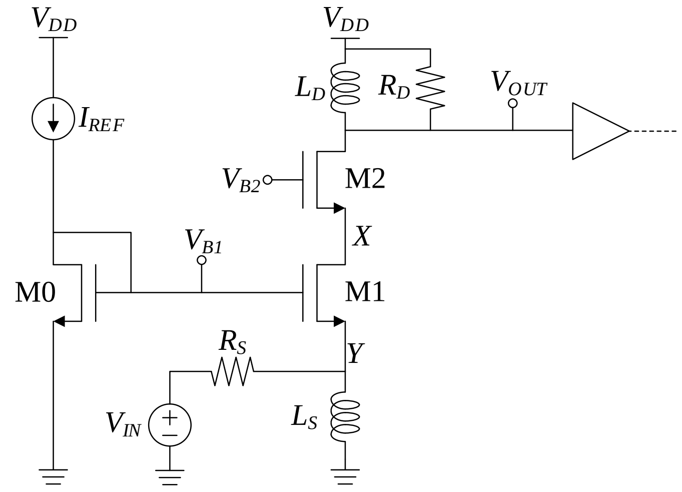


Figure 5.10 – Schematic of the common gate cascode LNA

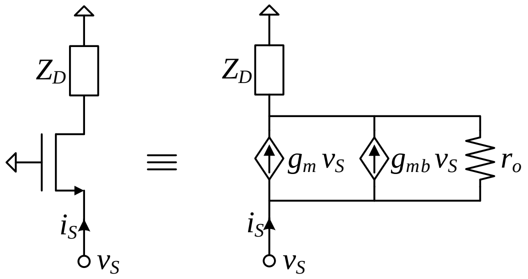


Figure 5.11 – Small-signal AC model used for finding the impedance seen from the source of a MOSFET while the drain is connected to an arbitrary impedance.

Load Impedance

In many integrated inductor implementations, the equivalent total parallel resistance of the load inductor at resonance frequency is too large to obtain the desired load resistance and voltage gain. Therefore, a parallel resistor is included for decreasing the load impedance and having a stronger control on the voltage gain. In this case, the load impedance can be expressed as

$$Z_L(j\omega) = Z_{LD}(j\omega) \, \Big\| \, R_D \, , \tag{5.11}$$

where $Z_{LD}(j\omega)$ is the impedance of the load inductor L_D and can be approximated by using (5.7). At f_0, the output impedance is purely real. By using (5.3) and (5.8) it can be expressed as

$$R_L \triangleq Z_L(j\omega_0) = R_{LP} \,\|\, R_D = R_S\left(1 + Q_S^2\right) \Big\| \, R_{HF} \, \Big\| \, R_D \, . \tag{5.12}$$

To have a more precise value for the voltage gain, it is always desired to have the first two terms much larger than R_D.

Input Impedance

At the resonance frequency, the real part of the input impedance of the LNA has to be set equal to the characteristic impedance of the transmission line, while the imaginary part is naturally zero. For this, it is necessary to find the expression of the impedance seen from node Y looking into the source of M1. Let us first find the impedance seen from node X looking into the source of M2. To find that, we can use the simplified small-signal AC circuit shown on Figure 5.11 where the drain of the device is connected to an arbitrary impedance.

The impedance seen from the source of the device on Figure 5.11 is equal to

$$Z_S = \frac{v_S}{i_S} = \frac{Z_D + r_o}{1 + \left(g_m + g_{mb}\right) r_o} = \frac{Z_D + r_o}{A_{vg}}, \tag{5.13}$$

where A_{vg} is the intrinsic gain of a device in common gate configuration which is calculated as

$$A_{vg} = 1 + \left(g_m + g_{mb}\right) r_o. \tag{5.14}$$

By applying (5.13) on the circuit of Figure 5.10, the impedance seen from node X looking upwards, $Z_{X,up}$, can be written as

$$Z_{X,up}(j\omega) = \frac{Z_{LD}(j\omega) \parallel R_D + r_{o2}}{1 + \left(g_{m2} + g_{mb2}\right) r_{o2}} = \frac{Z_{LD}(j\omega) \parallel R_D + r_{o2}}{A_{vg2}}. \tag{5.15}$$

At resonance frequency, Z_{LD} is equal to R_{LP}, therefore the input impedance becomes purely real and it is calculated as

$$R_{X,up} = \frac{R_L + r_{o2}}{1 + \left(g_{m2} + g_{mb2}\right) r_{o2}} = \frac{R_L + r_{o2}}{A_{vg2}}. \tag{5.16}$$

For a common gate LNA without a cascode device, the input resistance is equal to $R_{X,up}$, which is shown by (5.16). For a long channel device, we can assume $r_o \gg R_L$ and $g_m r_o \gg 1$; therefore, the input impedance can be approximated by

$$R_{X,up} \approx \frac{1}{g_{m2} + g_{mb2}}. \tag{5.17}$$

However, for short channel devices, this assumption does not hold. The input resistance increases as the gate length decreases and it becomes more dependent on R_L [139]. For this reason, the cascode topology is preferred in advanced technologies. However, addition of a series device brings a trade-off in the voltage headroom.

To find the input impedance, Z_{in}, of the common gate cascode LNA, (5.16) can be recursively used. Hence Z_{in} is calculated as

$$\begin{aligned} Z_{in}(j\omega) &= \frac{Z_{X,up}(j\omega) + r_{o1}}{1 + \left(g_{m1} + g_{mb1}\right) r_{o1}} = \frac{Z_{LD}(j\omega) \parallel R_D + r_{o2} + \left[1 + \left(g_{m2} + g_{mb2}\right) r_{o2}\right] r_{o1}}{\left[1 + \left(g_{m1} + g_{mb1}\right) r_{o1}\right]\left[1 + \left(g_{m2} + g_{mb2}\right) r_{o2}\right]} \\ &= \frac{Z_{LD}(j\omega) \parallel R_D + r_{o2}}{A_{vg1} A_{vg2}} + \frac{r_{o1}}{A_{vg1}}. \end{aligned} \tag{5.18}$$

At the resonance frequency, the input impedance can be written as:

$$R_{in} = \frac{R_L + r_{o2} + \left[1 + \left(g_{m2} + g_{mb2}\right) r_{o2}\right] r_{o1}}{\left[1 + \left(g_{m1} + g_{mb1}\right) r_{o1}\right]\left[1 + \left(g_{m2} + g_{mb2}\right) r_{o2}\right]} = \frac{R_L + r_{o2}}{A_{vg1} A_{vg2}} + \frac{r_{o1}}{A_{vg1}}. \tag{5.19}$$

It can be seen that for the cascode case, the load resistance is divided by the multiplication of the intrinsic gain of two devices. Therefore, for long channel devices the first term in (5.19) can be ignored and the input resistance can be approximated as

$$R_{in} \approx \frac{1}{g_{m1} + g_{mb1}}. \tag{5.20}$$

Nevertheless, in advanced technologies, the effect of the load impedance can be still observed at the input impedance, especially when devices with shortest channel length are used. For example, in 28 nm FDSOI technology, the intrinsic gain of a short channel device hardly exceeds 7-8. Moreover, the output resistance of the same device with a transconductance of 20 mA/V is around 250 Ω, which is even smaller than R_L. For instance, a voltage gain of 20 dB would necessitate an R_L value of at least 1 kΩ. In this case, R_L and r_{o2} together would bring an additional contribution of roughly 20 Ω ≈ (1000 Ω + 250 Ω)/(8 × 8), in addition to the transconductance of the input device M1, which is the main design parameter of the input impedance. Using a second cascade can further increase the isolation; however, this is not possible in advanced technologies due to low voltage supply values. Consequently, it is necessary to consider the additional undesired contribution arising from R_L and r_{o2}, and tune the parameters accordingly. To characterize how much the input is isolated from the load and the cascode device (M2), by using (5.19), we define the two undesired contributors of R_{in} in as

$$R_{RL} \triangleq \frac{R_L}{A_{vg1} A_{vg2}} \tag{5.21}$$

and

$$R_{ro2} \triangleq \frac{r_{o2}}{A_{vg1} A_{vg2}} \tag{5.22}$$

where R_{RL} represents the contribution of the load resistance R_L at the resonance frequency and R_{ro2} represents the contribution of r_{o2}. The last term in (5.19) is the desired contributor of R_{in} and it is represented by R_{gm} as follows:

$$R_{gm} \triangleq \frac{r_{o2}}{A_{vg1}}. \tag{5.23}$$

To examine how isolated the input impedance is, we define input impedance isolation parameter, $\mathbf{I}_{Rin}$, as

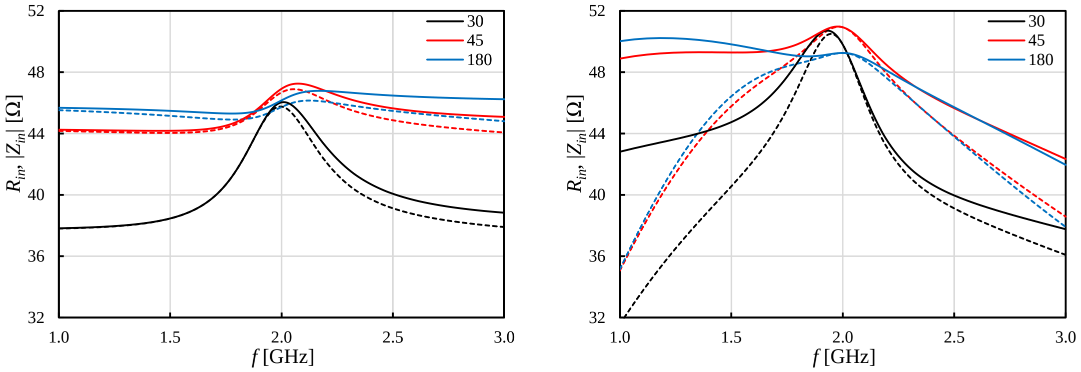
$$\mathbf{I}_{Rin} \triangleq \frac{R_{gm}}{R_{gm} + R_{RL} + R_{ro2}} = \frac{A_{vg2} r_{o2}}{A_{vg2} r_{o2} + r_{o2} + R_L}. \tag{5.24}$$

It can been that as $\mathbf{I}_{Rin}$ approaches to one, the input impedance becomes more isolated from the load and the cascode device. On the other hand, as it decreases, the input impedance becomes more dependent on the upper part of the circuit. Therefore, one needs to check $\mathbf{I}_{Rin}$ and target increasing its value as much as possible. One possible way of increasing $\mathbf{I}_{Rin}$ is to use long channel devices. However, this would either limit the maximum frequency of the system or the bandwidth. Consequently, in some applications, using shortest devices might be the only solution. However, for devices with very short channel, R_{RL} and R_{ro2} are comparable to R_{gm}. As a result of this, it might be impossible to have $\mathbf{I}_{Rin}$ close to one. At this point, one needs to further decrease R_{gm} so that R_{in} is still close to the characteristic impedance. This is only possible by having a larger g_{m1} compared to the long channel designs. However, increasing g_m would increase the power dissipation of the LNA.

The behaviour of $Z_{in}(j\omega)$ in the frequency range of interest can be understood by examining (5.18). Intuitively, the input impedance makes a peak at f_0 due to the contribution of the frequency dependent impedance of the inductor and stays relatively constant at the other frequencies. This peak is more observable especially for short channel devices due to their lower $\mathbf{I}_{Rin}$. To observe this, the input impedance for the three LNAs are plotted on Figure 5.12a.

It can be seen that the input impedance of the LNA with 30 nm devices have strong dependency on the load impedances. This dependency decreases with increased gate length. For the LNA with 180 nm devices, the input impedance is quite flat and independent of the frequency. This situation makes the design of LNAs with small gate length devices more difficult. Table 5.3 shows the values of different contributors at the resonance frequency. For a gate length of 30 nm, 25% of R_{in} is reflected either from the load resistance or from r_{o2}. Whereas, for the 180 nm LNA, this value is only less than 2%.

Until now, during the analysis of the input impedance of the common gate cascode LNA we have only focused on the impedance seen at Y while looking upwards into the source of M1 and ignoring the source inductor L_S. As long as the impedance of L_S is much greater than 50 Ω, its loading can be ignored. For this, the inductance value of L_S should be tuned in such a way that it resonates with the total equivalent capacitance seen at the input node, which contains the input pad capacitance, parasitic capacitance of L_S, parasitic capacitances of M1 and the interconnect parasitic capacitance. However, even if the input impedance is



(a)

(b)

Figure 5.12 – Total input impedance (dashed lines) and equivalent parallel resistance (continuous lines) of three different LNAs with different gate lengths 30 nm, 45 nm and 180 nm (a) while looking into the source of M1 (b) seen from the PCB.

Table 5.3 – Input resistance values of LNAs with different gate lengths and their contributors from different parts of the circuit at f_0.

Gate Length	r_{o1}	R_{in}	R_{RL}	R_{ro2}	R_{gm}	$\mathbf{I}_{Rin}$
30	0.34 kΩ	46.9 Ω	7.85 Ω	3.65 Ω	35.4 Ω	0.75
45	0.76 kΩ	47.7 Ω	2.27 Ω	2.25 Ω	43.2 Ω	0.90
180	3.76 kΩ	47.0 Ω	0.15 Ω	0.62 Ω	46.2 Ω	0.98

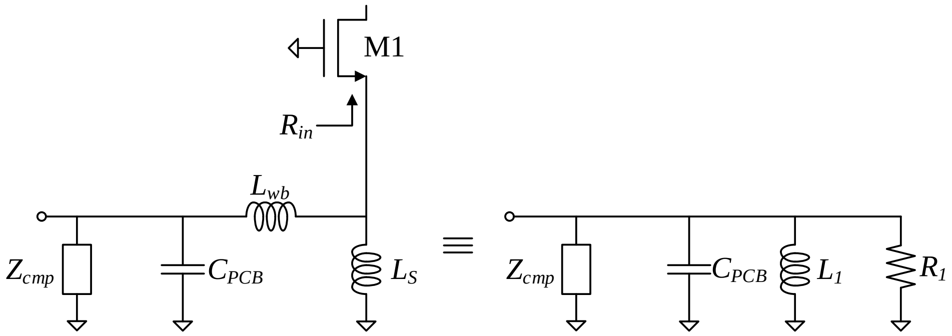


Figure 5.13 – Tuning the input impedance by considering the off-chip parasitics: wirebonding inductance, L_{wb} and parasitic capacitance of the PCB, C_{PCB}.

equal to 50 Ω by perfectly tuning L_S, it would be transformed to another value due to the wire-bonding inductance and maybe some capacitive loading of the PCB. The model including the wire-bonding inductance, L_{wb}, and the parasitic capacitance of the PCB, C_{PCB}, can be seen on Figure 5.13. It can be seen that the input impedance is no longer a purely real 50 Ω, but a 50 Ω resistor that is in series with the wire-bonding inductance. The series impedance $j\omega L_{wb} + R_{in}$ can be transformed to a parallel impedance at the resonance frequency. At f_0,

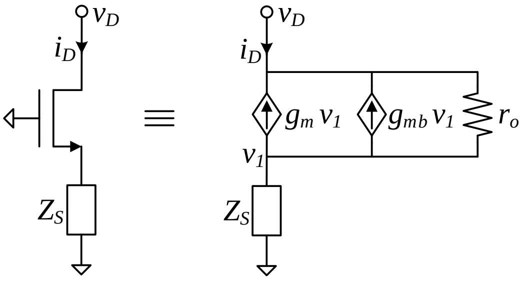


Figure 5.14 – Small-signal AC model used for finding the impedance seen from the drain of a MOSFET while the source is connected to an arbitrary impedance.

the parallel real and the imaginary part of the parallel impedance can be calculated as

$$R_1 = R_{in}\left[1+\left(\frac{\omega_0 L_{wb}}{R_{in}}\right)^2\right] \tag{5.25}$$

and

$$L_1 = L_{wb}\left[1+\left(\frac{R_{in}}{\omega_0 L_{wb}}\right)^2\right]. \tag{5.26}$$

It can be seen that the resistive part is larger than 50 Ω and how large it becomes depends on the value of the wire-bonding inductance. Obtaining a 50 Ω input impedance seen from the test board can be possible by tuning R_{in} to a value less than 50 Ω. Figure 5.12b shows what happens to the behaviour of Figure 5.12a when the parasitics of the PCB are taken into account and the wire-bonding inductance is assumed to be 1 nH. It can be seen that the resistive part is around 50 Ω, which is around 5-7% less on Figure 5.12a. Finally, the inductive part depends on the value of L_1 and C_{PCB} and their overall effect can be compensated by a parallel impedance Z_{cmp}.

Output Impedance

According to Figure 5.14, the impedance seen from the drain of a MOSFET when its source is connected to an arbitrary load is equal to

$$Z_D = \frac{v_D}{i_D} = \left[1+\left(g_m+g_{mb}\right)r_o\right]Z_S + r_o = A_{vg}Z_S + r_o. \tag{5.27}$$

Therefore, by using (5.27) recursively, the resistance seen from the output while looking into the drain of M2 is found as

$$\begin{aligned} R_{out,dw} &= \left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]R_S + \left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]r_{o1} + r_{o2} \\ &= A_{vg1}A_{vg2}R_S + A_{vg2}r_{o1} + r_{o2} \end{aligned} \tag{5.28}$$

Hence, by using (5.11) and (5.28), the output impedance, $Z_{out}(j\omega)$, can be calculated as

$$Z_{out} = Z_{LD}(j\omega) \,\|\, R_D \,\|\, R_{out,dw} \tag{5.29}$$

At resonance frequency, output impedance becomes purely real. Hence, the output resistance, R_{out} is equal to

$$R_{out} = R_L \,\|\, R_{out,dw} \tag{5.30}$$

For long channel devices, the impedance expression in (5.28) tends to be significantly larger than R_L. Therefore, for long channel devices, R_{out} can be approximated as

$$R_{out} \approx R_L \tag{5.31}$$

In 28nm FDSOI technology, even for very small channel devices, this approximation does not give bad results. For example $R_{out,dw}$ can easily exceed $5R_L$ for 30 nm gate length. Therefore, the loading of the devices does not influence the output impedance significantly.

Voltage Gain

The AC behaviour of the common gate cascade LNA at the resonance frequency can be understood by analysing the small-signal AC equivalent circuit shown on Figure 5.15.

The voltage gain, A_v, is equal to $\dfrac{v_{out}}{v_{in}}$. Hence, by analysing Figure 5.15 A_v can be written as

$$\begin{aligned} A_v &= \frac{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]R_L}{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]R_S+\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]r_{o1}+r_{o2}+R_L} \\ &= \frac{A_{vg1}A_{vg2}R_L}{A_{vg1}A_{vg2}R_S+A_{vg2}r_{o1}+r_{o2}+R_L}. \end{aligned} \tag{5.32}$$

When the input impedance is matched to the source impedance, $R_{in} = R_S$, we can use (5.19) and calculate the gain of the common gate cascode LNA as

$$A_v = \frac{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]R_L}{2\left\{\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]r_{o1}+r_{o2}+R_L\right\}} = \frac{A_{vg1}A_{vg2}R_L}{2\left[A_{vg2}r_{o1}+r_{o2}+R_L\right]} = \frac{R_L}{2R_S}. \tag{5.33}$$

The factor of two in the denominator is because the gain of the input stage is 1/2 due to the voltage division under perfect matching. (5.33) also shows that the gain of the common gate cascode amplifier is linearly proportional to the value of the load resistor. In Section 5.1, we have shown that the maximum value of R_L depends on the losses of the load inductor, L_D; therefore, proper design of the inductor is crucial especially for high frequencies.

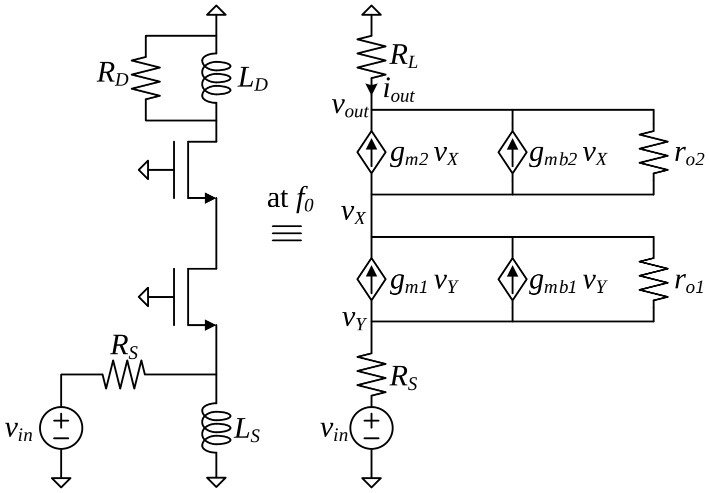


Figure 5.15 – Small-signal AC model of the common gate cascade LNA at f_0.

Transconductance

According to Figure 5.15, we calculate the transconductance of the common gate cascode LNA, $G_m = \left.\dfrac{i_{out}}{v_{in}}\right|_{R_L=0}$, at f_0 as

$$\begin{aligned} G_m &= \frac{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]}{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]R_S+\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]r_{o1}+r_{o2}} \\ &= \frac{A_{vg1}A_{vg2}}{A_{vg1}A_{vg2}R_S+A_{vg2}r_{o1}+r_{o2}}. \end{aligned} \tag{5.34}$$

When the input is matched to the source resistance, (5.34) simplifies to

$$G_m = \frac{\left[1+\left(g_{m1}+g_{mb1}\right)r_{o1}\right]\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]}{R_L+2\left\{r_{o2}+\left[1+\left(g_{m2}+g_{mb2}\right)r_{o2}\right]r_{o1}\right\}} = \frac{A_{vg1}A_{vg2}}{R_L+2\left[r_{o2}+A_{vg2}r_{o1}\right]}. \tag{5.35}$$

Under long channel assumption, G_m can be approximated as

$$G_m \approx \frac{1}{2\left(g_{m1}+g_{mb1}\right)}. \tag{5.36}$$

Note that the transconductance expression is also divided by two due to the voltage division at the input. Unfortunately, the transconductance does not have flexibility during the design since it value is determined by the value of the source resistance.

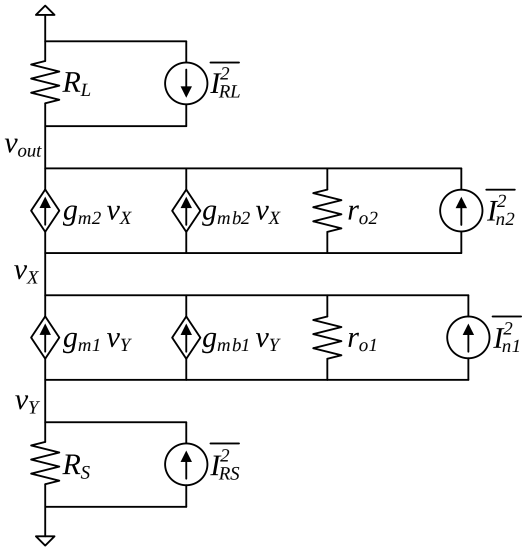


Figure 5.16 – Small-signal AC model of the common gate cascade LNA at f_0 with the independent noise sources.

Input Referred Noise Voltage and Noise Figure

The small-signal AC model of the common gate cascode LNA with its all critical noise sources at f_0 is shown on Figure 5.16.

The input referred noise voltage, $\overline{V^2_{in}}$ or $S_{v,in}$, can be determined by finding the output noise and dividing it by the square of the voltage gain of the amplifier. Hence, by analysing the model on Figure 5.16 and using (5.14), (5.32) and (5.34), the input referred noise voltage can be written as

$$\begin{aligned}\overline{V^2_{in}} &= R_S^2\overline{I^2_{RS}} + \frac{1}{G_m^2}\overline{I^2_{RL}} + \frac{r_{o1}^2}{A_{vg1}^2}\overline{I^2_{n1}} + \frac{r_{o2}^2}{A_{vg1}^2 A_{vg2}^2}\overline{I^2_{n2}} \\ &= 4kT\left(R_S + \frac{1}{G_m^2 R_L} + \frac{r_{o1}^2}{A_{vg1}^2}\gamma_1 g_{d01} + \frac{r_{o2}^2}{A_{vg1}^2 A_{vg2}^2}\gamma_2 g_{d02}\right)\end{aligned} \tag{5.37}$$

In many circuit simulators, only a single temperature value, T, which is the temperature of the chip, is used for all devices. Therefore, the individual temperature values of M1 and M2 devices are also set equal to each other and to the rest of the chip. However, as we have already shown, this assumption would be misleading in FDSOI. Therefore, we define the local temperatures of the M1 and M2 devices as T_1 and T_2 respectively, which might be different from the average temperature of the chip. By using the local temperature values of M1 and M2, (5.37) can be written in a more compact form as

$$\overline{V^2_{in}} = 4kT\left[R_S + \frac{1}{G_m^2 R_L} + \left(\frac{A_{vi1}}{A_{vg1}}\right)^2 \frac{\gamma_1}{\alpha_1}\frac{1}{g_{m1}}\left(\frac{T_1}{T}\right) + \left(\frac{1}{A_{vg1}}\frac{A_{vi2}}{A_{vg2}}\right)^2 \frac{\gamma_2}{\alpha_2}\frac{1}{g_{m2}}\left(\frac{T_2}{T}\right)\right] \tag{5.38}$$

where α is the ratio of the transconductance to the zero bias drain conductance,

$$\alpha = \frac{g_m}{g_{d0}} \tag{5.39}$$

and A_{vi} is the intrinsic gain of the device, which is expressed as

$$A_{vi} = g_m r_o. \tag{5.40}$$

The expression inside the brackets in 5.38 is equal to the equivalent noise resistance. The equivalent noise resistance of each term would define their contribution to the overall noise of the amplifier. It can be seen that the main contributor is the source resistance, R_S, and the thermal noise of M1. The noise contribution of M2 is divided by the square of the voltage gain of M1; therefore, it is quite negligible.

While analysing the noise in an RF system, we are mostly interested in how much noise an amplifier adds on top of the noise that is present at the input. For examining the noise contribution of an amplifier, we can use the noise factor, F, and the noise figure, NF. Noise factor is defined as the ratios of the SNR values of the signals at the output and at the input of the amplifier.

$$F = \frac{SNR_{out}}{SNR_{in}} \tag{5.41}$$

Noise figure is defined as

$$NF = 10\log(F). \tag{5.42}$$

It can be seen that both F and NF are figure of merits in terms of observing the additional noise contribution of the system normalized to the noise of the input signal. In our case, we can simply find the noise factor by dividing the input referred noise voltage by the noise voltage of the source resistor since the noise at the input is only due to the source resistance.

$$F_{LNA} = \frac{\overline{V_{in}^2}}{4kTR_S} \tag{5.43}$$

Therefore, by using (5.38) and (5.43), the noise factor of the common gate cascode LNA can be written as

$$F = 1 + \frac{1}{G_m^2 R_L R_S} + \left(\frac{A_{vi1}}{A_{vg1}}\right)^2 \frac{\gamma_1}{\alpha_1}\frac{1}{g_{m1}R_S}\left(\frac{T_1}{T}\right) + \left(\frac{1}{A_{vg1}}\frac{A_{vi2}}{A_{vg2}}\right)^2 \frac{\gamma_2}{\alpha_2}\frac{1}{g_{m2}R_S}\left(\frac{T_2}{T}\right). \tag{5.44}$$

It can be seen that the noise factor expression of the common gate cascode LNA is a function of temperature and might give different values depending on different self-heating conditions.

For long channel devices, we can assume $\alpha = 1$, $\gamma = 2/3$. Moreover, the self-heating effects would be less significant for long channel devices. Therefore, we can assume $T_1 = T_2 = T$. In addition to this, if we assume $g_m \gg g_{mb}$ and neglect the noise contribution of M2, (5.38) can be simplified as

$$F \approx 1 + \frac{2}{A_v} + \frac{2}{3}. \tag{5.45}$$

By observing (5.45) one can find the ultimate noise figure performance of a common gate cascode LNA which would be approximately equal to 10log(5/3) ≈ 2.22dB. Nevertheless, for short channel devices noise factor is mostly larger than this value mainly due to larger γ.

5.2.2 Optimum Bias Point for Lowest Noise Figure

In this section, the noise performance of the common gate cascode LNA is analysed both by including and excluding the self-heating effects and the minimum value of the noise figure is sought by changing different design parameters. Unfortunately, in this topology, most of the design parameters are set by the main design specifications like characteristic impedance, gain, power dissipation, frequency band etc. Therefore, there is not a large space of different design parameters to play with. One possible way of decreasing the noise figure is to use larger devices as it is shown by (5.45), although this would decrease the maximum operating frequency and bandwidth. Due to this reason, it can only be done for relatively low frequency applications. Changing the bias voltage of the cascode device, V_{B2}, can be another option. By changing V_{B2} it would be also possible to modify the power dissipation values of M1 and M2, hence their local temperature and individual degradation in their noise power. Moreover, by providing a sufficient headroom, it is possible to keep both devices in saturation. For this reason, the common gate cascode topology is quite interesting to see the self-heating effects on the noise performance. Another design parameter is the reference current, I_{REF}. Increasing the reference current might have two outcomes on the noise figure. First, the noise figure might decrease since transconductance of M1 increases. Second, the noise figure might increase because larger current results in larger heating. Considering the first statement, changing I_{REF} is not practically applicable since changing g_m would affect the input impedance and it has to be fixed for input matching. Nevertheless, its outcomes can still be studied to observe self-heating effects if the matching is not effected significantly. This can be provided by using velocity saturated devices where g_m becomes independent of the drain current. Like this, in the vicinity of the matching conditions, power consumption of the LNA can be modulated via I_{REF}.

To observe the effect of the listed parameters on the noise performance, three common gate cascode LNAs were designed and implemented with different gate lengths of 30 nm, 45 nm and 180 nm. The 180 nm LNA is used for observing the behaviour where self-heating effects are quite insignificant due to the large area of its active devices. On the other hand, with the 30 nm LNA it would be possible to observe the self-heating effects at its maximum due to its small active area. The 45 nm LNA is also implemented as a *back-up* for the short-channel case since the design of the 30 nm LNA contains many difficulties especially in terms of

Table 5.4 – Design parameters of three different common gate cascode LNAs implemented in 28 nm FDSOI technology.

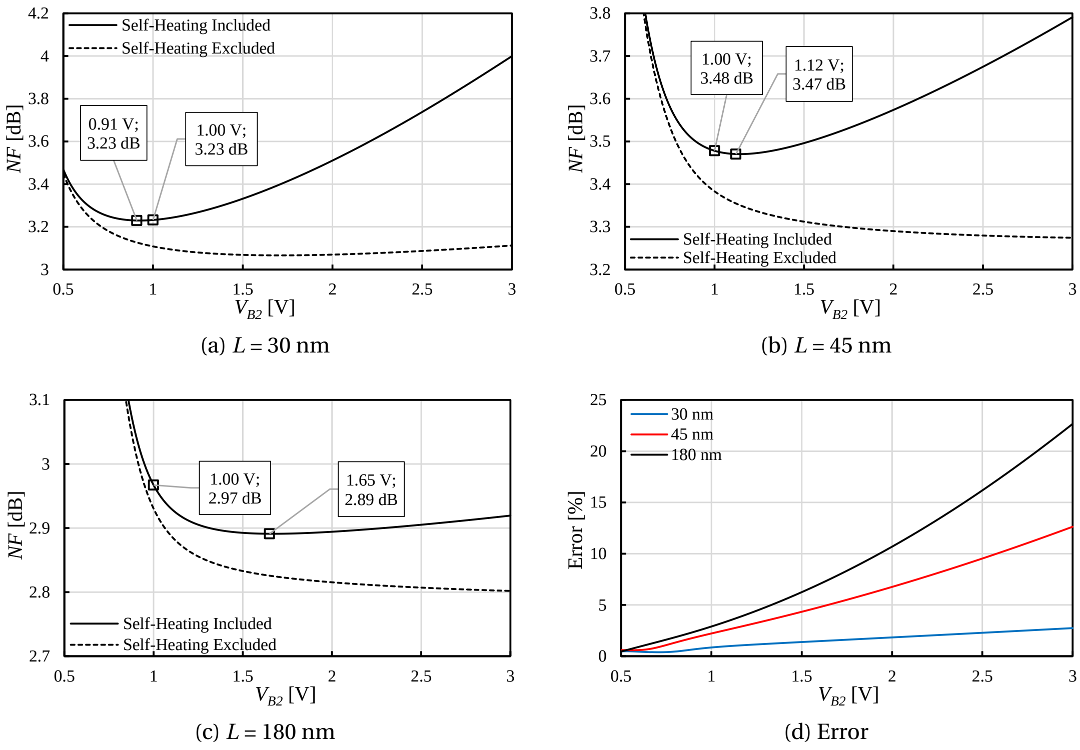
Implementation (M0, M1, M2)	$L = 30$ nm	$L = 45$ nm	$L = 180$ nm	Unit
I_{REF}	240	240	240	µA
W_0	4 × 420	4 × 370	4 × 680	nm
W_1	58 × 42	54 × 370	54 × 680	nm
W_2	58 × 420	54 × 420	54 × 1000	nm
V_{T1}	254	283	303	mV
V_{T2}	265	312	339	mV
g_{m1}	23.5	20.3	19.7	mA/V

setting its input impedance to 50 Ω due the previously explained reasons. Having these three implementations, it is possible to observe the impact of the self-heating effects on short-channel devices as well as the technology scaling. Some of the design parameters of the three different implementations are summarized on Table 5.4. The reference current for each implementation is set to 240 µA and the current amplification is set as 54/4 = 13.5. However, for the 30 nm implementation, this value is set to 54/8 = 14.5 for having a larger g_m for M1 device so that the input impedance can be set to 50 Ω by considering the effect of the load impedance on R_{in} according to (5.10). L_D and L_S of Figure 5.10 are implemented with custom designed inductors with the explained design approach of Section 5.1. With proper design, the center frequency of the LNAs are set to 2 GHz.

To perform proper measurements, a source follower is cascaded to the LNA to provide proper matching at the output (Chapter 6, Section 6.1.1). To compensate the process variations, a 3-bit cap bank and a fine-tune varactor are connected at the output of LNA so that the center frequency can be fixed to 2 GHz for each implementation during the measurements.

Effect of V_{B2} on Noise Figure

For lower values of V_{B2}, M1 enters into linear region and the transconductance of M1 decreases, hence the input impedance of the amplifier decreases. Consequently, the input matching condition is lost and the gain of the amplifier decreases. As a result of this, the noise figure also increases for lower V_{B2} values. V_{B2} can be increased as long as there is enough headroom under the given power supply voltage. For a range of V_{B2} values, both devices stay in saturation region. As V_{B2} further increases, M2 enters in linear region. In terms of noise performance, it is not expected to see a significant change in the noise figure when the temperature values of each device are assumed equal to each other. However, when individual temperature values of M1 and M2 are considered separately, it can be expected that larger V_{B2} levels would increase the power dissipation and local temperature of M1. Consequently, when (5.38) is considered,



(a) $L = 30$ nm

(b) $L = 45$ nm

(c) $L = 180$ nm

(d) Error

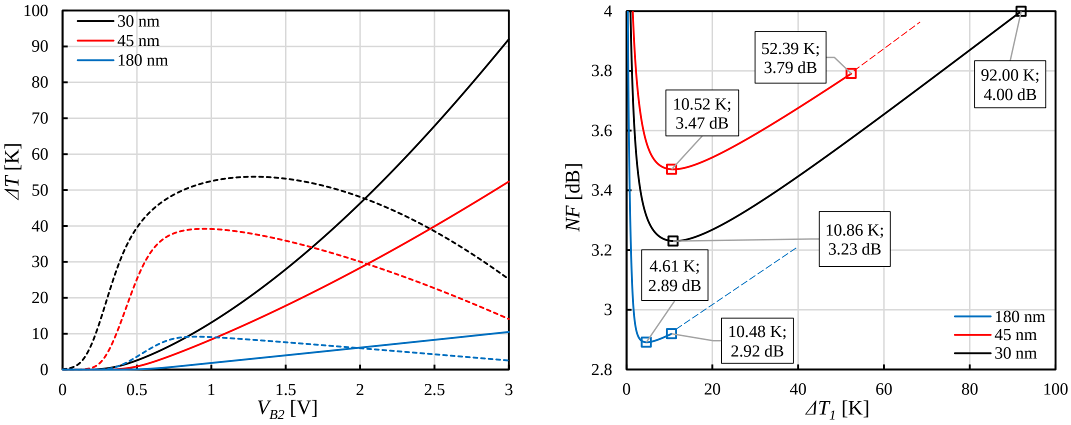
Figure 5.17 – Noise figure under different V_{B2} levels when the supply voltage is set to 3 V for three different the common gate cascode LNA implementations with gate lengths of (a) 30 nm, (a) 45 nm, (a) 180 nm. Continuous lines correspond to the case where self-heating effects are included and dashed lines are for the case where self-heating effects are excluded. (d) the calculation error when self-heating effects are not included.

it is obvious that the thermal noise of M1 would linearly increase with its temperature and this would be observed as a larger noise figure.

From the previous explanations, it can be seen that the noise figure of the common gate cascode LNA should have a lowest value for a specific V_{B2} and it tends to increase when V_{B2} moves away from this point in either direction. For finding the V_{B2} value corresponding to this minimum, the supply voltage, V_{DD}, is set to 3 V and V_{B2} is swept.

Figure 5.17 shows how the noise figure of the three LNAs behave as V_{B2} changes under V_{DD} = 3 V. The experiments are performed for two cases where self-heating effects are included and excluded. It can be seen that for each LNA, the noise figure curves are different for the two different cases. Moreover, noise figure is always larger when self-heating effects are active. With this, it can be concluded that excluding self-heating effects causes some amount of calculation error, which can be defined as

$$\text{Error} = \frac{NF_{SH} - NF_{noSH}}{NF_{noSH}}. \tag{5.46}$$



(a)

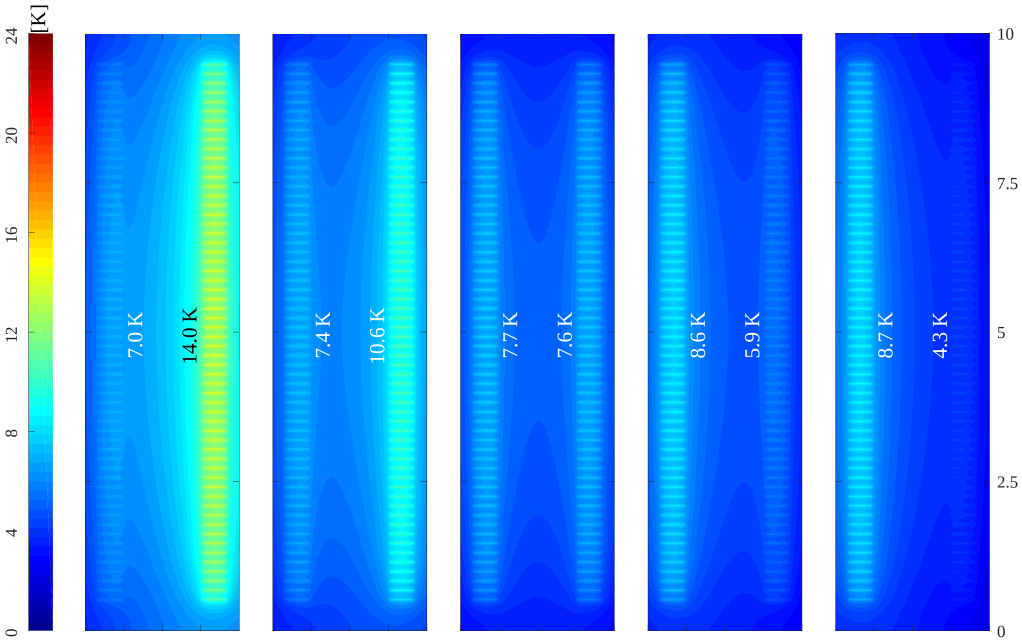
(b)

Figure 5.18 – (a) temperature rise of M1 (continuous) and M2 (dashed) as V_{B2} changes, (b) noise figure vs temperature rise of M1.

The error for each case is plotted on Figure 5.17d. It can be seen that as the gate length decreases, the error increases significantly and it reaches up to 23% for L = 30 nm.

When the noise figure performance of the LNA implementation with 30 nm gate length is examined, it can be seen that the noise figure is relatively constant for the isothermal experiment. However, the noise figure rapidly rises with V_{B2} when self-heating effects are considered and the temperature of each device is calculated separately. Figure 5.18a, which shows the individual temperature rise of each device, can help to understand this situation. It can be seen that the temperature of M1 rises with an increasing rate as V_{B2} increases and it goes up to 92 K. Consequently, M1 generates higher amount of noise. Since the noise figure of the common gate cascode LNA has a strong dependency on the noise generated by M1 according to (5.38), the noise figure of the LNA also increases. At his point, the actual noise figure becomes 0.9 dB greater than the isothermal case and this corresponds to a 23% calculation error. On the other hand, the noise figure has a minimum value of 3.23 dB at V_{B2} = 0.91 V and it stays quite constant in the vicinity of this point. Therefore, this point should be used for biasing the amplifier to obtain the lowest noise figure.

For the LNA implementation with 45 nm gate length, similar behaviours are observed. Nevertheless, the difference between the isothermal case and the case where self-heating effects are considered do not differ as much as they do for the LNA implementation with 30 nm gate length. In 45 nm gate length case, the bias point that gives the lowest noise figure is V_{B2} = 1.12 V. This value is roughly 200mV larger than the 30 nm gate length case. This can be explained by the larger threshold voltage values of the M1 and M2 devices (Table 5.4). The lowest possible noise figure for L = 45 nm case is 3.47 dB and it is roughly 0.25 dB larger than 30 nm gate length case. Normally, one would expect a decreasing noise figure trend as the gate length increases. This is because longer channels implementations have smaller γ and larger α, which is the case in these implementations. This means that the larger noise figure



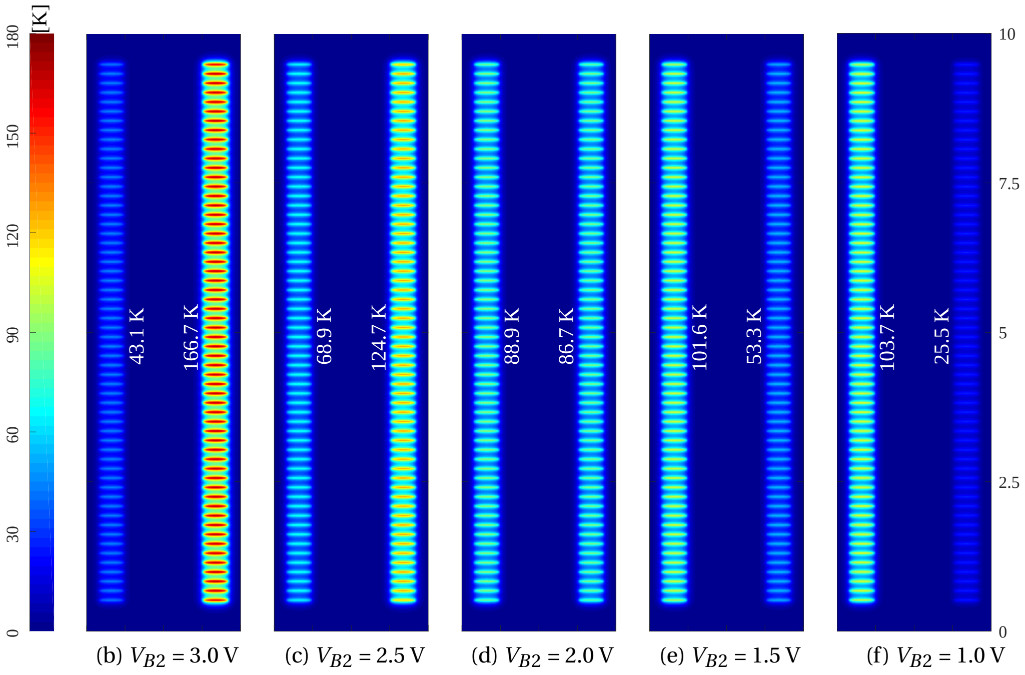
(b) $V_{B2} = 3.0$ V (c) $V_{B2} = 2.5$ V (d) $V_{B2} = 2.0$ V (e) $V_{B2} = 1.5$ V (f) $V_{B2} = 1.0$ V

Figure 5.19 – ΔT profile of the common gate cascode LNA with $L = 30$ nm for the bulk thermal properties under different bias conditions provided by V_{B2} while $V_{DD} = 3$ V. The observable devices are M2 and M1 from left to right. The spatial unit is µm.

does not originate from α and γ parameters. The main reason of larger noise figure in 45 nm gate length case compared to the 30 nm case is the lower transconductance of M1. As it was previously explained, the transconductance of M1 for the 30 nm case is set to a relatively large value (Table 5.4) as dictated by (5.10) in order to compensate for the influence of the load impedance on the input impedance. Consequently, the larger transconductance of M1 decreases the noise figure slightly.

For the long channel case, where the LNA is implemented by using devices with 180 nm gate length, the difference between the isothermal and self-heating case is not as dramatic as the short channel cases. The maximum temperature under largest V_{B2} is around 11 K, which is significantly smaller than the short channel implementations. This can be explained by the large area of the devices and the resulting low heat generation density. In this case, the lowest noise figure is obtained under a relatively large V_{B2}, which is 1.65 V. This is because of the large threshold voltages of long channel devices (Table 5.4). Nevertheless, the minimum noise figure is found to be the smallest for the long channel example, even smaller than the 30 nm gate length case whose g_{m1} is 20% larger. This is primarily because γ is significantly smaller in the long channel implementation.

Figure 5.18b shows the noise figure dependency on the temperature rise due to the self-heating of M1. It can be seen that the noise figure linearly increases with the temperature. The linear dependency suggests that the increase in the noise figure is primarily due to the increase of the



(b) $V_{B2} = 3.0$ V (c) $V_{B2} = 2.5$ V (d) $V_{B2} = 2.0$ V (e) $V_{B2} = 1.5$ V (f) $V_{B2} = 1.0$ V

Figure 5.20 – ΔT profile of the common gate cascode LNA with $L = 30$ nm for the FDSOI thermal properties under different bias conditions provided by V_{B2} while $V_{DD} = 3$ V. The observable devices are M2 and M1 from left to right. The spatial unit is µm.

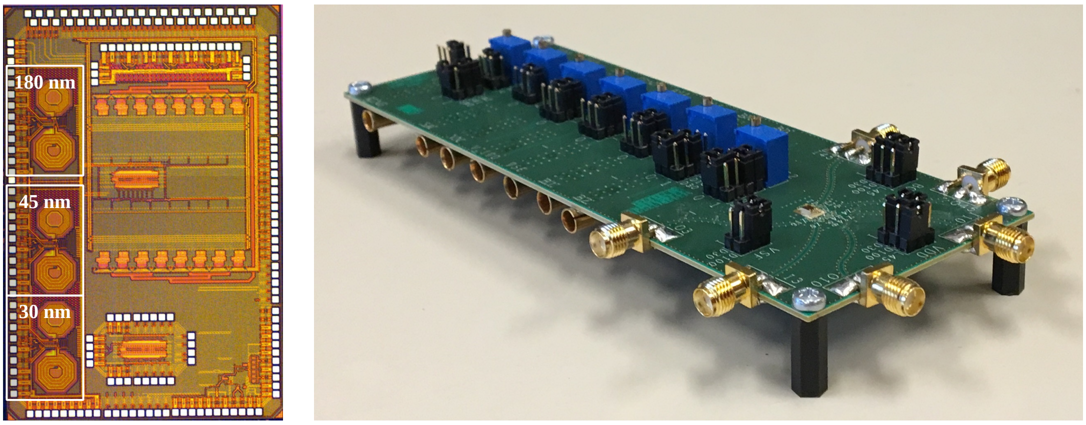
thermal noise of M1. The individual temperature values of M1 and M2 are also examined by more detailed thermal simulations rather than assuming a single device temperature. Figure 5.19 and Figure 5.20 show the thermal maps for the bulk and FDSOI geometry respectively of 30 nm implementation. It can be seen that the temperature profile is much flatter in bulk than FDSOI, which means that the same implementations in bulk does not show such dramatic dependencies on V_{B2}. On the other hand, the temperature values of more detailed thermal simulations are observed to be larger than the ones shown on Figure 5.18a. This suggest that the noise figure might rise even more dramatically with V_{B2} than what is observed on the simulations and the reliability of self-heating model provided by the design kit provider might be questionable.

5.3 Measurements

To observe the effect of different self-heating levels resulting from different bias conditions, the noise figure of LNAs were measured with the N8975A noise figure analyser and N4002A noise source of Keysight Technologies (Figure 5.21). Different self-heating levels were provided by

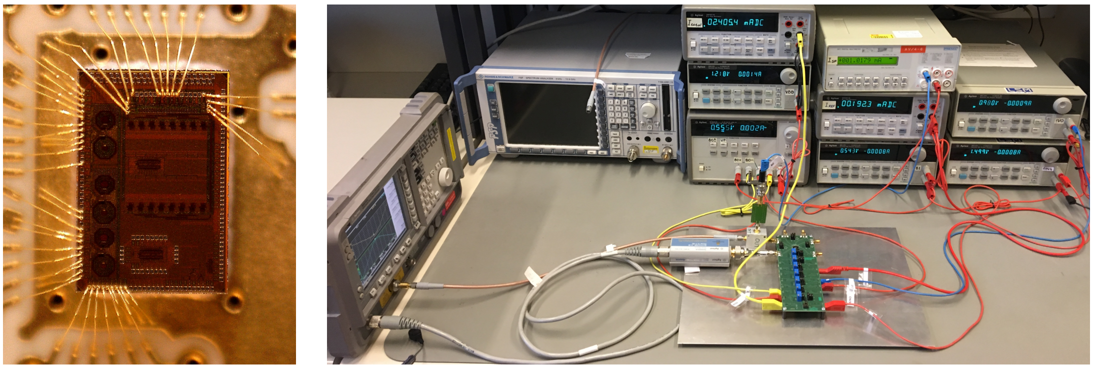
(a) modulating the power dissipation ratios of M1 and M2 through V_{B2} and

(b) changing the current consumption of LNA through I_{REF}.



(a) Micrograph

(b) RF PCB



(c) Bonded die

(d) Measurement set-up

Figure 5.21 – (a) micrograph of the taped-out die where the three LNA implementations are indicated with the channel length values, (b) RF test board, (c) bonded die and (d) the measurement set-up.

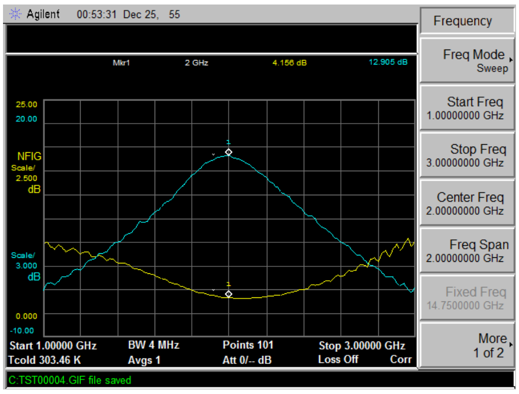
In case (a), the noise figure is expected to decrease until M1 is in saturation, mainly because the gain of the LNA increases rapidly with increasing V_{B2}. Once M1 is in saturation, the gain stays relatively constant. Further increasing V_{B2} increases the power density of M1 roughly by

$$P_{M1} = \frac{I_{D,M1} V_{DS,M1}}{WL}. \tag{5.47}$$

For long channel devices, the power density and the temperature of M1 have weaker dependency on V_{B2} than the short channel devices due to the device area factor in the denominator of (5.47). Therefore, for long channel implementations, the noise figure is expected to stay relatively constant with V_{B2}, whereas it is expected to increase rapidly in short channel devices due to the rapid increase of the local temperature of the FDSOI device.

In case (b), increasing I_{REF} increases the total power dissipation of the LNA according to

$$P_{LNA} = N I_{REF} V_{DD} \tag{5.48}$$



Freq Mode Sweep
Start Freq 1.00000000 GHz
Stop Freq 3.00000000 GHz
Center Freq 2.00000000 GHz
Freq Span 2.00000000 GHz
Fixed Freq 14.7500000 GHz

Start 1.00000 GHz
Tcold 303.46 K
BW 4 MHz
Avgs 1
Points 101
Att 0/-- dB
Stop 3.00000 GHz
Loss Off
Corr

C:TST00004.GIF file saved

Figure 5.22 – Screen-shot of the N8975A screen, where the noise figure and gain of the 45 nm LNA can be seen for the 1 GHz - 3 GHz band.

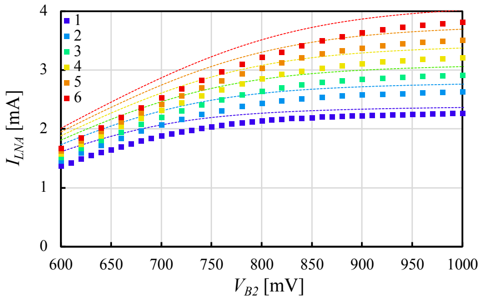
Table 5.5 – The measured I_{REF} and the applied V_{B1} values for each experiment. The units of I_{REF} and V_{B1} are µA and mV respectively.

LNA →	180 nm		45 nm		30 nm	
Experiment ↓	I_{REF}	V_{B1}	I_{REF}	V_{B1}	I_{REF}	V_{B1}
1	191	580	190	544	194	497
2	225	601	222	565	224	518
3	251	616	247	580	241	523
4	279	632	272	596	271	538
5	308	647	298	611	302	554
6	338	663	326	627	334	570

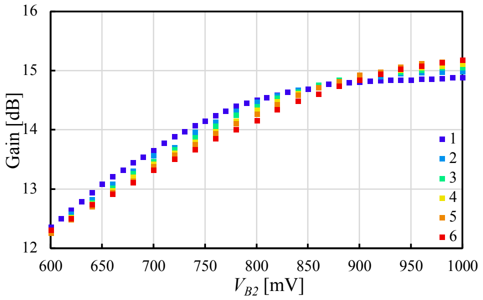
where N is the current multiplication factor of the current mirror constructed with M0 and M1 devices on Figure 5.10. Consequently, for larger I_{REF} values in short channel implementations, it is expected to observe a more rapid increase in the noise figure by increasing V_{B2}. Whereas, one expects to see similar trends under different I_{REF} values for long channel devices since the temperature of M1 would have a much weaker dependency on the power dissipation.

5.3.1 Results

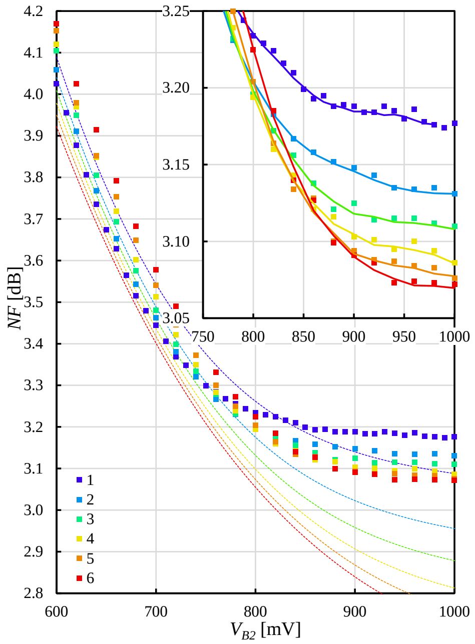
During the measurements, first, the behaviour of each LNA was observed for a large frequency band, mainly to see the location of the center frequency (Figure 5.22). For each case, the center frequency was fine-tuned to 2 GHz by externally changing the capacitor bank and varactor control signals. After observing the correct behaviour, the gain and noise figure data points were taken at 2 GHz. The measurement points can be observed on Figure 5.23 to Figure 5.25 for all LNAs. Three different parameters are provided, which are current consumption, gain and noise figure. Dashed lines show simulation results where the same bias values were applied. On the upper right corner of the noise figure plot, a zoomed graph can be observed



(a) Current Consumption

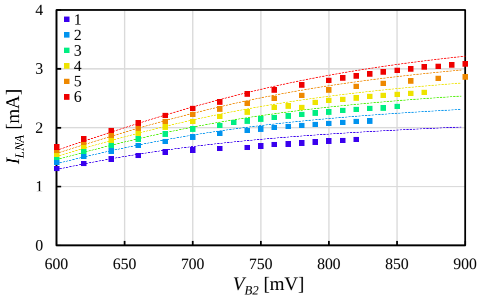


(b) Gain

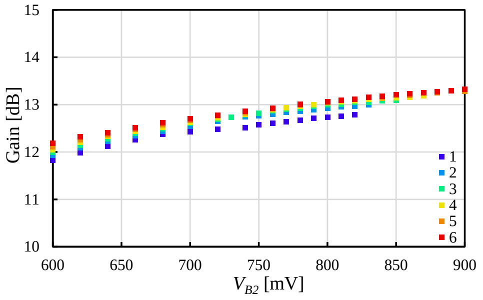


(c) Noise Figure

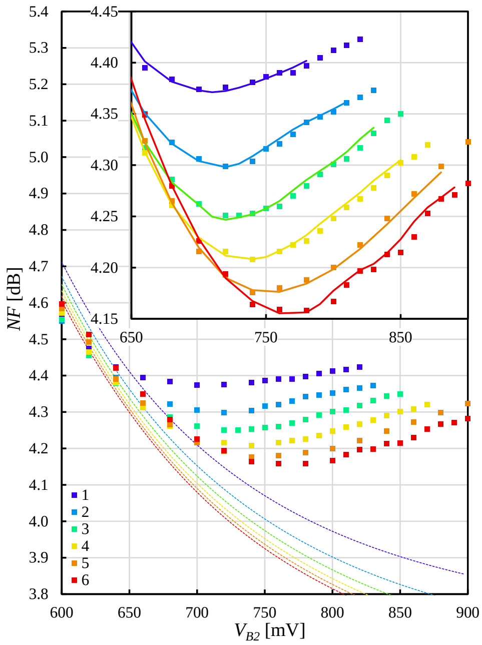
Figure 5.23 – Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 180$ nm.



(a) Current Consumption

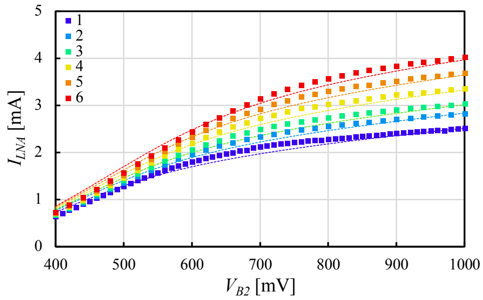


(b) Gain

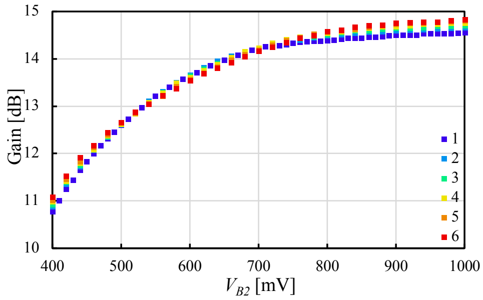


(c) Noise Figure

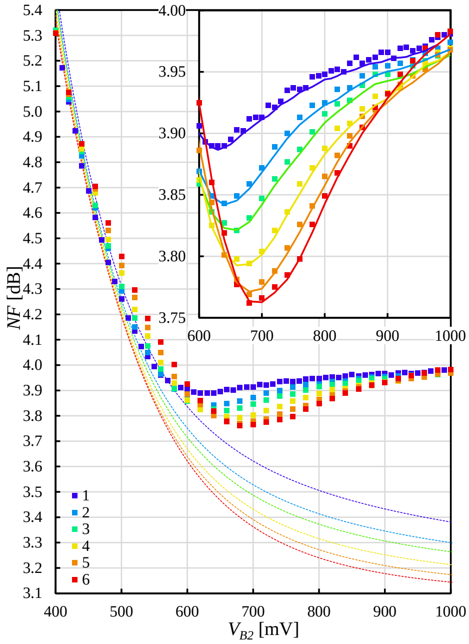
Figure 5.24 – Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 45$ nm.



(a) Current Consumption



(b) Gain



(c) Noise Figure

Figure 5.25 – Measurement results of (a) current consumption, (b) gain and (c) noise figure at 2 GHz of the LNA implemented with $L = 30$ nm.

where the noise figure is plotted in more detail for the linear-saturation transition regime. Here, the continuous lines correspond a trend-line that is constructed by taking a moving average of three points. While taking these data points, V_{B2} is swept from a value where M1 is in linear region up to the maximum allowed voltage level, which is 1 V, of the used 28 nm FDSOI technology. This was done for six different I_{REF} values by externally changing the bias voltage of M1 (V_{B1}). The applied V_{B1} values and the resulting I_{REF} values are summarized on Table 5.5 for each case.

Figure 5.23c shows that the noise figure of the long channel LNA (L = 180 nm) decreases continuously. The decrease in the noise figure is quite rapid until $V_{B2} \approx 800$ mV due to linear region operation of M1. For larger V_{B2} values, noise figure is relatively flat. Nevertheless, it does not start to increase by further increasing V_{B2}. This situation shows that the temperature rise due to the self-heating effects are not large enough to affect the noise figure. Even if the power dissipation is increased by increasing I_{REF}, no increase in the noise figure is observed by increasing V_{B2}, thanks to the large device area and low heat density.

The situation is quite different for the short channel implementations (Figure 5.24c and Figure 5.25c). In both 30 nm and 45 nm implementations, the noise figure decreases rapidly until M1 enters into saturation region. Immediately after the linear-saturation transition, the noise figure starts to increase due to increasing heat generation and temperature of M1. Moreover, the slope of the noise figure becomes larger with increasing I_{REF} since the heat generation and temperature increase more rapidly with V_{B2} according to (5.48). Another observation can be made regarding the fact that the minimum noise figure decreases with increasing I_{REF}. This is mainly because the transconductance of M1 increases under larger current bias. The noise figure can be decreased with this approach. The trade-off is the increased headroom, which can be also understood by observing that the V_{B2} value corresponding to the minimum noise gets larger with increasing I_{REF}. In addition, one also needs to be careful by considering the matching of the input impedance.

5.3.2 Comments

The measurement results clearly show that the noise figure is significantly influenced by the self-heating effects in FDSOI for small dimension implementations. Excluding the self-heating effects during simulations results in a very optimistic noise figure estimation and the measured noise figure can be up to 1 dB larger than the simulated value due to self-heating effects. Additionally, the obtained bias points are not the optimum to obtain the lowest noise figure as the noise figure does not continuously decrease with V_{B2}. To have the best noise figure performance, it is crucial to have realistic thermal models and include them separately for each device during the simulations.

5.4 Conclusion

Three different common gate cascode LNAs were implemented in 28 nm FDSOI technology with different gate length values, which are 30 nm, 45 nm and 180 nm in order to observe the self-heating effects on noise figure. Custom designed inductors were used for their source bias and drain load. Noise figure measurements were performed at 2 GHz. It was observed that increasing the bias voltage of the cascode device increases the noise figure of the LNAs with short channel devices due to the self-heating effects. Moreover, it was demonstrated that the noise figure increases more rapidly when the power consumption of the short channel LNAs are larger, which is due to larger increase of the temperature. The same situation was not observed in the case of LNAs with long channel devices since their power density values are not significantly large to create a dangerous hot-spot.

6 Flicker Noise Measurements

In Chapter 4, self-heating effects on the thermal and flicker noise performance of FDSOI devices are analysed. In this chapter, we present two experiments for physically observing how the flicker noise of an individual FDSOI device increases due to its self-heating. The main target in both experiments is to extract the flicker noise of a single FDSOI MOSFET under different bias conditions, hence different self-heating levels.

6.1 Test Blocks

For observing the influence of the self-heating effects on flicker noise, three parameters were applied as test variables. While choosing these parameters, attention was paid in order not to change the other parameters significantly while sweeping the self-heating. The three parameters are listed below:

- **Channel Length:** Under constant power dissipation, by decreasing the channel length, so the area, the power density increases roughly quadratically.
- **Drain Voltage:** By increasing the drain voltage under constant gate bias, the power density increases between quadratically to linearly depending on the value of output impedance in saturation.
- **Gate Voltage:** By increasing the gate voltage under constant drain bias, the power density increases between quadratically to linearly depending on the amount of velocity saturation.

For observing the noise levels under various values of these parameters, two different test structures were designed in 28 nm FDSOI technology for two experiments. The experiments are named as experiment-A (Figure 6.1a) and experiment-B (Figure 6.1b). In both cases, the goal is to measure the flicker noise of M1. To have an access to the flicker noise of M1, the main part of the test circuit is designed as a common source amplifier stage in experiment-A

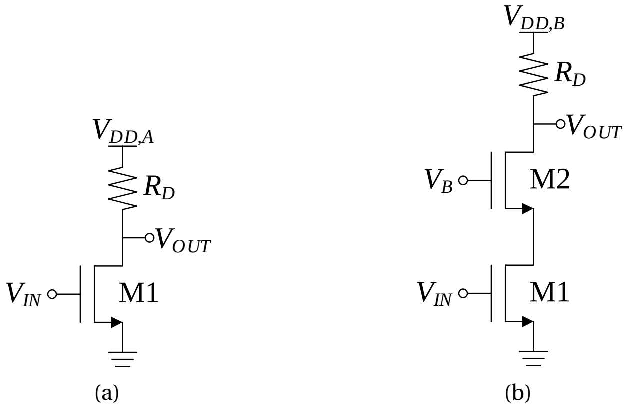


Figure 6.1 – Main parts of the two test structures for observing the flicker noise of a single MOSFET (M1). (a) experiment-A: common source and (b) experiment-B: common source with a cascode device.

and a common source amplifier stage with a cascode device in experiment-B. An additional load resistance, R_D, is included for (a) providing the desired bias on M1, (b) having drain of the device as an AC node and (c) having a reasonable gain so that the flicker noise of M1 is dominant in the system. For providing test conditions with different amounts of self-heating, in experiment-A the drain voltage of M1 is changed by changing the supply voltage V_{DD}, while the gate voltage is kept fixed at a constant value. However, especially for short channel devices, the output impedance

$$R_{outA} = r_o \parallel R_D \tag{6.1}$$

varies significantly due to the strong dependence of r_o on the drain bias. Consequently, the gain

$$A_{vA} = -g_m(r_o \parallel R_D) \tag{6.2}$$

also varies. Although this variation is removed from equation by referring the measured noise to the input, it is still an additional variation in the system. For this reason, experiment-B is also proposed since the cascode stage increases the output impedance. By using (5.27) and (5.14), the output impedance of the cascode stage can be written as

$$R_{outB} = \left\{ [1 + r_{o2}\left(g_{m2} + g_{mb2}\right)] \, r_{o1} + r_{o2} \right\} \Big\| R_D = \left(A_{vg2} r_{o1} + r_{o2}\right) \| R_D \tag{6.3}$$

where the output impedance seen while looking downwards into the drain of M2 is amplified by the intrinsic gain of M2, compared to the case of experiment-A. If the intrinsic gain of the cascode device M2 is sufficiently large, the output impedance becomes independent of the device parameters and converges to R_D. The gain of the amplifier in experiment-B is

$$A_{vB} = -\frac{g_{m1} r_{o1} R_D \, [1+(g_{m2} + g_{mb2}) r_{o2}]}{[1+(g_{m2} + g_{mb2}) r_{o2}] \, r_{o1} + r_{o2} + R_D} = -\frac{g_{m1} r_{o1} R_D A_{vg2}}{A_{vg2} r_{o1} + r_{o2} + R_D}. \tag{6.4}$$

With the same assumption, the gain approaches to $g_{m1}R_D$, which is roughly independent of the drain bias of M1. Another advantage of experiment-B is that the drain voltage and the self-heating of M1 can be controlled with the gate bias of M2 (V_B) rather than the power supply voltage of the amplifier. Since the gate voltage of M1 is fixed at a constant value, the drain current of M1 and M2 is roughly constant. Having a constant drain current, the source voltage of M2 follows its gate voltage (V_B). Like this, the power dissipation of M1 can be almost linearly controlled with V_B. Finally, the transconductance of the amplifier in experiment-A and experiment-B are

$$G_{mA} = g_m \tag{6.5}$$

and

$$G_{mB} = g_{m1}\frac{A_{vg2}r_{o1}}{A_{vg2}r_{o1} + r_{o2}}. \tag{6.6}$$

which will be used in the following sections during the derivation of the input referred noise voltage expressions.

6.1.1 Practical Considerations

In order to have a sufficiently wide band and bias flexibility during the electrical measurements, it is necessary to add additional blocks to the designs, which are shown on Figure 6.1. With the additional blocks, the complete test system can be observed on Figure 6.2. The functions of the additional blocks are explained in the following.

Matching

During the measurements, the signal input (v_{in} on Figure 6.2) for the test blocks is provided by different measurement devices such as network analyser and signal generator. The output is sensed by an oscilloscope or a spectrum analyser (v_{out} sensed by R_L on Figure 6.2). All of these devices are terminated by a 50 Ω resistor (R_S and R_L). For medium to high frequency range measurements, it is necessary to have the input and the output impedances of each stage matched to the characteristic impedance of the devices and the coaxial cables of the measurement set-up in order to prevent reflections and undesired attenuation. For this, a 50 Ω resistor is placed at the gate of the devices (R_{in} on Figure 6.2). The drawback of the additional matching resistor at the input is some amount of decrease in the gain due to the voltage division at the input according to

$$A_{vI} = \frac{R_{in}}{R_{in} + R_S} \tag{6.7}$$

which gives 6.02 dB attenuation under matching. For the output matching, placing a 50 Ω resistor is not possible. The loading of a 50 Ω resistor decreases the gain of the stage

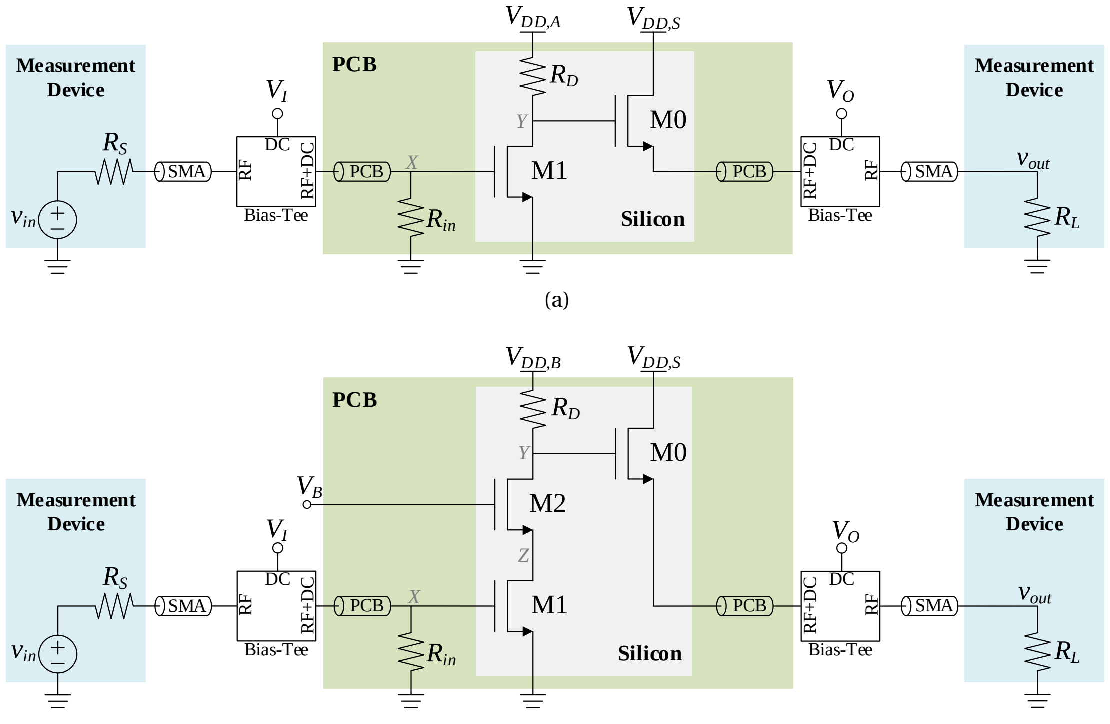


Figure 6.2 – Simple schematic representation of the complete test system for the designs of Figure 6.1.

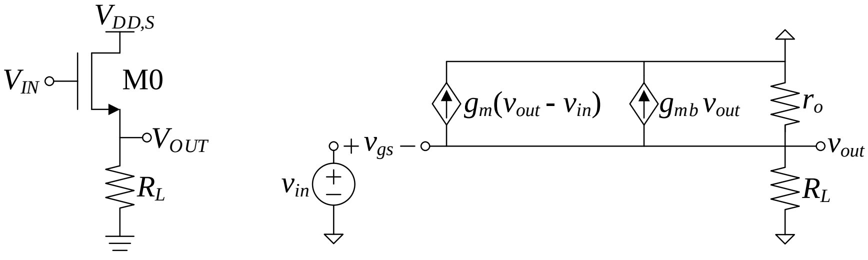
significantly as the output load of the common source stages is much larger than this value. For this reason, the output of the common source stage is followed by a source follower since the source follower stage does not bring any resistive loading (only capacitive) and the output impedance of the source follower can easily be tuned to 50 Ω with the used technology. The source follower schematic can be seen on Figure 6.3a. According to the model on Figure 6.3b, the output impedance of the unloaded source follower is

$$R_{outS} = \frac{1}{g_m + g_{mb} + \dfrac{1}{r_o}} = \frac{r_o}{1 + r_o\left(g_m + g_{mb}\right)} = \frac{r_o}{A_{vg}}. \tag{6.8}$$

For a perfect matching, this value has to be set equal to R_L. When the output is loaded with R_L, the gain of the stage is

$$A_{vS} = \frac{g_m}{g_m + g_{mb} + \dfrac{1}{r_o} + \dfrac{1}{R_L}} = \frac{g_m r_o R_L}{r_o + \left[1 + r_o\left(g_m + g_{mb}\right)\right] R_L} = g_m\left(R_{out} \parallel R_L\right) \tag{6.9}$$

The output impedance of the source follower is set equal to the characteristic impedance of the test equipment ($R_{out} = R_L = 50\ \Omega$ on Figure 6.3). In this case, the voltage gain of the stage



(a) Source follower

(b) Small-signal AC model

Figure 6.3 – (a) simple schematic of source follower and (b) its small-signal AC model.

becomes

$$A_{vS} = \frac{g_m}{2\left(g_m + g_{mb} + \dfrac{1}{r_o}\right)} = \frac{g_m R_L}{2} \tag{6.10}$$

independent of the channel length. For a long channel device, gain can be approximated as

$$A_{vS} \approx \frac{g_m}{2\left(g_m + g_{mb}\right)} \tag{6.11}$$

which gives a gain of 0.5 when g_{mb} is zero. For a positive g_{mb}, this value is even less. In this implementation, the voltage gain of the source follower stage is around 0.45, which corresponds to an attenuation of approximately 7 dB.

Biasing

As it has already been mentioned, the heat density of M1 is adjusted by changing its gate and drain bias voltages. For this, it is necessary to have DC access to the gate and the drain nodes of M1. It can be seen on Figure 6.2 that the gate of M1 is physically accessible through the test board (node X). However, since it carries a high frequency signal at the same time, a direct electrical access to this node is not possible and the DC level has to be provided separately. This is provided by a bias-tee. The bias-tee adds the AC and the DC voltages through large capacitive and inductive paths respectively while providing 50 Ω termination. Like this, the external DC voltage V_I appears directly at the gate of M1, without influencing the AC signal provided by v_{in}. The drain voltage of M1 in experiment-A can only be controlled through $V_{DD,A}$. However, since the drain of M1 is also the gate of M0 (node Y), changing the voltage at this node alters the biasing of the source follower. Therefore, it is also desired to control the bias of M0 externally. For this reason, another bias-tee is used at the output in order to provide the desired DC voltage (V_O) to the source of M0. Like this, the AC parameters of M0 (especially R_{outS} which has to be equal to R_L) can be adjusted without disturbing the AC output of the system. In experiment-B, the DC level at the input of the source follower is much more stable than the one of experiment-A. This is provided by the cascode device, M2. In case the drain voltage of M1 is changed through the gate bias of the cascode device (V_B), the DC voltage at Y does not change significantly thanks to the large output impedance seen at the drain of M2. Like this, the AC parameters of M0 can be set only with a little tuning.

6.1.2 Input Referred Noise

Experiment-A

The small-signal AC model of the complete system of experiment-A is shown on Figure 6.4a. According to the model, the gain of the complete system is

$$A_v = -\frac{R_{in}}{R_{in}+R_S}\cdot g_{m1}\left(R_D \parallel r_{o1}\right)\cdot \frac{g_{m0}}{g_{m0}+g_{mb0}+\dfrac{1}{r_{o0}}+\dfrac{1}{R_L}} = A_{vI}A_{vA}A_{vS} \tag{6.12}$$

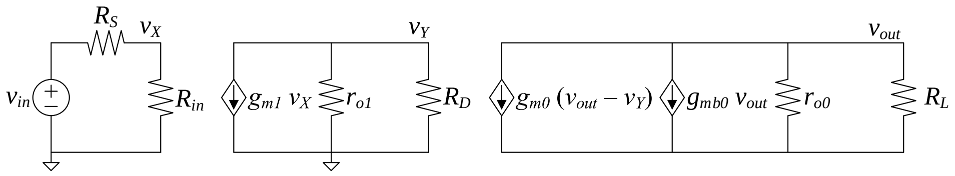
which reduces to

$$A_v = -\frac{1}{4}\cdot g_{m1}\left(R_D \parallel r_{o1}\right)\cdot g_{m0}R_L \tag{6.13}$$

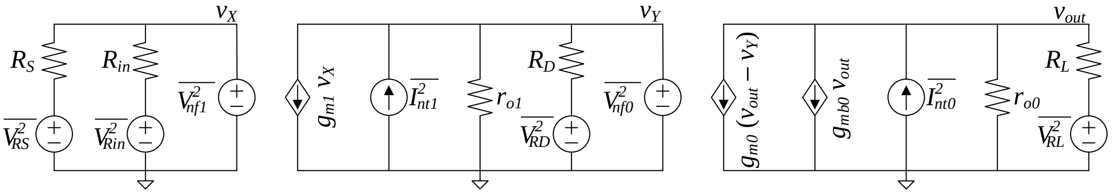
when both input and output are matched to the measurement system. It can be seen that the gain drops roughly by 75% due to the additional circuity for matching. For calculating the contribution of different noise mechanisms, one can analyse the circuit on Figure 6.4b where each independent noise source is represented either with a voltage or a current source. Without making any assumptions, the input referred noise voltage is calculated as

$$\begin{aligned}
\overline{V^2_{n,i}} &= \underbrace{\overline{V^2_{RS}} + \overline{V^2_{Rin}} + \frac{1}{A^2_{vI}}\left(\frac{1}{g^2_{m1}R^2_D}\overline{V^2_{RD}} + \frac{1}{A^2_{vA}g^2_{m0}R^2_L}\overline{V^2_{RL}}\right)}_{\text{Resistor Thermal Noise}} \\
&+ \underbrace{\frac{1}{A^2_{vI}}\left(\frac{1}{g^2_{m1}}\overline{I^2_{nt1}} + \frac{1}{A^2_{vA}g^2_{m0}}\overline{I^2_{nt0}}\right)}_{\text{MOS Thermal Noise}} + \underbrace{\frac{1}{A^2_{vI}}\left(\overline{V^2_{nf1}} + \frac{1}{A^2_{vA}}\overline{V^2_{nf0}}\right)}_{\text{MOS Flicker Noise}} \\
&= \underbrace{4kT\left[R_S + R_{in} + \frac{1}{A^2_{vI}}\left(\frac{1}{g^2_{m1}R_D} + \frac{1}{A^2_{vA}g^2_{m0}R_L} + \frac{1}{g_{m1}}\frac{\gamma_1}{\alpha_1} + \frac{1}{A^2_{vA}g_{m0}}\frac{\gamma_0}{\alpha_0}\right)\right]}_{\text{Thermal Noise}} \\
&+ \underbrace{\underbrace{\frac{1}{A^2_{vI}}\frac{K(V_{GS1})}{C'^2_{ox}f^{\gamma}W_1L_1}}_{\text{Targeted Noise}} + \frac{1}{A^2_{vI}A^2_{vA}}\frac{K(V_{GS0})}{C'^2_{ox}f^{\gamma}W_0L_0}}_{\text{Flicker Noise}}.
\end{aligned} \tag{6.14}$$

As it was already mentioned, the goal is to observe the behaviour of the flicker noise of M1, which is indicated as the **Targeted Noise** in (6.14). However, during the measurements, only the total noise of the complete system at the output can be accessed and it is not possible to

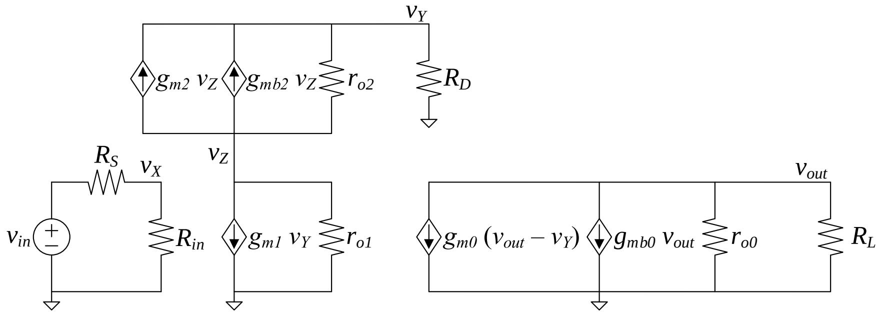


(a) Small-signal AC model

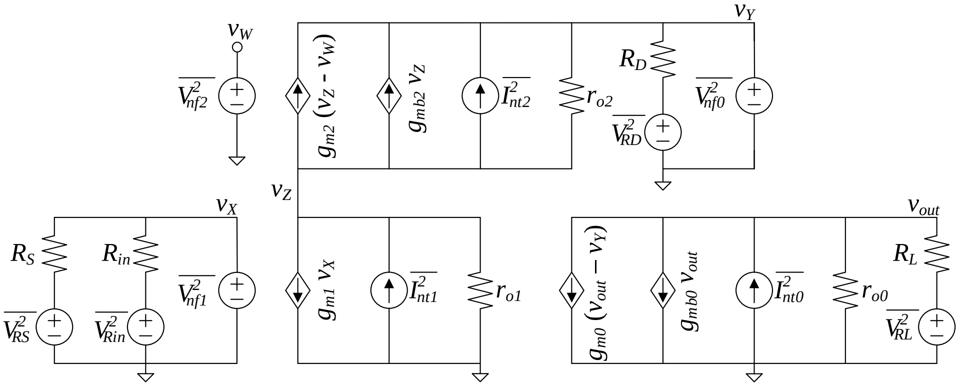


(b) Small-signal AC model including all independent noise sources

Figure 6.4 – Small signal AC models of the complete system of experiment-A.



(a) Small-signal AC model



(b) Small-signal AC model including all independent noise sources

Figure 6.5 – Small signal AC models of the complete system of experiment-B.

remove the contribution of any independent noise source. For this reason, one needs to make sure that the power of the flicker noise of M1 is sufficiently larger than the total noise power generated by other sources. We will discuss on this point in the following sections in more detail.

Experiment-B

Similarly, the small-signal AC model of the complete system of experiment-B is shown on Figure 6.5a. The gain of the complete system is

$$A_v = -\frac{R_{in}}{R_{in}+R_S}\,\frac{g_{m1}r_{o1}R_DA_{vg2}}{A_{vg2}r_{o1}+r_{o2}+R_D}\,\frac{g_{m0}}{g_{m0}+g_{mb0}+\frac{1}{r_{o0}}+\frac{1}{R_L}} = A_{vI}A_{vA}A_{vS} \tag{6.15}$$

which reduces to

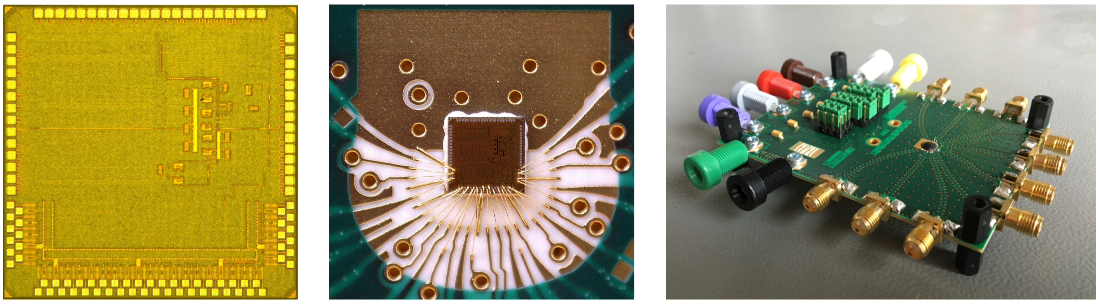
$$A_v = -\frac{1}{4}\,\frac{g_{m1}r_{o1}R_DA_{vg2}}{A_{vg2}r_{o1}+r_{o2}+R_D}\,g_{m0}R_L \tag{6.16}$$

when both input and output are matched to the measurement system. The same amount of reduction in the gain is also present in experiment-B. The input referred noise voltage is also similar to the one in experiment-A except the little contribution from M2 and the slight increase of the gain of the main stage:

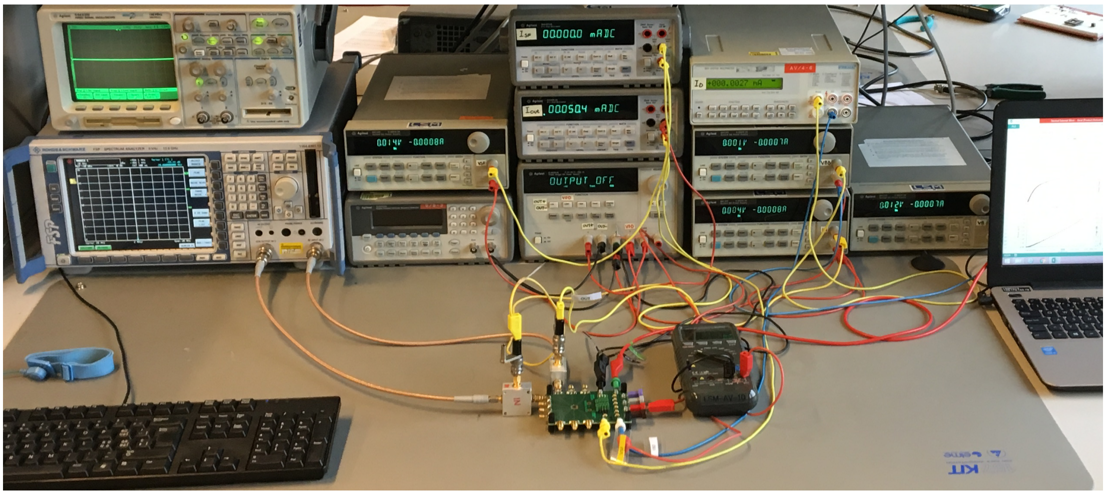
$$\begin{aligned}
\overline{V^2_{n,i}} &= \underbrace{\overline{V^2_{RS}}+\overline{V^2_{Rin}}+\frac{1}{A^2_{vI}}\left(\frac{1}{G^2_{mB}R^2_D}\overline{V^2_{RD}}+\frac{1}{A^2_{vA}g^2_{m0}R^2_L}\overline{V^2_{RL}}\right)}_{\text{Resistor Thermal Noise}}\\
&+\underbrace{\frac{1}{A^2_{vI}}\left(\frac{1}{g^2_{m1}}\overline{I^2_{nt1}}+\frac{r^2_{o2}}{A^2_{vi1}A^2_{vg2}}\overline{I^2_{nt2}}+\frac{1}{A^2_{vA}g^2_{m0}}\overline{I^2_{nt0}}\right)}_{\text{MOS Thermal Noise}}\\
&+\underbrace{\frac{1}{A^2_{vI}}\left(\overline{V^2_{nf1}}+\frac{A^2_{vi2}}{A^2_{vi1}A^2_{vg2}}\overline{V^2_{nf2}}+\frac{1}{A^2_{vA}}\overline{V^2_{nf0}}\right)}_{\text{MOS Flicker Noise}}\\
&=\underbrace{4kT\left[R_S+R_{in}+\frac{1}{A^2_{vI}}\left(\frac{1}{G^2_{mB}R_D}+\frac{1}{A^2_{vA}g^2_{m0}R_L}+\frac{1}{g_{m1}}\frac{\gamma_1}{\alpha_1}+\frac{A^2_{vi2}}{A^2_{vi1}A^2_{vg2}g_{m2}}\frac{\gamma_2}{\alpha_2}+\frac{1}{A^2_{vA}g_{m0}}\frac{\gamma_0}{\alpha_0}\right)\right]}_{\text{Thermal Noise}}\\
&+\underbrace{\underbrace{\frac{1}{A^2_{vI}}\frac{K(V_{GS1})}{C'^2_{ox}f^{\gamma}W_1L_1}}_{\text{Targeted Noise}}+\frac{1}{A^2_{vI}}\frac{1}{C'^2_{ox}f^{\gamma}}\left(\frac{A^2_{vi2}}{A^2_{vi1}A^2_{vg2}}\frac{K(V_{GS2})}{W_2L_2}+\frac{1}{A^2_{vA}}\frac{K(V_{GS0})}{W_0L_0}\right)}_{\text{Flicker Noise}}
\end{aligned} \tag{6.17}$$

6.2 Measurement Results

In both experiments, different samples with different gate length values were designed and taped out (Figure 6.6) in order to observe the effect of gate length and heat density. In experiment-A two samples were designed for L = 60 nm and L = 100 nm. On the other



(a) Micrograph (b) Bonded die (c) RF PCB



(d) Measurement set-up

Figure 6.6 – (a) Micrograph of the taped-out die, (b) bonded die: gold wire-bonding is used for better electrical conductivity, (c) RF test board and (d) the measurement set-up. Many power supplies and multi-meters were necessary to provide the bias and supply voltages (V_I, V_O, V_B, $V_{DD,A}$, $V_{DD,B}$, $V_{DD,S}$) and measure different currents for each data point.

Table 6.1 – Properties and bias voltages of different samples used in experiment-A and experiment-B.

	Experiment-A		**Experiment-B**			Unit
Length	60	100	30	60	100	nm
Width	24×625	28×780	24×470	24×625	28×780	nm
Low Heating V_I	400	400	325	400	420	mV
High Heating V_I	550	550	500	550	570	mV

hand, an additional sample of $L = 30$ nm was included in experiment-B. In experiment-A the 30 nm sample was not used mainly due to its very low output impedance, which results in a very low gain that is strongly dependent on the power supply voltage. The width of the samples were sized in order to get a transconductance of 20 mA/V at a gate drive of 0.6 V. The values are summarized on Table 6.1.

During the measurements, two different gate bias levels were applied on the samples to provide a low heating and a high heating in order to compare their behaviours. For low heating, a drain current of 0.4 mA was targeted and the input bias (V_I on Figure 6.2) was set accordingly. On the other hand, for the high heating case, a drain current of 2 mA was targeted so that the heating is roughly more than five times larger. These bias levels are listed on Table 6.1. After finding the desired gate bias, V_I was fixed at this constant level and the second parameter was swept in order to modify the drain voltage, hence the power dissipation and heat density of M1. In experiment-A, the power supply voltage ($V_{DD,A}$) was swept. The sweep was started from a sufficiently low voltage where M1 is in linear region and it was increased until the point where the drain voltage of M1 passes the maximum allowed voltage of the technology (1 V). In experiment-B, the gate voltage of the cascode device, V_B, was swept while the power supply voltage, $V_{DD,B}$, was kept constant. The lower border of the sweep was again set to the linear region operation of M1, while the upper border was set to the constant power supply ($V_{DD,B}$) in order to prevent the linear region operation of M2. The sweep was performed for both values of V_I for each sample. Unfortunately, the sweep had to be performed manually in order to prevent a damage on the samples, since very large supply voltages had to be applied[1].

6.2.1 Measurement Procedure

The measurement procedure consists of two main steps, which are (1) measurement of the gain and (2) measurement of the total noise at the output.

Gain Measurement

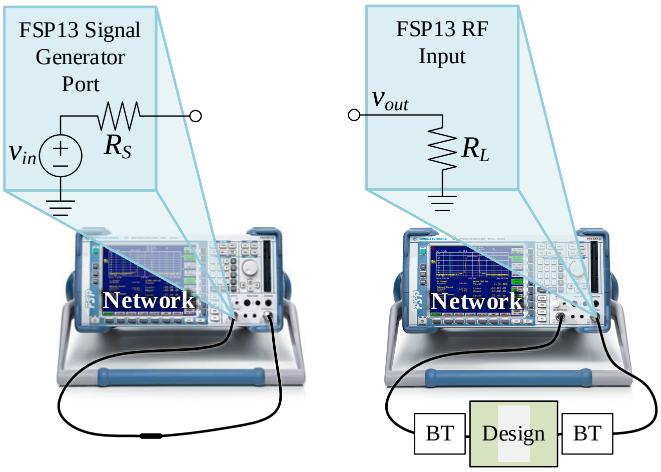
For performing the gain measurements, the network analyser function of the Rohde&Schwarz FSP13 spectrum analyser is used [143]. Before characterizing the network properties of a design by using FSP, a calibration is necessary in order to exclude any additional attenuation in the system. As it can be seen on Figure 6.7a, during the calibration, the signal generator and the RF input ports of the spectrum analyser are connected to each other with the SMA cables which are already being used in the test set-up. Each stage on Figure 6.2 except the bias-tees and what is covered by PCB is included in the calibration. The reason of excluding the bias-tees in the calibration is that they have to be present during the noise measurements for providing the same bias values of the gain measurements. If excluded, the gain of the noise measurement system differs from the measured gain by the attenuation of two bias-tees.

During the calibration step, FSP measures the power gain of the calibration set-up, which is equal to

$$A_{pC}^{[\mathrm{dB}]} = 10\log\left(\frac{P_{out}}{P_{av}}\right) = 10\log\left(\frac{v_{out}^2}{R_L}\right) - 10\log\left(\frac{v_{in}^2}{4R_S}\right) \tag{6.18}$$

---

[1] Very large supply voltages (up to 6 V) appear only on the load resistors (R_D) in order to provide the maximum allowed V_{DS} (1 V) and a large heating for M1. While doing this, attention was paid in order not to have a V_{DS} or V_{GS} value that is larger than 1 V.



(a) Calibration (b) Gain Measurement

Figure 6.7 – Configurations of FSP for different steps of the gain measurement. Bias-tees are indicated with BT.

where P_{av} is the maximum available source power that is equal to the power dissipated on a load that is complex conjugate of the source resistance. In case the load resistance (R_L) is set equal to the source resistance (R_S), $v_{out} = v_{in}/2$ and the power gain is 0 dB. However, the non-ideal behaviour of the test equipment brings some attenuation especially at higher frequencies. The calibration removes this up to some extent. After the calibration finishes, one can proceed with the network characterization. As it can be seen on Figure 6.7b, during the gain measurements, the gain from the signal generator and the RF input ports of the network analyser (covering the components on the light blue region indicated by **Measurement Device** on Figure 6.2) are connected to the output and the input of the test blocks through SMA connectors. Finally, the calibration power gain A_{pC} is removed from the measured power gain A_{pM} and the power gain of the design is extracted as

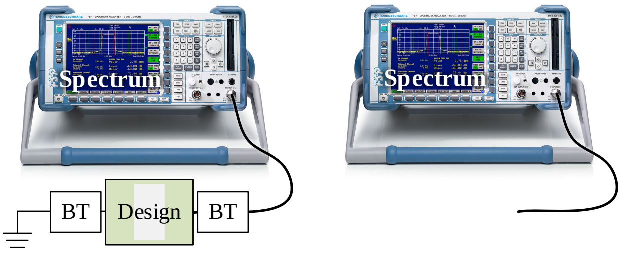
$$A_p^{[\mathrm{dB}]} = A_{pM}^{[\mathrm{dB}]} - A_{pC}^{[\mathrm{dB}]} \tag{6.19}$$

Since during the power gain calculation, the output power is normalized with respect to the maximum available source power, the relation between the voltage gain and the power gain is written as

$$A_p = \frac{P_{out}}{P_{av}} = \frac{4R_S}{R_L} A_v^2 \tag{6.20}$$

In case everything is matched at each stage (i.e. $R_L = R_{in} = R_{outS} = R_S$), the power gain of the complete system is expressed as

$$A_p^{[\mathrm{dB}]} = 20\log(A_{vA}) + 20\log(A_{vS}) \approx 20\log\left[g_{m1}\left(R_D \parallel r_{o1}\right)\right] - 7\,\mathrm{dB} \tag{6.21}$$



(a) Measurement of total noise (b) Measurement of FSP noise

Figure 6.8 – Configurations of FSP for different steps of the noise measurement.

for experiment-A and

$$A_p^{[\mathrm{dB}]} = 20\log(A_{vB}) + 20\log(A_{vS}) \approx 20\log\left(\frac{g_{m1} r_{o1} R_D A_{vg2}}{A_{vg2} r_{o1} + r_{o2} + R_D}\right) - 7\,\mathrm{dB} \tag{6.22}$$

for experiment-B.

Noise Measurement

During the noise measurements, the spectrum analyser mode of the FSP is used. As it can be seen on Figure 6.8a, the AC input of the design is grounded and the output of the design is connected to the RF input of the FSP. Like this, the internally generated noise of the design is sampled at the RF input without any externally applied signal. It should also be noted that the noise of the source resistance R_S does not appear at the output since it is disconnected from the system. The same is true for the input resistance R_{in} since the input, hence both terminals of R_{in} are shorted to ground. Therefore, they do not appear in the measured output noise power of the design, $S_{PnM,o}$. In case there is no additional noise in the measurement set-up, the input referred noise power, $S_{Pn,i}$, can be obtained by dividing the measured output noise power by the power gain according to

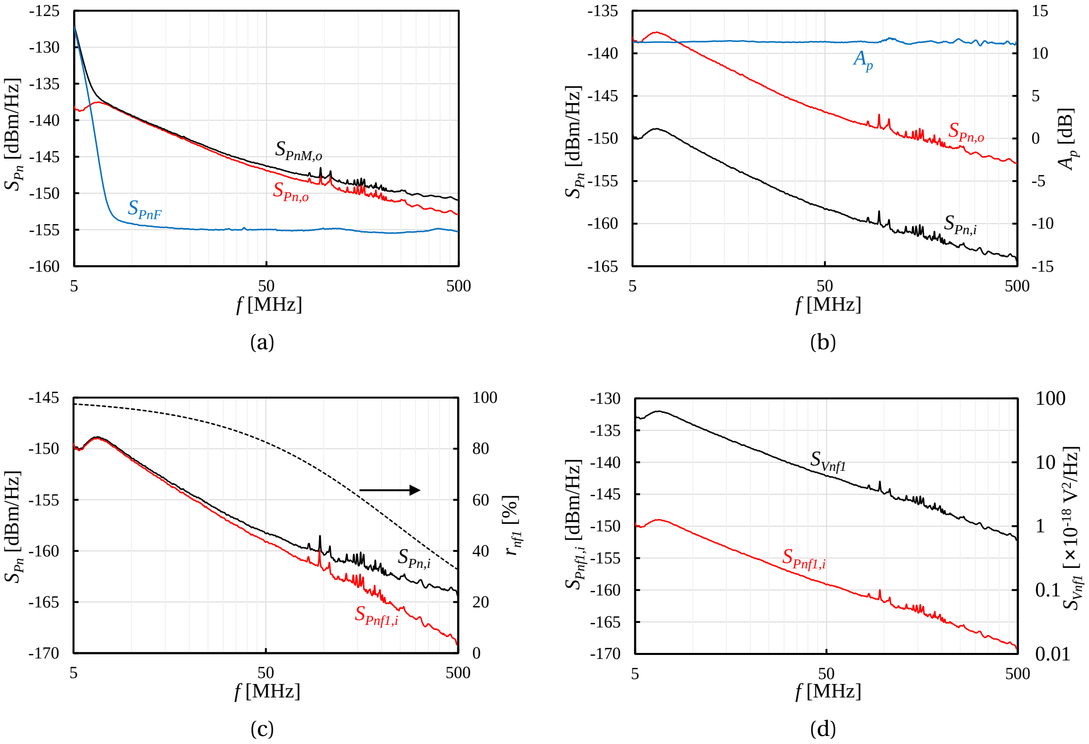
$$S_{Pn,i}^{[\mathrm{dBm}]} = S_{PnM,o}^{[\mathrm{dBm}]} - A_p^{[\mathrm{dB}]}. \tag{6.23}$$

However, (6.23) would contain some error since FSP has some internal noise and it is included in $S_{PnM,o}$ in addition to the amplified noise of the design. For this, it is mandatory to measure the input referred noise of the FSP, ${S_{PnF}}^2$, and remove it from the measured noise ($S_{PnM,o}$) in order to obtain the output noise of the design:

$$S_{Pn,i}^{[\mathrm{dBm}]} = 10\log\left(\frac{S_{PnM,o} - S_{PnF}}{1\,\mathrm{mW}}\right) - A_p^{[\mathrm{dB}]} = S_{Pn,o}^{[\mathrm{dBm}]} - A_p^{[\mathrm{dB}]}. \tag{6.24}$$

---

[2] S_{PnF} is the noise power of the spectrum analyser referred to its RF input (i.e. output of the design), not to the input of the design.



(a)

(b)

(c)

(d)

Figure 6.9 – Noise power spectral densities at different stages of the set-up for obtaining the flicker noise voltage of M1 from the measured noise power.

This situation is illustrated on Figure 6.9a where the power spectral densities of the measured output noise power ($S_{PnM,o}$), input referred internal noise power of FSP (S_{PnF}) and the extracted output noise power of the design ($S_{Pn,o}$) can be observed. It can be seen that S_{PnF} is quite flat for frequencies larger than 10 MHz. However, it is quite large for lower frequencies. This brings some uncertainty if one desires perform a measurement for lower frequencies. The reason is that the desired signal might be lost when it is added to the noise of FSP. This depends on the power of the signal that is desired to be measured. The situation can be observed for frequencies close to 5 MHz where $S_{Pn,o}$ deviates significantly from $S_{PnM,o}$ since S_{PnF} is much larger than $S_{Pn,o}$. Same observations can also be made by focusing on higher frequencies. As the frequency increases, $S_{PnM,o}$ decreases mainly due to the decrease in the flicker noise. Consequently, $S_{PnM,o}$ becomes comparable to S_{PnF} and the accuracy of the measurement is small also at very high frequencies. On the other hand, in medium frequency range, $S_{PnM,o}$ and $S_{Pn,o}$ are almost equal to each other and (6.23) provides a good approximation[3]. Therefore, 10-50 MHz band would be the most accurate region in terms of sampling the flicker noise.

After obtaining the input referred noise power spectral density, $S_{Pn,i}$, (Figure 6.9b) the next step is to extract the flicker noise. Since there are also other noise sources appearing in the input

---

[3] (6.24) is still used while extracting the flicker noise even though the influence of the internal noise of FSP is quite negligible.

referred noise expressions and since they have zero correlation with the flicker noise of M1, practically it is not possible to obtain the exact contribution of the flicker noise. Nevertheless, the contribution of the flicker noise of M1 in the total noise can be obtained by performing noise simulations. Like this, it might be possible to get an acceptable approximation by using the simulation results. For obtaining the contribution of the flicker noise at the input referred noise power, we first calculate $S_{Pn,i}$ as

$$S_{Pn,i} = \frac{1}{R_S}\left\{4kT\left[\frac{1}{g_{m1}}\left(\frac{1}{g_{m1}R_D}+\frac{\gamma_1}{\alpha_1}\right)+\frac{1}{A_{vA}^2 g_{m0}}\left(\frac{1}{g_{m0}R_L}+\frac{\gamma_0}{\alpha_0}\right)\right]+S_{Vnf1}+\frac{1}{A_{vA}^2}S_{Vnf0}\right\} \tag{6.25}$$

for experiment-A and

$$\begin{aligned} S_{Pn,i} = {} & \frac{1}{R_S}4kT\left[\frac{1}{g_{m1}}\left(\frac{1}{g_{m1}R_D}+\frac{\gamma_1}{\alpha_1}\right)+\frac{A_{vi2}^2}{A_{vi1}^2 A_{vg2}^2 g_{m2}}\left(\frac{\gamma_2}{\alpha_2}\right)+\frac{1}{A_{vA}^2 g_{m0}}\left(\frac{1}{g_{m0}R_L}+\frac{\gamma_0}{\alpha_0}\right)\right] \\ & + \frac{1}{R_S}\left(S_{Vnf1}+\frac{A_{vi2}^2}{A_{vi1}^2 A_{vg2}^2}S_{Vnf2}+\frac{1}{A_{vA}^2}S_{Vnf0}\right) \end{aligned} \tag{6.26}$$

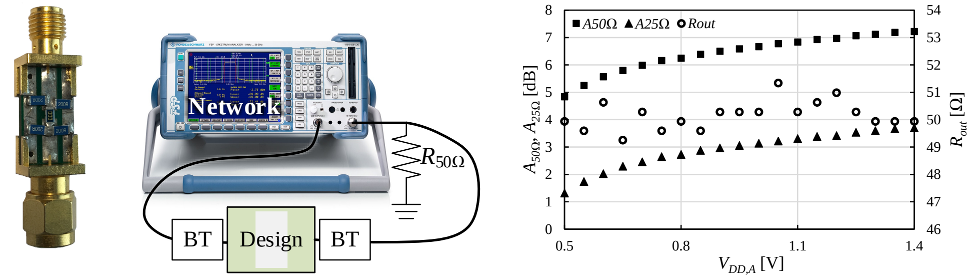
for experiment-B. By observing (6.25) and (6.26), it can be seen that some of the noise sources have very little contribution since they are located at the subsequent stages of the system and their power is divided by a large gain. On the other hand, the contributions of the noise source at the first amplifying stage depends on different parameters like transconductance and resistance. By properly adjusting them, their contribution can also be manipulated although cannot be removed. Finally, the contribution of the flicker noise sources depend also on the frequency. Therefore, by selecting a frequency band where flicker noise is dominant, one can get a more accurate measurement of the flicker noise of M1. Considering these facts, it is decided to focus on $f = 20$ MHz during the noise measurements at different bias conditions. Having the expressions of $S_{Pn,i}$, the contribution of the flicker noise of M1 is

$$r_{nf1} = \frac{S_{Pnf1,i}}{S_{Pn,i}} \tag{6.27}$$

where $S_{Pnf1,i}$ is the input referred flicker noise power of M1 and r_{nf1} stands for the ratio of $S_{Pnf1,i}$ to $S_{Pn,1}$. As it has already been explained, r_{nf1} is calculated by using the device parameters with (6.25) and (6.26). On Figure 6.9c, it can be seen that the flicker noise power of M1 is roughly 90% of the total noise power, which is sufficiently large. Having $S_{Pn,i}$ from measurements and $r_{n1,f}$ from analytic expressions and already characterized device parameters, the flicker noise of M1 is extracted by

$$S_{Vnf1} = r_{nf1} S_{Pn,i} R_S \tag{6.28}$$

which can be observed on Figure 6.9d.



(a) $R_{50\Omega}$ (b) Measurement of $A_{25\Omega}$ (c) $A_{50\Omega}$, $A_{25\Omega}$ and R_{out}.

Figure 6.10 – (a) external SMA 50 Ω feed-through termination, (b) configuration of FSP for tuning the output resistance to 50 Ω and (c) measured gain and output impedance values.

Output Matching

It has been already mentioned that the input and the output of the designs should provide a 50 Ω AC resistance for a sufficiently large band in order to prevent reflections. The matching at the input is provided by a 50 Ω external resistor (R_{in}). For this, it is guaranteed that the input is matched. On the other hand, the AC resistance at output is provided by the source follower according to (6.8). Its dynamic resistance depends on the DC operating point of the output stage. In case the device parameters of M0 deviate from their mean value due to process variations or temperature, the output resistance might deviate from 50 Ω and mismatch might occur. For this reason, at each measurement point, the bias voltages of M0 has to be tuned in order to adjust its AC resistance to 50 Ω. Measuring the AC resistance is not as straightforward as measuring the DC resistance by connecting an ohmmeter. This is due to that the ohmmeter measures the static resistance rather than the dynamic resistance. Moreover, the externally applied voltage of the ohmmeter shifts the DC bias from its desired value. For this, another technique is applied in two steps. Firstly, the gain of the design is measured at a sufficiently low frequency by using the configuration of Figure 6.7b. The value of this gain is denoted by $A_{50\Omega}$ since the load resistance is equal to 50 Ω. Then, an additional external 50 Ω resistor (shown by Figure 6.10a) is connected parallel to R_L, which reduces the load resistance to 25 Ω. The gain is measured once again and its value is denoted by $A_{25\Omega}$ with the configuration shown by Figure 6.10b. By using the gain values from these two configurations, the output impedance of the source follower is calculated as

$$R_{out} = \frac{A_{50\Omega} - A_{25\Omega}}{2A_{25\Omega} - A_{50\Omega}} 50\,\Omega \tag{6.29}$$

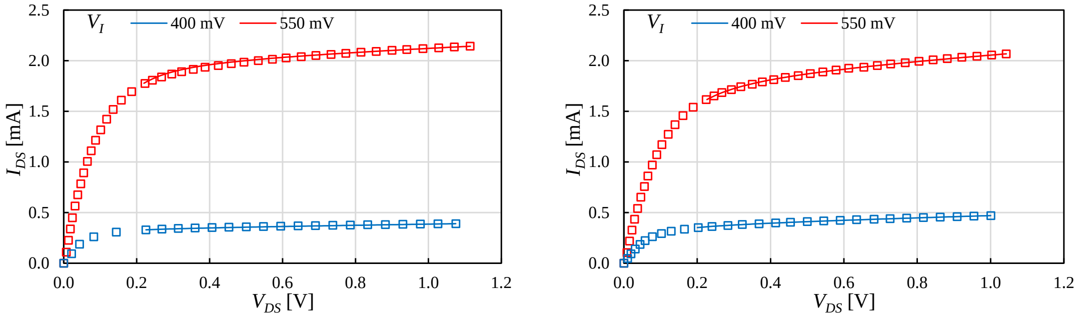
The source voltage of M0, which is provided through V_O, is adjusted until R_{out} is equal to 50 Ω. It can be also shown that once R_{out} = 50 Ω, $A_{50\Omega}$ = 1.5 $A_{25\Omega}$; in other words

$$A_{25\Omega}^{[\mathrm{dB}]} \approx A_{50\Omega}^{[\mathrm{dB}]} - 3.522\,\mathrm{dB}. \tag{6.30}$$

The measured values of $A_{50\Omega}$, $A_{25\Omega}$ and R_{out} can be observed on Figure 6.10c for the 100 nm sample under V_I = 400 mV and different supply voltage values.

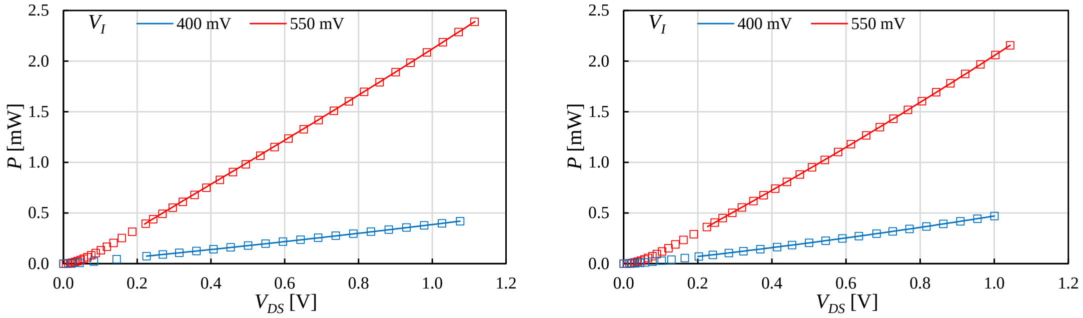
6.2.2 Results of Experiment-A

The measurements were performed according to the previously explained procedures. The results are shown on the graphs from Figure 6.11 to Figure 6.18 for the 100 nm and the 60 nm samples. The data points were measured at f = 20 MHz and they are shown by squares, while the simulation results are plotted by continuous curves on the same chart. The high heating case is always plotted with red colour, where blue is used for the low heating case. The drain current of the samples can be seen on Figure 6.11. For low heating case, V_I = 400 mV provided the desired current values, which is around 0.4 mA, for both samples. For the high heating case, V_I was set to 550 mV and the drain current was observed to be around 2 mA. Under these bias voltages, the power dissipation for the high heating case was observed to be more than five times larger than the power dissipation of the low heating case (Figure 6.12). The measured gain of the system and the simulation results according to (6.21) and (6.22) can be observed on Figure 6.13. It can be seen that the gain drops rapidly for low drain voltages due to linear region operation of M1. However, it is sufficiently large in saturation and does not change significantly for a large range of V_{DS}. The measured output noise power is plotted on Figure 6.14, which looks quite similar to the measured gain. Figure 6.15 shows the simulated temperature rise due to self-heating. It can be seen that the temperature rise of the 60 nm length sample is roughly two times the temperature of the 100 nm length sample. Figure 6.16 shows the portion of flicker noise power in the total measured noise power at f = 20 MHz according to the simulations. It can be seen that the flicker noise is at least 60% of the total noise power, which is a sufficiently large ratio. On Figure 6.17 and Figure 6.18, a second simulation data is also shown where self-heating effects are deactivated. The total input referred noise power can be observed on Figure 6.17. It can be seen that a larger amount of noise is observed when self-heating effects are activated, especially for high heating case. Moreover, the noise increases more rapidly for the 60 nm length sample than the 100 nm one. This is due to the larger temperature rise of the 60 nm sample as it has already been shown in Chapter 4. By looking at the measured data points, a similar trend can be observed, where the total input referred noise power generally increases with the drain voltage and the rate of change is larger for the 60 nm length sample. Moreover, for the same sample, the rate of change is larger in high heating compared to low heating case mainly since the temperature increase with the drain voltage is more rapid in high heating case. By comparing the simulation data with the measured noise, it can be easily seen that the input referred noise increases much more rapidly in measurements. In Chapter 2 we have already shown by device level electro-thermal simulations that the peak temperature of an FDSOI device, which is located close to the drain side of the channel, is much larger than its average temperature. Consequently, the larger increase in the noise with the power dissipation is not a surprising result as the real peak temperature in the device is larger than the estimation obtained from the circuit level simulations where the self-heating effects are activated. Finally, Figure 6.18 shows the extracted flicker noise voltage of M1. This situation in flicker noise is quite similar to the one in the total input referred noise power since a large portion of the input referred noise power is the flicker noise.



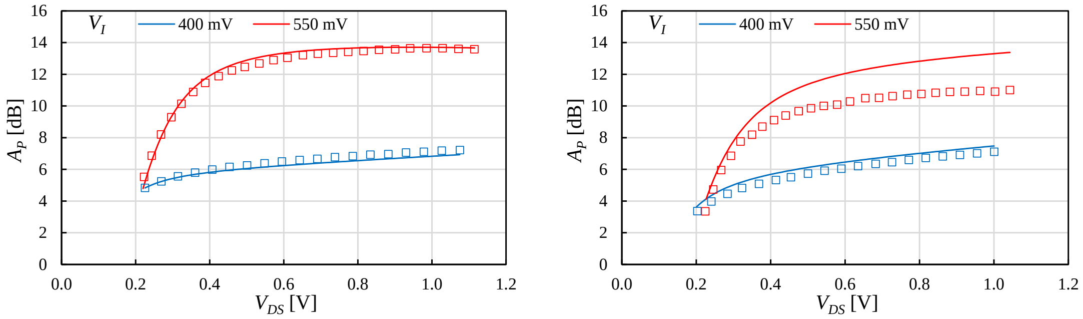
(a) $L = 100$ nm (b) $L = 60$ nm

Figure 6.11 – Drain current of M1.



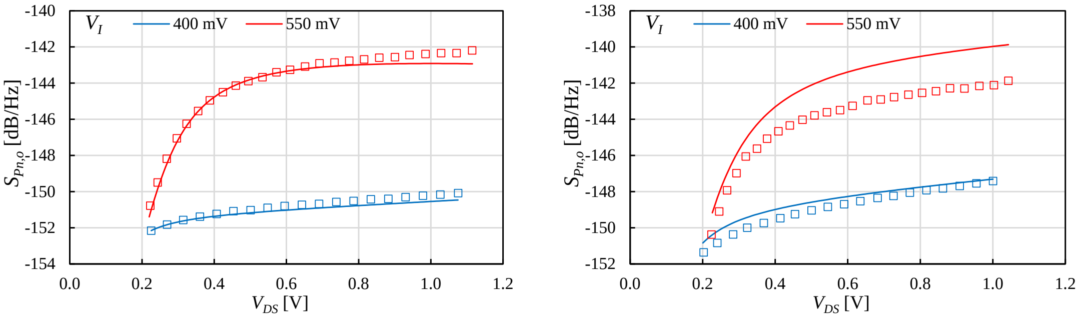
(a) $L = 100$ nm (b) $L = 60$ nm

Figure 6.12 – Power dissipation, P, of M1 ($V_{DS,M1} \times I_{D,M1}$).



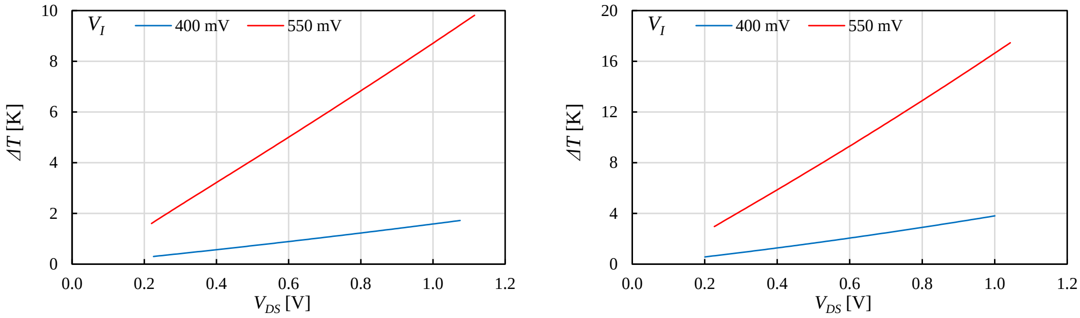
(a) $L = 100$ nm (b) $L = 60$ nm

Figure 6.13 – Gain, A_p, of M1 at 20 MHz.



(a) $L = 100$ nm (b) $L = 60$ nm

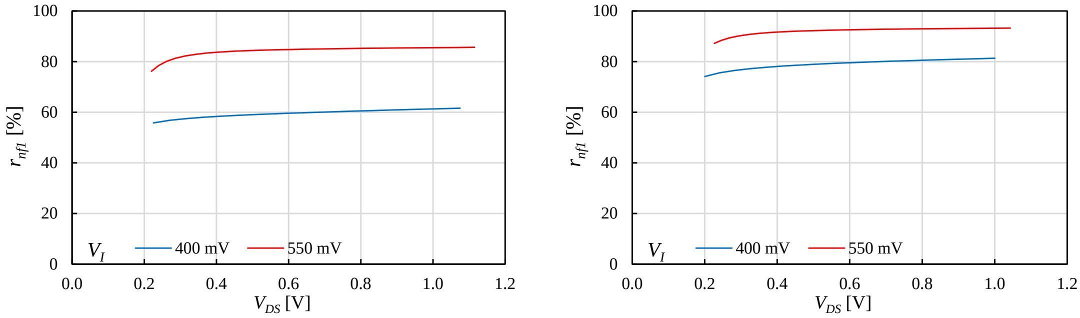
Figure 6.14 – Output noise power spectral density, $S_{Pn,o}$, of M1 at 20 MHz.



(a) $L = 100$ nm

(b) $L = 60$ nm

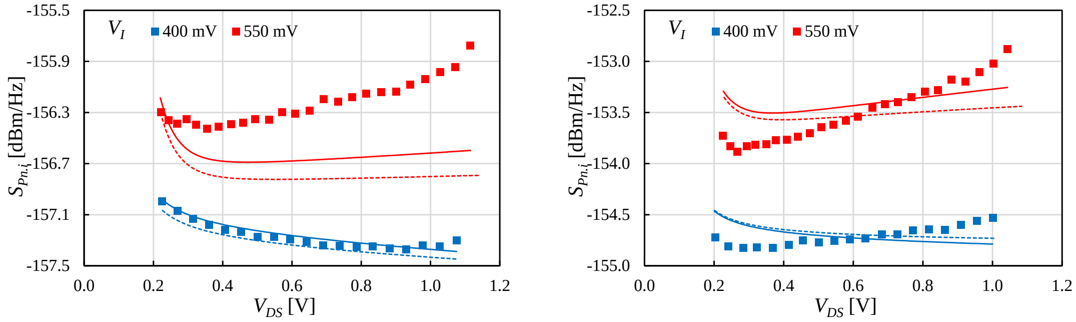
Figure 6.15 – Temperature rise, ΔT, of M1 due to its self-heating (simulation).



(a) $L = 100$ nm

(b) $L = 60$ nm

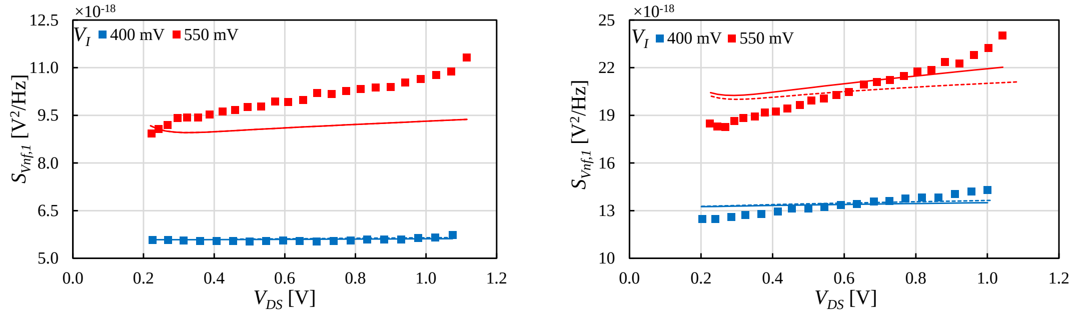
Figure 6.16 – Percentage of the flicker noise of M1 in the total noise at 20 MHz (simulation).



(a) $L = 100$ nm

(b) $L = 60$ nm

Figure 6.17 – Input referred noise power spectral density, $S_{Pn,i}$, of the entire design at 20 MHz.



(a) $L = 100$ nm

(b) $L = 60$ nm

Figure 6.18 – The power spectral density of flicker noise voltage, S_{Vnf1}, of M1 at 20 MHz.

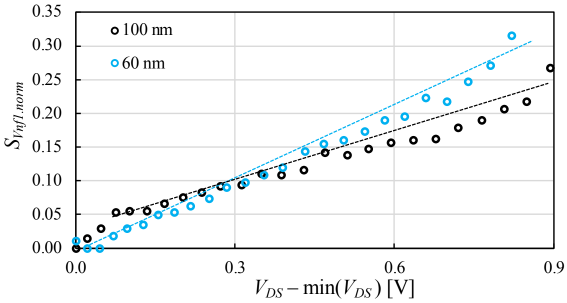


Figure 6.19 – Normalized flicker noise voltage power spectral densities of the 100 nm length and the 60 nm length devices for high heating case. The values of the horizontal axis are shifted by the smallest V_{DS} value in the measurement set.

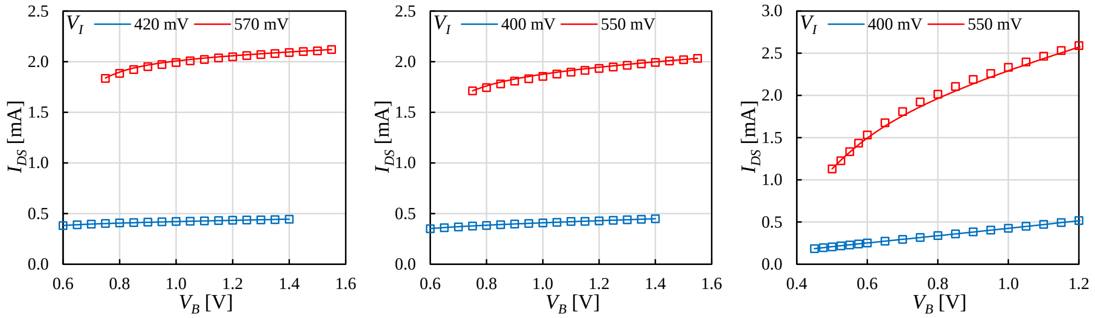
For comparing the effect of device scaling on the flicker noise degradation due to the self-heating, the high heating data on Figure 6.18 is normalized with respect to the minimum value of the flicker noise voltage according to

$$S_{Vnf1,norm} = \frac{S_{Vnf1}(V_{DS}) - \min\left[S_{Vnf1}(V_{DS})\right]}{S_{Vnf1}(V_{DS})} \tag{6.31}$$

The slope of (6.31) shows how rapidly the flicker noise increases in each case. Figure 6.19 plots the normalized flicker noise voltage of the two devices. It can be seen that the normalized flicker noise of the 60 nm length device increases more rapidly than the noise of the 100 nm length device. The situation is same also for the temperature (Figure 6.15). By comparing the noise power and the temperature data, a strong correlation between the flicker noise degradation in the temperature rise can be observed, which shows the influence of self-heating on the flicker noise.

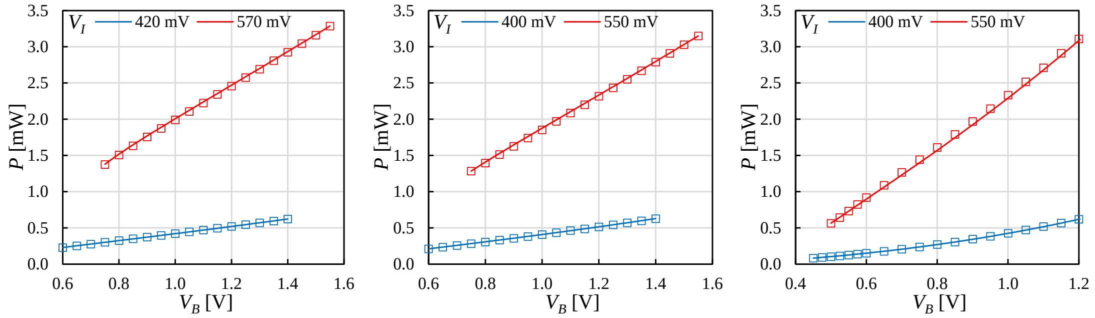
6.2.3 Results of Experiment-B

The results of experiment-B are shown on the graphs from Figure 6.20 to Figure 6.27 for the three devices with different gate lengths. The test procedure of experiment-B is quite similar to the one of experiment-A except some slight differences. The data points are measured also at f = 20 MHz. However, the main sweep parameter is the gate voltage of M2 (V_B) rather than the power supply ($V_{DD,B}$). $V_{DD,B}$ is set to a fixed value for each V_I. The desired bias conditions are precisely obtained for the 100 nm and the 60 nm samples (Figure 6.20 and Figure 6.22) similar to experiment-A. Moreover, a much flatter gain is observed with respect to V_B (Figure 6.22) thanks to the cascode device. However, for the 30 nm sample, a large dependency on V_B can be observed even with the cascode device due to its short channel effects, mainly DIBL (Figure 6.20c to Figure 6.23c). The temperature plots on Figure 6.24 show that the temperature increases as the gate length decreases, as expected. The portion of flicker noise power in the total measured noise is sufficiently large at f = 20 MHz, which is shown on Figure 6.25. Figure 6.26 shows that in each case there is a V_B bias where input referred noise power is minimum.



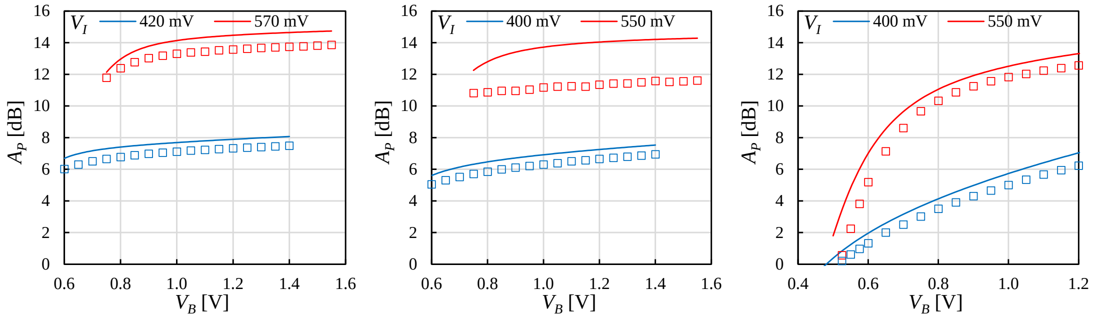
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.20 – Drain current of M1.



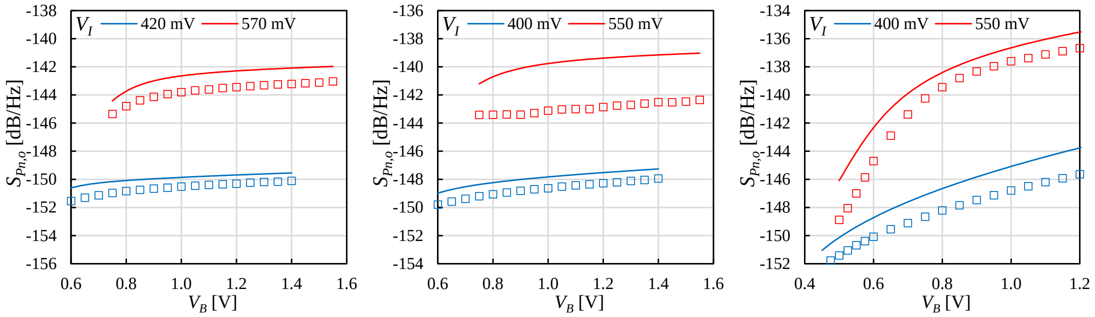
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.21 – Power dissipation, P, of M1 ($V_{DS,M1} \times I_{D,M1}$).



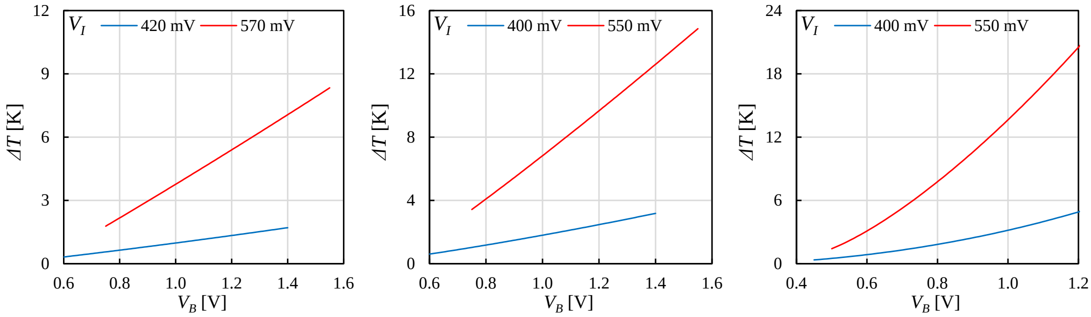
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.22 – Gain, A_p, of M1 at 20 MHz.



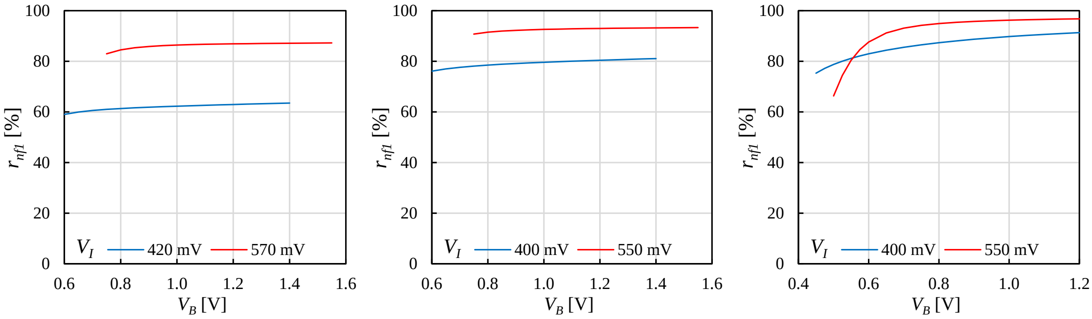
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.23 – Output noise power spectral density, $S_{Pn,o}$, of M1 at 20 MHz.



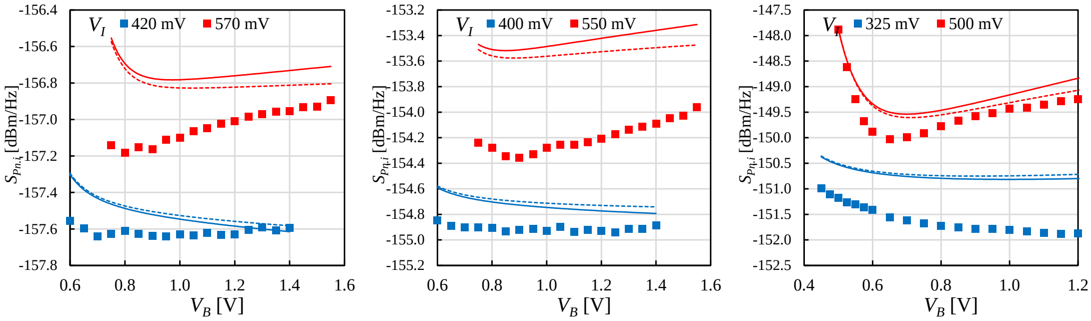
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.24 – Temperature rise, ΔT, of M1 due to its self-heating (simulation).



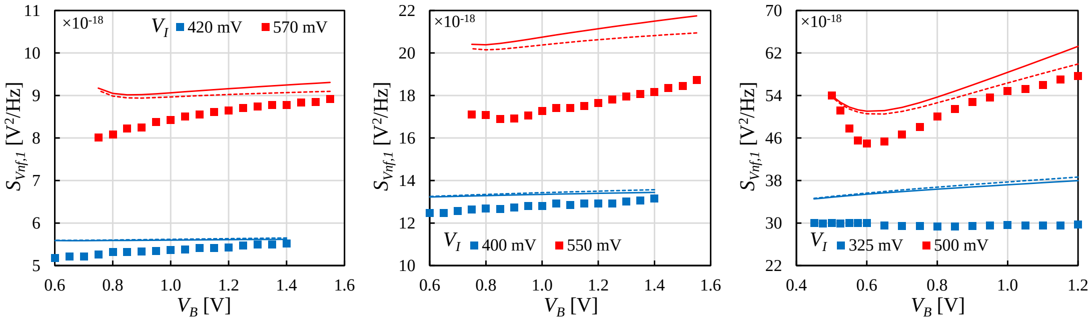
(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.25 – Percentage of the flicker noise of M1 in the total noise at 20 MHz (simulation).



(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.26 – Input referred noise power spectral density, $S_{Pn,i}$, of the entire design at 20 MHz.



(a) $L = 100$ nm (b) $L = 60$ nm (c) $L = 30$ nm

Figure 6.27 – The power spectral density of flicker noise voltage, S_{Vnf1}, of M1 at 20 MHz.

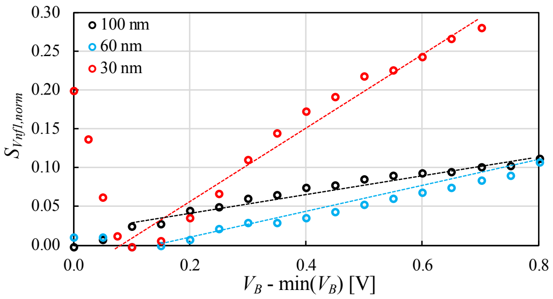


Figure 6.28 – Normalized flicker noise voltage power spectral densities of the 100 nm, 60 nm and 30 nm length devices for high heating case. The values of the horizontal axis are shifted by the smallest V_B value in the measurement set.

Moreover, the input referred noise power increases with V_B. According to Figure 6.27 the flicker noise also increases with respect to the bias voltage of the cascode device. The increase rate of both the input referred noise voltage and the flicker noise is larger as the gate length is smaller. This result is quite similar to the result obtained in experiment-A. Different than experiment-A, in experiment-B the increase rate of the flicker noise is observed to be relatively similar to the simulation results for the 60 nm (Figure 6.27b) and the 30 nm (Figure 6.27c) samples. However, in the 100 nm sample experiments, the flicker noise increases more rapidly than the simulation results. Nevertheless, the increase rate is still smaller than the 60 nm sample.

To make a fairer comparison between the samples, the normalized flicker noise is calculated according to (6.31) also in experiment-B. However, V_{DS} in (6.31) is changed by V_B since the sweep parameter is V_B rather than $V_{DD,A}$. The normalized flicker noise of each sample is plotted on Figure 6.28. It can be seen that the noise increases more rapidly as the gate length decreases. In fact, the slope is much higher for the 30 nm case compared to the larger gate length samples due to its very large heating. Considering these results, one should provide a relatively low drain bias in order to prevent excessive flicker noise and consider the self-heating calculations in order to obtain the lowest noise figure.

6.3 Conclusion

To observe the flicker noise behaviour under self-heating effects, electrical measurements on two different experiments are performed. It is observed that the input referred noise power and flicker noise increases more rapidly as the drain voltage and the power dissipation of a device is increased, mainly due to the larger value of device temperature resulting from self-heating. It is also shown that the increase rate of the flicker noise is larger for smaller gate length devices. This is due to their larger temperature values resulting from their large heat density under the same power dissipation due to their smaller areas. Finally, it is observed that the increase in noise is larger than the simulation results due to that the temperature of the hot spot is expected to be larger than the average temperature of the device.

7 Conclusion

In this thesis, we have investigated the self-heating effects in nanometer scale bulk and FDSOI devices. As it has been explained, the self-heating effects have become much more prominent in the advanced geometries than the traditional bulk devices due to the larger thermal resistance of the constitutive materials. Moreover, it became more prominent as the technology advanced. Consequently, ignoring self-heating effects in a design became out of question in advanced technologies. For this reason, we have provided different circuit examples where self-heating effects change the performance and reliability considerably. During the exploration of self-heating effects, we have first focused on a single device, where we have performed device level electro-thermal simulations. Second, we have explored the reliability of high-speed arithmetic blocks under the influence of self-heating effects. In the third part, we explored the noise figure performance of low noise amplifiers considering self-heating effects.

The significance of self-heating effects were first shown by device level electrothermal simulations. A comparison between the FDSOI and bulk geometries was performed in 40 nm and 28 nm technology nodes. It was shown that the temperature increase in FDSOI was much larger than bulk. Moreover, the simulations in two different technology nodes showed that the self-heating effect became more pronounced in FDSOI with device scaling. It was demonstrated that the peak temperature in a device operating in saturation region was close to its drain terminal, which made the drain contact more critical in terms of long-term reliability. With the electrothermal simulations on FDSOI devices under different bias conditions, it was shown that the spatial heat generation inside a device had significant dependency on the region of operation (linear or saturation). It was pointed out that deep saturation operation had to be prevented since the heat generation profile was flatter in linear region, whereas it was concentrated locally on the drain terminal in saturation region.

Second contribution of this thesis is to investigate the reliability of high-speed digital circuits considering self-heating effects. To this aim, two 64-bit Kogge-Stone parallel prefix adders were implemented in 40 nm bulk and 28 nm FDSOI technologies. Thermal simulations were performed on both adders using bulk and FDSOI thermal geometries. It was shown that the temperature profile throughout the entire circuit was much flatter in bulk than FDSOI, thanks to the large thermal conductance of bulk silicon. It was also demonstrated that the average heat density and the peak temperature in FDSOI had increased while passing from 40 nm to 28 nm technology. The comparison between FDSOI and bulk also showed that it was necessary to increase thermal simulation resolution in FDSOI to get more accurate results, as the hot-spots were much smaller in FDSOI than bulk. Another point to be made was that the simulation time and the resource utilization was significantly larger in FDSOI than bulk, primarily due to the different layers and the complexity of the FDSOI geometry. By performing extensive heat density analyses considering each device on both implementations, it was observed that some group of devices had much larger heat density than the others. By using these results, it was explained that the heating profile could be made much flatter by proper sizing of the devices in the critical groups to increase the reliability.

The third and final contribution of this thesis is to show the degradation in the noise performance of analog implementations, where noise is a critical parameter. For this, first the thermal noise current and the input referred thermal noise voltage of a single device in FDSOI was examined. It was shown that both the thermal noise current and input referred thermal noise voltage increased considerably by decreasing the device gate length and increasing the drain bias voltage. To explore self-heating effects in thermal noise further, three common gate cascode LNAs were implemented with different device sizes. The measurements indicated that the noise figure of the LNAs with short channels were significantly degraded by increasing the reference bias current and bias voltage of the cascode device, due to their large heat density. On the other hand, no significant dependence on these parameters were observed on the long channel implementations since their heat density was not very large. Similar investigations and measurements were also performed to observe the self-heating effects on the flicker noise. The measurements showed that the flicker noise performance was also degraded more in short channel implementations under large bias conditions.

The different analyses and experiments on different circuit examples demonstrated that the self-heating effects do have a significant influence on the circuit reliability and performance in FDSOI. Moreover, these effects will become more prominent in the future technologies. For this, the technology providers are urged to provide realistic self-heating models and the designers are encouraged to consider the outcomes in the design in order to obtain better results.

Bibliography

[1] J. R. Yuan, C. Svensson, and P. Larsson. New domino logic precharged by clock and data. *Electronics Letters*, 29(25):2188–2189, Dec 1993.

[2] Rozeau O, M. A. Jaud, T. Poiroux, and M. Benosman. *UTSOI Model 1.1.4: Surface Potetntial Model for Ultra Thin Fully Depleted SOI MOSFET.* LETI: Laboratoire d'électronique et de technologie de l'information, November 2012.

[3] S. Vangal, N. Borkar, E. Seligman, V. Govindarajulu, V. Erraguntla, H. Wilson, A. Pangal, V. Veeramachaneni, M. Anders, J. Tschanz, Y. Ye, D. Somasekhar, B. Bloechel, G. Dermer, R. Krishnamurthy, S. Narendra, M. Stan, S. Thompson, V. De, and S. Borkar. 5ghz 32b integer-execution core in 130nm dual-v/sub t/ cmos. In *2002 IEEE International Solid-State Circuits Conference. Digest of Technical Papers (Cat. No.02CH37315)*, volume 2, pages 334–535, Feb 2002.

[4] G. E. Moore. Cramming more components onto integrated circuits. *Electronics,*, 38(3): 114–117, April 1965.

[5] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc. Design of ion-implanted mosfet's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, Oct 1974.

[6] S.M. Sze. *Physics of Semiconductor Devices.* Wiley-Interscience publication. John Wiley & Sons, 1981.

[7] R.F. Pierret. *Semiconductor Device Fundamentals.* Addison-Wesley, 1996.

[8] *Medici Two-Dimensional Device Simulation Program User Manual.* Synopsys, Feb 2003.

[9] Paul Ampadu David Wolpert. *Managing Temperature Effects in Nanoscale Adaptive Systems.* Springer, New York, NY, USA, 2nd edition, 2012.

[10] E. Pop and K. E. Goodson. Thermal phenomena in nanoscale transistors. In *Thermal and Thermomechanical Phenomena in Electronic Systems, 2004. ITHERM '04. The Ninth Intersociety Conference on*, pages 1–7 Vol.1, June 2004.

[11] D. Vasileska. Modeling self-heating in nanoscale devices. In *2015 IEEE 15th International Conference on Nanotechnology (IEEE-NANO)*, pages 200–203, July 2015.

[12] *Pulsed I-V Testing for Components and Semiconductor Devices*. Keithley A Tetronix Company, Jan 2014.

[13] K. A. Jenkins, J. Y. C. Sun, and J. Gautier. Characteristics of soi fet's under pulsed conditions. *IEEE Transactions on Electron Devices*, 44(11):1923–1930, Nov 1997.

[14] C. Anghel, A. M. Ionescu, N. Hefyene, and R. Gillon. Self-heating characterization and extraction method for thermal resistance and capacitance in high voltage mosfets. In *European Solid-State Device Research, 2003. ESSDERC '03. 33rd Conference on*, pages 449–452, Sept 2003.

[15] C. Anghel, R. Gillon, and A. M. Ionescu. Self-heating characterization and extraction method for thermal resistance and capacitance in hv mosfets. *IEEE Electron Device Letters*, 25(3):141–143, March 2004.

[16] N. Rodriguez, C. Navarro, F. Andrieu, O. Faynot, F. Gamiz, and S. Cristoloveanu. Self-heating effects in ultrathin fd soi transistors. In *IEEE 2011 International SOI Conference*, pages 1–2, Oct 2011.

[17] Y. Qu, X. Lin, J. Li, R. Cheng, X. Yu, Z. Zheng, J. Lu, B. Chen, and Y. Zhao. Ultra fast (<1 ns) electrical characterization of self-heating effect and its impact on hot carrier injection in 14nm finfets. In *2017 IEEE International Electron Devices Meeting (IEDM)*, pages 39.2.1–39.2.4, Dec 2017.

[18] Eric Pop. *Self Heating and Scaling of Thin Body Transistors*. PhD thesis, -Stanford University, 2005.

[19] E. Pop, S. Sinha, and K. E. Goodson. Heat generation and transport in nanometer-scale transistors. *Proceedings of the IEEE*, 94(8):1587–1601, Aug 2006. ISSN 0018-9219.

[20] Claudio Fiegna, Yang Yang, Enrico Sangiorgi, and Anthony G O'Neill. Analysis of self-heating effects in ultrathin-body soi mosfets by device simulation. *Electron Devices, IEEE Transactions on*, 55(1):233–244, 2008.

[21] M Braccioli, G Curatola, Y Yang, E Sangiorgi, and C Fiegna. Simulation of self-heating effects in different soi mos architectures. *Solid-State Electronics*, 53(4):445–451, 2009.

[22] Toufik Sadi, Robert W Kelsall, and Neil J Pilgrim. Electrothermal monte carlo simulation of submicrometer si/sige modfets. *Electron Devices, IEEE Transactions on*, 54(2):332–339, 2007.

[23] D. Vasileska, K. Raleva, and S. M. Goodnick. Self-heating effects in nanoscale fd soi devices: The role of the substrate, boundary conditions at various interfaces, and the dielectric material type for the box. *IEEE Transactions on Electron Devices*, 56(12): 3064–3071, Dec 2009. ISSN 0018-9383.

[24] Katerina Raleva Dragica Vasileska and Stephen M. Goodnick. Heating effects in nanoscale devices. In Dragica Vasileska, editor, *Cutting Edge Nanotechnology*, chapter 3, pages 33–60. InTech, 2010.

[25] Dragica Vasileska, Katerina Raleva, and Stephen M Goodnick. *Heating effects in nanoscale devices*. INTECH Open Access Publisher, 2010.

[26] K. Raleva, D. Vasileska, A. Hossain, S.-K. Yoo, and S. M. Goodnick. Study of self-heating effects in soi and conventional mosfets with electro-thermal particle-based device simulator. *Journal of Computational Electronics*, 11(1):106–117, 2012.

[27] Sreekant V Narumanchi, Jayathi Y Murthy, and Cristina H Amon. Comparison of different phonon transport models for predicting heat conduction in silicon-on-insulator transistors. *Journal of Heat Transfer*, 127(7):713–723, 2005.

[28] Tianjiao Wang. *Sub-micron thermal transport in ultra-scaled metal oxide semiconductor (MOS) devices*. ProQuest, 2007.

[29] Chunjian Ni. *Phonon transport models for heat conduction in sub-micron geometries with application to microelectronics*. PhD thesis, School of Mechanical Engineering, Purdue University, 2009.

[30] Taichi Misawa, Shusuke Oki, and Yuji Awano. Quasi self-consistent monte carlo particle simulations of local heating properties in nano-scale gallium nitride fets. In *Simulation of Semiconductor Processes and Devices (SISPAD), 2013 International Conference on*, pages 308–311. IEEE, 2013.

[31] T Azuhata, T Matsunaga, K Shimada, K Yoshida, T Sota, K Suzuki, and S Nakamura. Optical phonons in gan. *Physica B: Condensed Matter*, 219:493–495, 1996.

[32] Chunjian Ni, Zlatan Aksamija, Jayathi Y. Murthy, and Umberto Ravaioli. Coupled electro-thermal simulation of mosfets. *Journal of Computational Electronics*, 11(1):93–105, 2012.

[33] M. Mohamed, Z. Aksamija, W. Vitale, F. Hassan, Kyeong-Hyun Park, and U. Ravaioli. A conjoined electron and thermal transport study of thermal degradation induced during normal operation of multigate transistors. *IEEE Transactions on Electron Devices*, 61(4): 976–983, April 2014. ISSN 0018-9383.

[34] C. Baltacı and Y. Leblebici. Thermal issues in deep sub-micron fdsoi circuits. In *2016 12th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME)*, pages 1–4, June 2016.

[35] Jie Chen, Gang Zhang, and Baowen Li. Thermal contact resistance across nanoscale silicon dioxide and silicon interface. *Journal of Applied Physics*, 112(6):064319, 2012.

[36] E. Lampin, Q.-H. Nguyen, P. A. Francioso, and F. Cleri. Thermal boundary resistance at silicon-silica interfaces by molecular dynamics simulations. *Applied Physics Letters*, 100 (13):131906, 2012.

[37] Yannis Tsividis. *Operation and Modeling of the MOS Transistor.* McGraw-Hill, Inc., New York, NY, USA, 1987.

[38] Christian C. Enz and Eric A. Vittoz. *Static Drain Current*, pages 33–54. John Wiley & Sons, Ltd, 2006.

[39] S. S. Bhattacharya, S. K. Banerjee, B. Y. Nguyen, and P. J. Tobin. Temperature dependence of the anomalous leakage current in polysilicon-on-insulator mosfet's. *IEEE Transactions on Electron Devices*, 41(2):221–227, Feb 1994.

[40] Haihua Su, F. Liu, A. Devgan, E. Acar, and S. Nassif. Full chip leakage-estimation considering power supply and temperature variations. In *Low Power Electronics and Design, 2003. ISLPED '03. Proceedings of the 2003 International Symposium on*, pages 78–83, Aug 2003.

[41] M. L. Mui, K. Banerjee, and A. Mehrotra. Power supply optimization in sub-130 nm leakage dominant technologies. In *Quality Electronic Design, 2004. Proceedings. 5th International Symposium on*, pages 409–414, 2004.

[42] M. Pedram and S. Nazarian. Thermal modeling, analysis, and management in vlsi circuits: Principles and methods. *Proceedings of the IEEE*, 94(8):1487–1501, Aug 2006.

[43] J. R. Black. Electromigration-a brief survey and some recent results. *IEEE Transactions on Electron Devices*, 16(4):338–347, Apr 1969.

[44] Ping-Chung Li and T. K. Young. Electromigration: the time bomb in deep-submicron ics. *IEEE Spectrum*, 33(9):75–78, Sep 1996.

[45] Werner Steinhögl, Günther Schindler, Gernot Steinlesberger, and Manfred Engelhardt. Size-dependent resistivity of metallic wires in the mesoscopic range. *Phys. Rev. B*, 66: 075414, Aug 2002.

[46] Sungjun Im, N. Srivastava, K. Banerjee, and K. E. Goodson. Scaling analysis of multilevel interconnect temperatures for high-performance ics. *IEEE Transactions on Electron Devices*, 52(12):2710–2719, Dec 2005.

[47] W. Steinhögl, G. Schindler, G. Steinlesberger, M. Traving, and M. Engelhardt. Comprehensive study of the resistivity of copper wires with lateral dimensions of 100 nm and smaller. *Journal of Applied Physics*, 97(2):023706, 2005.

[48] A. H. Ajami, M. Pedram, and K. Banerjee. Effects of non-uniform substrate temperature on the clock signal integrity in high performance designs. In *Proceedings of the IEEE 2001 Custom Integrated Circuits Conference (Cat. No.01CH37169)*, pages 233–236, 2001.

[49] Toshiki Kanamoto, Takaaki Okumura, Katsuhiro Furukawa, Hiroshi Takafuji, Atsushi Kurokawa, Koutaro Hachiya, Tsuyoshi Sakata, Masakazu Tanaka, Hidenari Nakashima, Hiroo Masuda, Takashi Sato, and Masanori Hashimoto. Impact of self-heating in wire interconnection on timing. *IEICE Transactions*, 93-C(3):388–392, 2010.

[50] A. H. Ajami, K. Banerjee, and M. Pedram. Modeling and analysis of nonuniform substrate temperature effects on global ulsi interconnects. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(6):849–861, June 2005.

[51] S. Lin and H. Yang. Analytical thermal analysis of on-chip interconnects. In *2006 International Conference on Communications, Circuits and Systems*, volume 4, pages 2776–2780, June 2006.

[52] Kaustav Banerjee, Massoud Pedram, and Amir H. Ajami. Analysis and optimization of thermal issues in high-performance vlsi. In *Proceedings of the 2001 International Symposium on Physical Design*, ISPD '01, pages 230–237, New York, NY, USA, 2001.

[53] A. H. Ajami, K. Bnerjee, M. Pedram, and L. P. P. P. van Ginneken. Analysis of non-uniform temperature-dependent interconnect performance in high performance ics. In *Proceedings of the 38th Design Automation Conference (IEEE Cat. No.01CH37232)*, pages 567–572, June 2001.

[54] Toshiki KANAMOTO, Takaaki OKUMURA, Katsuhiro FURUKAWA, Hiroshi TAKAFUJI, Atsushi KUROKAWA, Koutaro HACHIYA, Tsuyoshi SAKATA, Masakazu TANAKA, Hidenari NAKASHIMA, Hiroo MASUDA, Takashi SATO, and Masanori HASHIMOTO. Impact of self-heating in wire interconnection on timing. *IEICE Transactions on Electronics*, E93.C (3):388–392, 2010.

[55] Yijiang Shen, Ngai Wong, and E. Y. Lam. Interconnect thermal simulation with higher order spatial accuracy. In *APCCAS 2008 - 2008 IEEE Asia Pacific Conference on Circuits and Systems*, pages 566–569, Nov 2008.

[56] A. Chakraborty, K. Duraisami, A. Sathanur, P. Sithambaram, L. Benini, A. Macii, E. Macii, and M. Poncino. Dynamic thermal clock skew compensation using tunable delay buffers. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 16(6):639–649, June 2008.

[57] T. Ragheb, A. Ricketts, M. Mondal, S. Kirolos, G. M. Links, V. Narayanan, and Y. Massoud. Design of thermally robust clock trees using dynamically adaptive clock buffers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 56(2):374–383, Feb 2009.

[58] X. Chen, W. R. Davis, and P. D. Franzon. Adaptive clock distribution for 3d integrated circuits. In *2011 IEEE 20th Conference on Electrical Performance of Electronic Packaging and Systems*, pages 91–94, Oct 2011.

[59] D. Wolpert and P. Ampadu. Exploiting programmable temperature compensation devices to manage temperature-induced delay uncertainty. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 59(4):735–748, April 2012.

[60] D. Pham, S. Asano, M. Bolliger, M. N. Day, H. P. Hofstee, C. Johns, J. Kahle, A. Kameyama, J. Keaty, Y. Masubuchi, M. Riley, D. Shippy, D. Stasiak, M. Suzuoki, M. Wang, J. Warnock, S. Weitzel, D. Wendel, T. Yamazaki, and K. Yazawa. The design and implementation of a first-generation cell processor - a multi-core soc. In *2005 International Conference on Integrated Circuit Design and Technology, 2005. ICICDT 2005.*, pages 49–52, May 2005.

[61] F. Kaplan, C. De Vivero, S. Howes, M. Arora, H. Homayoun, W. Burleson, D. Tullsen, and A. K. Coskun. Modeling and analysis of phase change materials for efficient thermal management. In *2014 IEEE 32nd International Conference on Computer Design (ICCD)*, pages 256–263, Oct 2014.

[62] J. D. Davis, P. A. Bunce, D. M. Henderson, Y. H. Chan, U. Srinivasan, D. Rodko, P. Patel, T. J. Knips, and T. Werner. 7ghz l1 cache srams for the 32nm zenterprise ec12 processor. In *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, pages 324–325, Feb 2013.

[63] J. Warnock. 4.1 22nm next-generation ibm system z microprocessor. In *Solid- State Circuits Conference - (ISSCC), 2015 IEEE International*, pages 1–3, Feb 2015.

[64] Y. C. Chen, S. Ladenheim, H. Kalargaris, M. Mihajlović, and V. F. Pavlidis. Computationally efficient standard-cell fem-based thermal analysis. In *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 490–495, Nov 2017.

[65] Jie Lai and Arun Majumdar. Concurrent thermal and electrical modeling of submicrometer silicon devices. *Journal of Applied Physics*, 79(9):7353–7361, 1996.

[66] S. Sinha, E. Pop, R. W. Dutton, and K. E. Goodson. Non-equilibrium phonon distributions in sub-100nm silicon transistors. *Journal of Heat Transfer*, 128(7):638–647, Dec 2005.

[67] N. Allec, Z. Hassan, L. Shang, R. P. Dick, and R. Yang. Thermalscope: Multi-scale thermal analysis for nanometer-scale integrated circuits. In *2008 IEEE/ACM International Conference on Computer-Aided Design*, pages 603–610, Nov 2008.

[68] Z. Hassan, N. Allec, L. Shang, R. P. Dick, V. Venkatraman, and R. Yang. Multiscale thermal analysis for nanometer-scale integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 28(6):860–873, June 2009.

[69] Z. Hassan, N. Allec, F. Yang, L. Shang, R. P. Dick, and X. Zeng. Full-spectrum spatial-temporal dynamic thermal analysis for nanometer-scale integrated circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 19(12):2276–2289, Dec 2011.

[70] Krishna Pradeep. Exploring the nano-scale self-heating mechanisms in soi/bulk mos devices. Master's thesis, -EPFL, 2015.

[71] M.N. Ozisik. *Boundary Value Problems of Heat Conduction*. Dover phoenix editions. Dover Publications, 2002.

[72] Z. Yu, D. Yergeau, and R. W. Dutton. Full chip thermal simulation. In *Quality Electronic Design, 2000. ISQED 2000. Proceedings. IEEE 2000 First International Symposium on*, pages 145–149, 2000.

[73] Peng Li, L. T. Pileggi, M. Asheghi, and R. Chandra. Ic thermal simulation and modeling via efficient multigrid-based approaches. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 25(9):1763–1776, Sept 2006.

[74] D. Oh, C. C. P. Chen, and Y. H. Hu. Efficient thermal simulation for 3-d ic with thermal through-silicon vias. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(11):1767–1771, Nov 2012.

[75] W. Yu, T. Zhang, X. Yuan, and H. Qian. Fast 3-d thermal simulation for integrated circuits with domain decomposition method. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(12):2014–2018, Dec 2013.

[76] Wei Huang, M. R. Stan, and K. Skadron. Parameterized physical compact thermal modeling. *IEEE Transactions on Components and Packaging Technologies*, 28(4):615–622, Dec 2005.

[77] M. Asheghi, M. N. Touzelbaev, K. E. Goodson, Y. K. Leung, and S. S. Wong. Temperature-dependent thermal conductivity of single-crystal silicon layers in soi substrates. *Journal of Heat Transfer*, 120:30–36, Feb 1998.

[78] Kjell Bløtekjær. Transport equations for electrons in two-valley semiconductors. *Electron Devices, IEEE Transactions on*, 17(1):38–47, 1970.

[79] R Stratton. Diffusion of hot and cold electrons in semiconductor barriers. *Physical Review*, 126(6):2002, 1962.

[80] *Sentaurus Device User Guide*. Synopsys, 2012.

[81] DBM Klaassen. A unified mobility model for device simulation—i. model equations and concentration dependence. *Solid-State Electronics*, 35(7):953–959, 1992.

[82] R.F. Pierret. *Advanced Semiconductor Fundamentals*. Modular series on solid state devices. Addison-Wesley Publishing Company, 1987.

[83] E Hi Sondheimer. The mean free path of electrons in metals. *Advances in physics*, 1(1): 1–42, 1952.

[84] Wei Huang, K. Skadron, S. Gurumurthi, R. J. Ribando, and M. R. Stan. Differentiating the roles of ir measurement and simulation for power and temperature-aware design. In *Performance Analysis of Systems and Software, 2009. ISPASS 2009. IEEE International Symposium on*, pages 1–10, April 2009.

[85] Can Baltacı and Yusuf Leblebici. Thermal aware design and comparative analysis of a high performance 64-bit adder in fd-soi and bulk cmos technologies. *Integration, the VLSI Journal*, 58:421 – 429, 2017.

[86] T.M. Tritt. *Thermal Conductivity: Theory, Properties, and Applications.* Physics of Solids and Liquids. Springer, 2004.

[87] Tsuneyuki Yamane, Naoto Nagai, Shin-ichiro Katayama, and Minoru Todoki. Measurement of thermal conductivity of silicon dioxide thin films using a 3 omega method. *Journal of Applied Physics*, 91(12):9772–9776, 2002.

[88] S.-M. Lee and David G. Cahill. Heat transport in thin dielectric films. *Journal of Applied Physics*, 81(6):2590–2595, 1997.

[89] P. M. Kogge and H. S. Stone. A parallel algorithm for the efficient solution of a general class of recurrence equations. *IEEE Transactions on Computers*, C-22(8):786–793, Aug 1973. ISSN 0018-9340.

[90] F. K. Gurkaynak, Y. Leblebicit, L. Chaouati, and P. J. McGuinness. Higher radix kogge-stone parallel prefix adder architectures. In *Circuits and Systems, 2000. Proceedings. ISCAS 2000 Geneva. The 2000 IEEE International Symposium on*, volume 5, pages 609–612 vol.5, 2000.

[91] Jaehong Park, H. C. Ngo, J. A. Silberman, and S. H. Dhong. 470 ps 64-bit parallel binary adder [for cpu chip]. In *2000 Symposium on VLSI Circuits. Digest of Technical Papers (Cat. No.00CH37103)*, pages 192–193, June 2000.

[92] Y. Shimazaki, R. Zlatanovici, and B. Nikolic. A shared-well dual-supply-voltage 64-bit alu. *IEEE Journal of Solid-State Circuits*, 39(3):494–500, March 2004. ISSN 0018-9200.

[93] S. Mathew, M. Anders, R. Krishnamurthy, and S. Borkar. A 4 ghz 130 nm address generation unit with 32-bit sparse-tree adder core. In *2002 Symposium on VLSI Circuits. Digest of Technical Papers (Cat. No.02CH37302)*, pages 126–127, June 2002.

[94] S. Mathew, R. Krishnamurthy, M. Anders, R. Rios, K. Mistry, and K. Soumyanath. Sub-500 ps 64 b alus in 0.18 /spl mu/m soi/bulk cmos: Design amp; scaling trends. In *2001 IEEE International Solid-State Circuits Conference. Digest of Technical Papers. ISSCC (Cat. No.01CH37177)*, pages 318–319, Feb 2001.

[95] R. Zlatanovici and B. Nikolic. Power-performance optimal 64-bit carry-lookahead adders. In *Solid-State Circuits Conference, 2003. ESSCIRC '03. Proceedings of the 29th European*, pages 321–324, Sept 2003.

[96] S. K. Mathew, R. K. Krishnamurthy, M. A. Anders, R. Rios, K. R. Mistry, and K. Soumyanath. Sub-500-ps 64-b alus in 0.18- mu;m soi/bulk cmos: design and scaling trends. *IEEE Journal of Solid-State Circuits*, 36(11):1636–1646, Nov 2001.

[97] S. Mathew, M. Anders, R. K. Krishnamurthy, and S. Borkar. A 4-ghz 130-nm address generation unit with 32-bit sparse-tree adder core. *IEEE Journal of Solid-State Circuits*, 38(5):689–695, May 2003.

[98] R. Rafati, S. M. Fakhraie, and K. C. Smith. Low-power data-driven dynamic logic (d3l) [cmos devices]. In *2000 IEEE International Symposium on Circuits and Systems. Emerging Technologies for the 21st Century. Proceedings (IEEE Cat No.00CH36353)*, volume 1, pages 752–755 vol.1, 2000.

[99] F. Frustaci, M. Lanuzza, P. Zicari, S. Perri, and P. Corsonello. Designing high-speed adders in power-constrained environments. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 56(2):172–176, Feb 2009.

[100] P. I. J. Chuang, D. Li, M. Sachdev, and V. Gaudet. A 148ps 135mw 64-bit adder with constant-delay logic in 65nm cmos. In *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference*, pages 1–4, Sept 2012.

[101] S. K. Chang and C. L. Wey. A fast 64-bit hybrid adder design in 90nm cmos process. In *2012 IEEE 55th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 414–417, Aug 2012.

[102] Y. S. Wang, M. H. Hsieh, C. M. Liu, Y. C. Wu, B. F. Lin, H. C. Chiu, and C. C. P. Chen. A 1.2v 6.4ghz 181ps 64-bit cd domino adder with dll measurement technique. In *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, pages 1423–1426, May 2011.

[103] Y. S. Wang, M. H. Hsieh, J. C. M. Li, and C. C. P. Chen. An at-speed test technique for high-speed high-order adder by a 6.4-ghz 64-bit domino adder example. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 59(8):1644–1655, Aug 2012.

[104] R. Zlatanovici, S. Kao, and B. Nikolic. Energy delay optimization of 64-bit carry-lookahead adders with a 240 ps 90 nm cmos design example. *IEEE Journal of Solid-State Circuits*, 44(2):569–583, Feb 2009.

[105] J. S. Shah, S. M. Jahinuzzman, D. Li, P. Chuang, and M. Sachdev. A 64-bit, 2.4 ghz adder with se detection capabilities employing time redundancy. In *2009 2nd Microsystems and Nanoelectronics Research Conference*, pages 37–40, Oct 2009.

[106] S. Mathew, M. Anders, R. Krishnamurthy, and S. Borkar. A 6.5ghz 54mw 64-bit parity-checking adder for 65nm fault-tolerant microprocessor execution cores. In *2007 IEEE Symposium on VLSI Circuits*, pages 46–47, June 2007.

[107] S. K. Mathew, M. A. Anders, B. Bloechel, Trang Nguyen, R. K. Krishnamurthy, and S. Borkar. A 4-ghz 300-mw 64-bit integer execution alu with dual supply voltages in 90-nm cmos. *IEEE Journal of Solid-State Circuits*, 40(1):44–51, Jan 2005.

[108] Zhanpeng Jin, Xubang Shen, and Yongqiang Bai. A 64-bit fast adder with 0.18 um cmos technology. In *IEEE International Symposium on Communications and Information Technology, 2005. ISCIT 2005.*, volume 2, pages 1207–1211, Oct 2005.

[109] 2013. URL https://lappc4.epfl.ch/wiki/Main_Page. FPGA4U web-page.

[110] Cyclone ii fpga starter development board reference manual, October 2006.

[111] Usbee rx user's manual, version 1.5, 2005.

[112] Ting-Yen Chiang, K. Banerjee, and K. C. Saraswat. Effect of via separation and low-k dielectric materials on the thermal characteristics of cu interconnects. In *Electron Devices Meeting, 2000. IEDM '00. Technical Digest. International*, pages 261–264, Dec 2000.

[113] Zeng Wang, Gang Dong, YinTang Yang, and JianWei Li. Effect of dummy vias on interconnect temperature variation. *Chinese Science Bulletin*, 56(21):2286–2290, 2011.

[114] A. Van der Ziel. *Noise; sources, characterization, measurement*. Prentice-Hall information and system sciences series. Prentice-Hall, 1970.

[115] P. Klein. An analytical thermal noise model of deep submicron mosfet's. *IEEE Electron Device Letters*, 20(8):399–401, Aug 1999.

[116] A. J. Scholten, H. J. Tromp, L. F. Tiemeijer, R. Van Langevelde, R. J. Havens, P. W. H. De Vreede, R. F. M. Roes, P. H. Woerlee, A. H. Montree, and D. B. M. Klaassen. Accurate thermal noise model for deep-submicron cmos. In *International Electron Devices Meeting 1999. Technical Digest (Cat. No.99CH36318)*, pages 155–158, Dec 1999.

[117] G. Knoblinger, P. Klein, and M. Tiebout. A new model for thermal channel noise of deep submicron mosfets and its application in rf-cmos design. In *2000 Symposium on VLSI Circuits. Digest of Technical Papers (Cat. No.00CH37103)*, pages 150–153, June 2000.

[118] R. P. Jindal. Hot-electron effects on channel thermal noise in fine-line nmos field-effect transistors. *IEEE Transactions on Electron Devices*, 33(9):1395–1397, Sep 1986.

[119] A. A. Abidi. High-frequency noise measurements on fet's with small dimensions. *IEEE Transactions on Electron Devices*, 33(11):1801–1805, Nov 1986.

[120] A. Van Der Ziel. *Noise in Solid State Devices and Circuits*. Wiley, 1986.

[121] A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. A. Zegers van Duijnhoven, and V. C. Venezia. Noise modeling for rf cmos circuit simulation. *IEEE Transactions on Electron Devices*, 50(3):618–632, March 2003.

[122] M. Malits, A. Svetlitza, E. Manzhosov, N. Rotman, and Y. Nemirovsky. The influence of thermoelectric effects on the self-heating of nanometer cmos-soi devices. In *2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel*, pages 1–5, Nov 2012.

[123] A. Ruangphanit, A. Poyai, R. Muanghlua, S. Niemcharoen, and W. Titiroongruang. Development a novel model of threshold voltage of nmos with temperature dependence and

narrow channel width. In *2016 13th International Conference on Electrical Engineering/-Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–5, June 2016.

[124] V. H. Nguyen, Y. M. Niquet, F. Triozon, I. Duchemin, O. Nier, and D. Rideau. Quantum modeling of the carrier mobility in fdsoi devices. *IEEE Transactions on Electron Devices*, 61(9):3096–3102, Sept 2014.

[125] A. L. McWhorter. *Semiconductor surface physics*, chapter 1 /f noise and germanium surface properties, page 207. Series in physics. University of Pennsylvania Press, 1957.

[126] F.N. Hooge. $1/f$ noise is no surface effect. *Physics Letters A*, 29(3):139 – 140, 1969.

[127] S. T. Hsu, D. J. Fitzgerald, and A. S. Grove. Surface state related l/f noise in mos transistors. *Applied Physics Letters*, 12(9):287–289, 1968.

[128] F. Berz. Theory of low frequency noise in si most's. *Solid-State Electronics*, 13(5):631 – 647, 1970.

[129] Gerard Ghibaudo. On the theory of carrier number fluctuations in mos devices. *Solid-State Electronics*, 32(7):563 – 565, 1989.

[130] S.M. Sze and K.K. Ng. *Physics of Semiconductor Devices*. Wiley, 2006.

[131] R. Jayaraman and C. G. Sodini. A 1/f noise technique to extract the oxide trap density near the conduction band edge of silicon. *IEEE Transactions on Electron Devices*, 36(9): 1773–1782, Sep 1989.

[132] K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng. A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors. *IEEE Transactions on Electron Devices*, 37 (3):654–665, Mar 1990.

[133] J. H. Scofield, N. Borland, and D. M. Fleetwood. Reconciliation of different gate-voltage dependencies of 1/f noise in n-mos and p-mos transistors. *IEEE Transactions on Electron Devices*, 41(11):1946–1952, Nov 1994.

[134] G. Reimbold. Modified 1/f trapping noise theory and experiments in mos transistors biased from weak to strong inversion-influence of interface states. *IEEE Transactions on Electron Devices*, 31(9):1190–1198, Sept 1984.

[135] M. J. Deen and Y. Zhu. 1/f noise in n-channel mosfets at high temperatures. *AIP Conference Proceedings*, 282(1):165–188, 1992.

[136] Jimmin Chang, A. A. Abidi, and C. R. Viswanathan. Flicker noise in cmos transistors from subthreshold to strong inversion at various temperatures. *IEEE Transactions on Electron Devices*, 41(11):1965–1971, Nov 1994.

[137] Swastik Kar and A K Raychaudhuri. Temperature and frequency dependence of flicker noise in degenerately doped si single crystals. *Journal of Physics D: Applied Physics*, 34 (21):3197, 2001.

[138] R. R. Troutman. Vlsi limitations from drain-induced barrier lowering. *IEEE Journal of Solid-State Circuits*, 14(2):383–391, Apr 1979.

[139] Behzad Razavi. *RF Microelectronics (2Nd Edition) (Prentice Hall Communications Engineering and Emerging Technologies Series)*. Prentice Hall Press, Upper Saddle River, NJ, USA, 2nd edition, 2011.

[140] W. B. Kuhn and N. M. Ibrahim. Analysis of current crowding effects in multiturn spiral inductors. *IEEE Transactions on Microwave Theory and Techniques*, 49(1):31–38, Jan 2001.

[141] Mutlu Avci and Serhan Yamacli. An improved elmore delay model for vlsi interconnects. *Mathematical and Computer Modelling*, 51(7):908 – 914, 2010. 2008 International Workshop on Scientific Computing in Electronics Engineering (WSCEE 2008).

[142] C. P. Yue and S. S. Wong. On-chip spiral inductors with patterned ground shields for si-based rf ics. *IEEE Journal of Solid-State Circuits*, 33(5):743–752, May 1998.

[143] Rohde&schwarz operating manual spectrum analyser r&s fsp13, May 2003.

List of Publications

Books

- Stéphane Badel, **Can Baltacı**, Alessandro Cevrero, Yusuf Leblebici Design Automation for Differential MOS Current-Mode Logic Circuits Springer, New York, NY, USA, 2019.

Journals

- **C. Baltacı** Yusuf Leblebici, "Thermal aware design and comparative analysis of a high performance 64-bit adder in FD-SOI and bulk CMOS technologies," Integration, the VLSI Journal, Volume 58, 2017, pp. 421-429.

Conferences

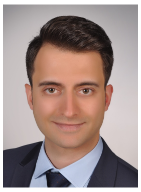
- **C. Baltacı** and Y. Leblebici, "Self-heating effects on the thermal noise of deep sub-micron FD-SOI MOSFETs," 2017 13th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), Giardini Naxos, 2017, pp. 229-232.
- **C. Baltacı** and Y. Leblebici, "Thermal issues in deep sub-micron FDSOI circuits," 2016 12th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), Lisbon, 2016, pp. 1-4.

Written

- **C. Baltacı**, K. Pradeep, F. Kaplan, A. K. Coskun, T. Toifl and Y. Leblebici, "Comparison of the Self-Heating Behaviours of Bulk and FDSOI Technologies".
- **C. Baltacı** and Y. Leblebici, "High-Q Integrated Inductor Design in 28 nm FDSOI".

Can BALTACI

Chemin des Bruyères 5, 1007 Lausanne
canbaltaci@gmail.com
+41 78 669 32 21



STRENGTHS

- Analog Integrated Circuit Design
- Semi-Custom Design Flow (RTL, Synthesis, P&R)
- Self-Heating Effects in Advanced Technologies
- Low Power Design with Moderate- and Sub-threshold
- Transistor Level High-Speed Digital Design
- Hands-on Lab Measurements and Design Verification

PROFESSIONAL EXPERIENCE

Nov 13 – May 18 **MICROELECTRONIC SYSTEMS LABORATORY** *Analog Design Engineer* Lausanne, SWITZERLAND

- Designed and implemented different LNA topologies in 28 nm, proposed techniques for minimizing noise figure considering individual temperatures of each device and the resulting degradation in their noise performance
- Designed and tested high-speed adders using different techniques (parallel-prefix; dynamic and static logic) in 28 nm and 40 nm, proposed circuit level solutions to increase their reliability and lifetime against self-heating
- Designed complete test systems for observing the thermal and flicker (1/f) noise of specific devices in 28 nm
- Designed integrated inductors (stacked, symmetric), T-coils, capacitors, varactors; created their pCells in Skill
- Designed PCBs for tests of various chips; coordinated manufacture, dicing, bonding with external stakeholders
- Performed hands-on measurements and verification of digital ASICs via FPGA (VHDL, C) and characterization of analog RF circuits (noise, gain, matching) with spectrum/network/NF analyzer and other measurement tools
- Investigated self-heating effects of nanometer scale MOSFETs, analyzed the performance degradation due to self-heating in 28 nm and 40 nm FD-SOI and bulk by performing device/circuit level electro-thermal simulations
- Involved in the semi-custom design automation project for differential logic, managed three master students

Jul 14 – Aug 14 **SAMSUNG ELECTRONICS** *Visiting Mixed-Signal Design Engineer* Hwaseong, SOUTH KOREA

- Proposed and implemented novel voltage level shifters (one solution proposed to be patented by Samsung)

Aug 12 – Aug 13 **SWATCH GROUP – EM MICROELECTRONIC** *Hardware Design Engineer* Marin, SWITZERLAND

- Designed three sub-threshold CML standard cell libraries, their bias generators and CMOS/CML-CML/CMOS voltage level shifters in 90 nm and 180 nm technologies for the applications where low switching noise is crucial
- Implemented (full-custom) and tested three ASICs for cryptographic applications covering DES, Grain-128a
- Investigated differential logic against side channel attacks (DPA, electromagnetic etc.), proposed solutions

Jun 10 – Jul 11 **MIKRO-TASARIM (MICRO-DESIGN)** *Hardware Design Engineer* Ankara, TURKEY

- Involved in the tape-out of the second ASIC of the company which was a 576×7 TDI read-out chip
- Created behavioural models of analog and digital blocks, performed mixed signal simulations
- Designed LVDS transmitter and receiver front-ends for 1 Gbps chip to chip in 180 nm for camera links

EDUCATION

Nov 13 – **EPFL** *Ph.D.* Lausanne, SWITZERLAND
Advisor: Prof. Yusuf Leblebici, co-advisor: Thomas Toifl (IBM) (GPA: 6 / 6)
Thesis Title: *Self-Heating Aware Design of Integrated Circuits in Deep Sub-Micron FDSOI and Bulk Technologies*

Sep 11 – Sep 13 **EPFL** *M.Sc.* Lausanne, SWITZERLAND
Advisor: Prof. Yusuf Leblebici (GPA: 5.56/6)

Sep 05 – Jun 10 **METU** *B.Sc.* Ankara, TURKEY
Specialization on Telecommunications (GPA: 3.89 / 4, Ranked: 7th / 200+)

HONOURS AND AWARDS

Jun 17	**PRIME Silver Leaf Award** (10%-20% of best papers in PRIME Conference)
Jun 16	**PRIME Gold Leaf Award** (10% of best papers in PRIME Conference)
Jul 11	**International M.Sc. Scholarship** given by Scientific and Technological Research Council of Turkey
Jul 10	**National M.Sc. Scholarship** given by Scientific and Technological Research Council of Turkey
Jun 10	**Best Senior Design Project** given to a team of five students out of 200+ (led and managed four B.Sc. team-mates)
Mar 10, Oct 07	**Dean's List** given to highest grade each semester

ADDITIONAL INFORMATION

Languages	Turkish (Native), English (fluent), French (Intermediate), German and Spanish (Elementary, currently learning),
Computer Skills	Cadence Virtuoso, VHDL, ModelSim, Synopsys, Encounter, Skill, Ocean ADS, HotSpot, TCAD, Altium, MATLAB
Free Time	Snowboarding, kiteboarding, BodyPump/Combat/Attack, GRIT, literature, foreign languages, guitar, cooking
Personal Info	30, single, Turkish citizenship