

GALS methodology for substrate noise reduction in mixed-signal integrated circuits

Von der Fakultät MINT- Mathematik, Informatik, Physik, Elektro- und Informationstechnik
der Brandenburgischen Technischen Universität Cottbus–Senftenberg
genehmigte Dissertation zur Erlangung des akademischen Grades eines Dr.-Ing.

vorgelegt von

Milan Babić

geboren am 20.11.1987 in Trebinje, Bosnien-Herzegowina

Vorsitzender: Prof. Dr. rer. nat. Peter Langendörfer
Gutachter: Prof. Dr.-Ing. Rolf Kraemer
Gutachter: Prof. Dr.-Ing. Heinrich Theodor Vierhaus
Gutachter: Prof. Dr. Miloš Krstić

Tag der mündlichen Prüfung: 06.10.2021

Abstract

Mixed signal integrated circuits (MSICs) contain both analog and digital circuitry, integrated together on the same die. While this integration brings many benefits, it also gives rise to the issue of substrate noise coupling between the noisy digital circuitry and noise-sensitive analog circuitry. In order to counteract this issue, various substrate noise reduction methodologies had been developed so far.

This thesis explores a new approach in substrate noise reduction – using a GALS (globally-synchronous, locally-asynchronous) design strategy for the digital part of a MSIC, in order to reduce the noise generation at its source. GALS architectures consist of several locally synchronous modules (LSMs) which communicate asynchronously to each other. By converting an initially synchronous architecture of digital circuitry into a GALS architecture, simultaneous switching noise generated by this circuitry can be reduced. While GALS had already been used for reducing other types of simultaneous switching noise, this is, to the best of the author’s knowledge, the first attempt to develop a GALS-based methodology for substrate noise reduction.

In order to be able to theoretically analyze GALS-based methodologies for substrate noise suppression, corresponding models at high abstraction level for substrate noise generation and substrate noise propagation in lightly doped substrates (which is a type of substrate mostly used for MSICs) have been developed. These models have further been used for developing two new GALS-based substrate noise reduction methodologies: harmonic-balanced plesiochronous GALS partitioning (HB) and harmonic-and-area-based plesiochronous GALS partitioning with power domain separation (HAB).

A theoretical analysis has shown that HB can reach substrate noise attenuation of up to $20\log(M)$, where M is the number of LSMs of the resulting GALS system. On the other hand, the attenuation achievable by HAB depends on the distribution of switching current harmonics and area among the partitions, as well as from the substrate itself. For each of the two methodologies, a suitable partitioning procedure for a practical application has been developed; these partitioning procedures have been numerically evaluated in MATLAB. HB has further been embedded within the EMIAS CAD tool, where it has been evaluated on a real design example – a wireless sensor node. A special case of HAB for low frequencies has been applied for developing a test chip called SGE (*power domain Separation and Galsification Experiment*). The measurements on silicon have proved the applicability of the methodology.

While both methodologies presented in this thesis have their limitations, they also show some significant advantages compared to the existing substrate noise reduction methodologies. The author hopes that this work will lead to further research, improvement and finally wider acceptance of GALS-based methodologies for substrate noise reduction.

Zusammenfassung

Integrierte Schaltkreise mit gemischten Signalen (MSICs) enthalten sowohl analoge als auch digitale Schaltungen, die zusammen auf demselben Chip integriert sind. Während diese Integration viele Vorteile gewährt, führt sie andererseits zur Kopplung des Substratrauschens zwischen den rauschgenerierenden digitalen Schaltungen und rauschempfindlichen analogen Schaltungen. Um diesem Problem entgegenzuwirken, wurden bisher verschiedene Methoden zur Reduzierung des Substratrauschens entwickelt.

In dieser Dissertation wird ein neuer Ansatz zur Reduzierung des Substratrauschens untersucht – Verwendung der GALS (global-asynchron, lokal-synchron) Entwicklungsstrategie für den digitalen Teil des MSICs, um die Erzeugung des Rauschens an seiner Quelle zu reduzieren. GALS Architekturen bestehen aus mehreren lokal-synchronen Modulen (LSMs), die miteinander asynchron kommunizieren. Durch Umwandlung der synchronen Architektur eines digitalen Schaltkreises in eine GALS Architektur kann das vom Schaltkreis erzeugte simultane Schaltungsrauschen signifikant reduziert werden. Während GALS schon zur Reduzierung von anderen Arten des simultanen Schaltrauschens eingesetzt wurde, ist das, nach bestem Wissen des Autors, der erste Versuch, eine GALS-basierte Methodik zur Reduzierung des Substratrauschens zu entwickeln.

Um GALS-basierte Methoden zur Unterdrückung des Substratrauschens theoretisch analysieren zu können, wurden die entsprechenden Modellen zur Erzeugung des Substratrauschens und zur Übertragung des Substratrauschens durch schwach-dotierte Substrate (die meistgenutzte Art der Substrate für MSICs) auf hohem Abstraktionsniveau entwickelt. Diese Modellen wurden zur Entwicklung der zwei neuen GALS-basierten Methoden zur Reduzierung des Substratrauschens benutzt: „Harmonic-balanced plesiochronous GALS partitioning“ (HB) und „Harmonic-and-area-based plesiochronous GALS partitioning with power domain separation“ (HAB).

Eine theoretische Analyse hat gezeigt, dass durch HB eine Dämpfung des Substratrauschens bis $20\log(M)$ erreicht werden kann, wobei M die Anzahl der LSMs des resultierenden GALS Systems ist. Auf der anderen Seite, hängt die durch HAB erreichbare Dämpfung von der Verteilung der harmonischen Schwingungen des Schaltstroms und der Fläche unter den LSMs, sowie von dem Substrat ab. Zur praktischen Anwendung jeder dieser zwei Methoden wurde ein entsprechendes Verteilungsverfahren entwickelt; diese Verteilungsverfahren wurden dann in MATLAB numerisch evaluiert. HB wurde in den EMIAS CAD Werkzeug eingebunden und so auf einem Beispiel des realen Entwurfs – einem drahtlosen Sensorknoten – evaluiert. Ein Sonderfall des HABs für tiefe Frequenzen wurde angewendet, um ein Testchip namens SGE (*power domain Separation and Galsification Experiment*) zu entwickeln. Die Anwendbarkeit der Methode wurde durch die Messungen auf dem Silizium nachgewiesen.

Während die beiden Methoden, die in dieser Dissertation vorgestellt wurden, ihre Beschränkungen haben, weisen sie auch manche wesentliche Vorteile gegenüber den bestehenden Methoden zur Unterdrückung des Substratrauschens auf. Der Autor hofft, dass diese Arbeit zur weiteren Forschung, Verbesserung und schließlich Verbreitung der GALS-basierten Methoden zur Unterdrückung des Substratrauschens beitragen wird.

List of Abbreviations

ADC – Analog to Digital Converter
ASIC – Application Specific Integrated Circuit
CAD – Computer Aided Design
CBL – Current Balanced Logic
CDC – Clock Domain Crossing
CM – Clock Modulation
CMOS – Complementary Metal Oxide Semiconductor
CSL – Current Steering Logic
DAC – Digital to Analog Converter
DFF – D-type Flip-Flop
DNW – Deep NWell
DTI – Deep Trench Isolation
ECL – Emitter Coupled Logic
EDA – Electronic Design Automation
EMI – ElectroMagnetic Interference
FIFO – First In First Out
FF – Flip-Flop
FFT – Fast Fourier Transformation
FM – Frequency Modulation
FPGA – Field Programmable Gate Array
FSCL – Folded Source Coupled Logic
GALS – Globally Asynchronous, Locally Synchronous
GPSOI – Ground-Plane Silicon On Insulator
GR – Guard Ring
HB-pGp – Harmonic Ballanced Plesiochronous GALS Partitioning
LNA – Low Noise Amplifier
LSM – Locally Synchronous Module
MSIC – Mixed-Signal Integrated Circuit
PB-pGp – Power Ballanced Plesiochronous GALS Partitioning
PD – Power Domain
PDN – Power Delivery Network
PDS – Power Domain Separation
PRBS – Pseudo-Random Binary Sequencer
PRNG – Pseudo-Random Number Generator
PWL – PieceWise Linear
SCL – Source Coupled Logic
SCS – Switching Current Shaping
SNDR – Signal to Noise and Distortion Ratio
SNR – Signal to Noise Ratio
SOI – Sillicon On Insulator
SPA – Spectral Peak Attenuation
SSN – Simultaneous Switching Noise
TSV – Through Silicon Via

RBW – Resolution BandWidth

RMS – Root Mean Square

VCO – Voltage Controlled Oscillator

Table of Contents

- 1. Introduction..... 8
 - 1.1. Motivation, main objectives and structure of the thesis 8
 - 1.2. Publications list..... 10
- 2. Fundamentals of substrate noise coupling and related work..... 12
 - 2.1. Simultaneous switching noise in mixed signal integrated circuits and substrate noise as its form 12
 - 2.2. An Overview of Existing Models for the Simultaneous Switching Noise (SSN) Analysis 15
 - 2.2.1. Modeling of Switching Current..... 16
 - 2.2.2. Modeling of Supply and Ground Bounce 17
 - 2.2.3. Modeling of Substrate Noise Coupling..... 19
 - 2.3. An Overview of Existing Substrate Noise Reduction Methodologies..... 23
 - 2.3.1. Physical- and circuit-level methodologies for substrate noise reduction 24
 - 2.3.2. System-level methodologies for substrate noise reduction 42
 - 2.4. Summary..... 56
- 3. The modeling approach for theoretical analysis of system level methodologies for substrate noise suppression..... 58
 - 3.1. A coarse model for simultaneous switching noise propagation in lightly doped substrates 59
 - 3.2. A frequency domain model for ground bounce in synchronous and GALS systems 67
 - 3.3. An analysis of applicability of resistive approximation in substrate noise propagation model 72
 - 3.4. Summary..... 77
- 4. Harmonic-balanced plesiochronous GALS partitioning methodology for substrate noise reduction 79
 - 4.1. General requirements for a substrate noise reduction methodology in MSICs 80
 - 4.2. The theoretical foundations of a GALS-based methodology for substrate noise reduction 83
 - 4.2.1. The choice of a GALS clocking scheme 83
 - 4.2.2. Mathematical analysis of the optimal plesiochronous GALS partitioning for substrate noise reduction..... 85
 - 4.2.3. Avoiding the harmonic overlapping 90
 - 4.2.4. The impact of the plesiochronous approximation 91
 - 4.3. An algorithm for harmonic-balanced GALS partitioning, and numerical evaluation of the methodology in MATLAB..... 95
 - 4.3.1. An algorithm for harmonic-balanced GALS partitioning 95

4.3.2.	Numerical evaluation of the methodology in MATLAB.....	99
4.3.3.	Substrate noise spectral power attenuation for frequency bands containing more harmonics.....	103
4.4.	A design flow for the GALS-based substrate noise reduction methodology	107
4.4.1.	Incorporation of the methodology into a GALS design flow, and an application on a real design example.....	107
4.4.2.	The impact of granularity	110
4.5.	Comparison of harmonic-balanced plesiochronous GALS partitioning to the other system-level methodologies for substrate noise reduction	111
4.5.1.	Main limitations of harmonic-balanced plesiochronous GALS partitioning	111
4.5.2.	Comparison to power-balanced plesiochronous GALS partitioning	112
4.5.3.	Comparison to other system level methodologies for substrate noise reduction	115
4.6.	Summary.....	117
5.	Substrate noise reduction methodology based on a combination of GALS partitioning and physical placement.....	121
5.1.	Introduction.....	121
5.2.	Substrate noise reduction by power domain separation in the digital aggressor	122
5.3.	Additional improvement by applying GALS design approach	131
5.4.	The impact of decoupling capacitance.....	134
5.5.	An algorithm for substrate noise reduction by GALS partitioning and power domain separation, and a numerical evaluation of the methodology in MATLAB	142
5.6.	Some special cases of the methodology	148
5.6.1.	Targeting low order substrate noise harmonics	148
5.6.2.	Power approximately uniformly distributed per area.....	153
5.7.	A comparison to other physical and system level substrate noise reduction methodologies	155
5.7.1.	Main limitations of the substrate noise reduction methodology based on power domain separation and galsification	155
5.7.2.	Comparison to other physical-level methodologies for substrate noise reduction....	156
5.7.3.	Comparison to other system-level methodologies for substrate noise reduction	157
5.8.	Summary.....	158
6.	Case study: <i>SGE</i> chip.....	160
6.1.	Architecture of the <i>SGE</i> chip and GALS partitioning strategy applied	160
6.1.1	Architecture of the reference design	160
6.1.2	GALS partitioning goal, the applied partitioning methodology and the architecture of the resulting GALS design.....	162

6.1.3	Floorplan, layout and packaging of the test chip	167
6.2.	Testing and measurement results	168
6.3.	Summary.....	173
7.	Conclusion	175
	List of References	177

1. Introduction

1.1. Motivation, main objectives and structure of the thesis

Mixed-signal systems, containing both analog and digital circuitry, are common in modern electronics. While signal processing is predominantly digital today, the modules providing an interface of an electronic system to the outside world, which is analog by its nature, generally stay analog [1]. Integrating both analog and digital modules into a single mixed-signal integrated circuit (MSIC) leads to minimization of system dimensions, reduction of power consumption and (especially for the systems produced in high volume) it may provide vast cost savings.

However, integration of both analog and digital circuitry on the same die also leads to the problem of noise coupling. Simultaneous switching of a large number of gates in the digital circuitry causes the so called simultaneous switching noise (SSN). The SSN has three main forms:

- Electromagnetic interference (EMI) due to switching currents
- Supply and ground noise (also called ground bounce)
- Substrate noise

Substrate noise is a general term for the electrical disturbances which propagate from the noise generating circuitry (also called „aggressor“) to the noise-sensitive circuitry (also called „victim“) through the common die substrate. Digital circuitry is usually not much sensitive to the substrate noise. On the other hand, when substrate noise reaches analog circuitry, its performance may deteriorate.

Substrate noise within the frequency band of a low noise amplifier (LNA) can couple to its output and thus disturb its output signal. This can happen both for single-ended LNAs [2] and for differential LNAs [3]. In analog-digital converters (ADCs), substrate noise may cause a drop in signal to noise ratio (SNR). [4][5] Substrate noise introduces timing uncertainty into the sampling process of an ADC [6][7], which also manifests itself as an undesired frequency modulation (FM) [8][9]. In voltage controlled oscillators (VCOs), substrate noise can cause a frequency modulation of the oscillator frequency [10][11] and a degradation of phase noise [12][13]. It also introduces spurious side-band tones into the output spectrum of a VCO. [14] In the extreme case of a large digital aggressor, even an injection locking of a VCO to a wrong frequency may occur: instead of locking to the resonant frequency of its LC tank, the VCO may lock to the substrate noise at a higher order clock harmonic. [15]

In order to prevent the detrimental impact of substrate noise to the operation of analog circuits, plenty of methodologies for substrate noise suppression and reduction have been developed. However, none of these methodologies is perfect and there is still no universal solution for suppressing the substrate noise. One subcategory of substrate noise reduction methodologies are system level methodologies, which are based on distributing the switching in digital aggressor circuitry in time. System level methodologies can also be used for combatting the other two forms of switching noise as well. Moreover, they were usually initially developed to target other forms of switching noise, before being optimized to specifically target the substrate noise.

One of the ways to distribute switching in digital circuitry in time is by applying a GALS (globally asynchronous, locally synchronous) design concept, instead of a standard synchronous digital design concept. A GALS system consists of several locally synchronous modules (LSMs), each of them representing a separate clock domain. LSM clocks are generated independently from each other, which means that they are naturally desynchronized. In order for desynchronized LSMs to communicate between each other, they are connected by an asynchronous interface. A simplified block diagram of a GALS system consisting of two LSMs is shown in Fig. 1.1.

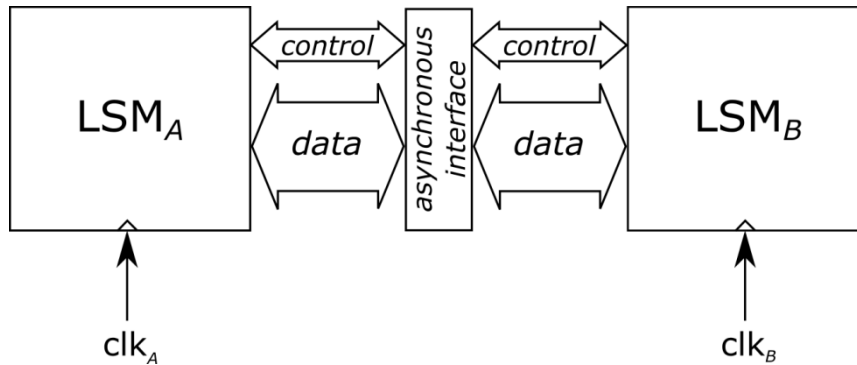


Figure 1.1 – A simplified block diagram of a GALS system with two LSMs

GALS-based switching noise reduction methodologies have already been developed for noise reduction of other switching noise forms – EMI radiation originating from switching current [16], and ground bounce [17]. These noise reduction methodologies are based on converting an initially synchronous system into an equivalent GALS system, i.e. on „galsification“ of an initially synchronous system. In [16], the galsification procedure was optimized for spectral peak reduction of the fundamental harmonic of the switching current (relevant for the reduction of EMI radiation originating from switching current), while in [17] the procedure was optimized for the time-domain ground bounce peak reduction. However, to the best of the author’s knowledge, the application of GALS as a means for substrate noise reduction has not been examined yet, and no GALS-based methods optimized for substrate noise reduction exist so far.

The main goal of this work is to explore the possibility of using GALS design approach for reduction of substrate noise, and to develop new GALS-based methodologies for substrate noise reduction. Since lightly doped substrates are dominantly used for MSICs, the methodology should specifically provide substrate noise reduction in such substrates, and possibly use the advantages which this substrate type provides.

In the next Chapter, **Chapter 2**, some fundamental theory with an overview of related work is provided. First, the basic principles of substrate noise coupling are explained, including noise generation by a digital aggressor, propagation through the substrate, and reception by the analog victim. After that, an overview of the existing approaches for modeling the three forms of switching noise (switching current, ground bounce and substrate noise) is presented. Further, a detailed survey of existing substrate noise reduction methodologies is presented, with special attention being given to the system level methodologies. Finally, the principles of GALS design approach, and the existing methodologies for reducing other SSN types by using GALS are presented.

In order to be able to develop a GALS based methodology for substrate noise reduction, appropriate high abstraction level models, enabling the theoretical analysis of the impact of system

level configurations to the substrate noise, are needed. Thus, as an additional, secondary task of this work, a coarse, high abstraction level model for substrate noise propagation in lightly doped substrates is developed. Noise generation by a GALS system is also modeled. **Chapter 3** deals with these modeling tasks. Models developed within this chapter are further used for the analysis of GALS-based noise reduction methodologies represented in the following chapters.

Once the appropriate analysis models are available, the main goal of this work can be pursued. Initially, a general GALS-based substrate noise reduction methodology, applicable to any kind of substrates, is developed, and presented in **Chapter 4**. The general requirements for substrate noise reduction are set, the frequency scheme is chosen, and the corresponding partitioning strategy is theoretically derived. An algorithm which provides the application of the proposed partitioning strategy is developed and numerically evaluated. Further, the algorithm is integrated in the design flow, and applied to a realistic design. The proposed methodology is compared to existing system-level methodologies for substrate noise reduction; its advantages, limitations and possibilities for future improvements are identified and discussed.

As the next step, a methodology which also takes into account the properties of lightly doped substrates is developed. This methodology is presented in **Chapter 5**. It combines power domain separation, floorplanning and galsification in order to reduce the substrate noise. The methodology has been analyzed theoretically and numerically, and an algorithm providing its application for realistic designs has been developed. Some special cases, where the application of this methodology becomes significantly simplified, are also identified and further developed. Finally, the methodology is compared to existing methodologies.

A special case of the methodology presented in Chapter 5 has also been verified in silicon. The test chip which has been fabricated for this purpose and the measurement setup are described in **Chapter 6**. The measurement results are presented and discussed.

Finally, in **Chapter 7**, the conclusions are drawn and the contributions of this work are summarized.

1.2. Publications list

In scope of the research conducted for this thesis, six conference articles have been published with me being the first author, i.e. the main contributor. Here is the list of these publications:

- [P1] Milan Babić, Miloš Krstić, “*A coarse model for estimation of switching noise in lightly doped substrates*”, Proc. of 18th IEEE Symposium on Design and Diagnostics of Electronic Circuits and Systems (DDECS), pp. 217-222, Belgrade, Serbia 2015.
- [P2] Milan Babić, Xin Fan, Miloš Krstić, “*Frequency-domain modeling of ground bounce and substrate noise in lightly doped substrates*”, Proc. of 25th International Workshop on Power and Timing Modeling, Optimization and Simulation (PATMOS), Salvador, Bahia, Brazil 2015.

- [P3] Milan Babić, Miloš Krstić, *“GALS methodology for substrate noise reduction in BiCMOS technologies”*, Proc. of Workshop on Manufacturable and Dependable Multicore Architectures at Nanoscale (MEDIAN), pp. 46-47, Tallinn, Estonia 2015.
- [P4] Milan Babić, Miloš Krstić, *“Modeling of substrate noise coupling and ground bounce for GALS systems”*, Proc. of the Annual DCPS (Dependable Cyber-Physical Systems) Evaluation Workshop, pp. 70-73, Cottbus, Germany 2015.
- [P5] Milan Babić, Steffen Zeidler, Miloš Krstić, *“GALS partitioning methodology for substrate noise reduction in mixed-signal integrated circuits”*, Proc. of 22nd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC), pp. 67-74, Porto Alegre, Brazil 2016.
- [P6] Milan Babić, Miloš Krstić, *“A substrate noise reduction methodology based on power domain separation of GALS subcomponents”*, Proc. of 27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS), Thessaloniki, Greece 2017.

Additionally, one more conference article has been published and one journal article has been accepted for publication, with me being the coauthor. Here is the list of these publications:

- [P7] Miloš Krstić, Xin Fan, Milan Babić, Eckhard Grass, Tobias Bjerregaard and Alex Yakovlev, *“Reducing switching noise effects by advanced clock management,”* 2017 11th International Workshop on the Electromagnetic Compatibility of Integrated Circuits (EMCCompo), pp. 3-8, St. Petersburg, Russia, 2017.
- [P8] Xin Fan, Milan Babić, Shutao Zhang, Eckhard Grass and Miloš Krstić, *“Plesiochronous Spread Spectrum Clocking With Guaranteed QoS for In-Band Switching Noise Reduction”*, IEEE Transactions on Circuits and Systems I: Regular Papers (accepted for publication), 2021.

2. Fundamentals of substrate noise coupling and related work

This chapter provides the theoretical fundamentals of substrate noise coupling, together with an overview of related work regarding substrate noise modeling and reduction strategies.

In Section 1, the phenomenon of simultaneous switching noise in digital circuits is presented, including its main forms, one of them being substrate noise. Further, the basic principles of substrate noise coupling are explained, including noise generation by a digital aggressor, propagation through the substrate, and reception by the analog victim.

In Section 2, an overview of the existing approaches for modeling the three forms of switching noise (switching current, ground bounce and substrate noise) is presented.

In Section 3, a detailed survey of existing substrate noise reduction methodologies is presented, with special attention being given to the system level methodologies. Also, the principles of GALS design approach and the existing GALS-based methodologies for reducing other SSN types are presented.

Finally, in Section 4, the conclusions are drawn.

2.1. Simultaneous switching noise in mixed signal integrated circuits and substrate noise as its form

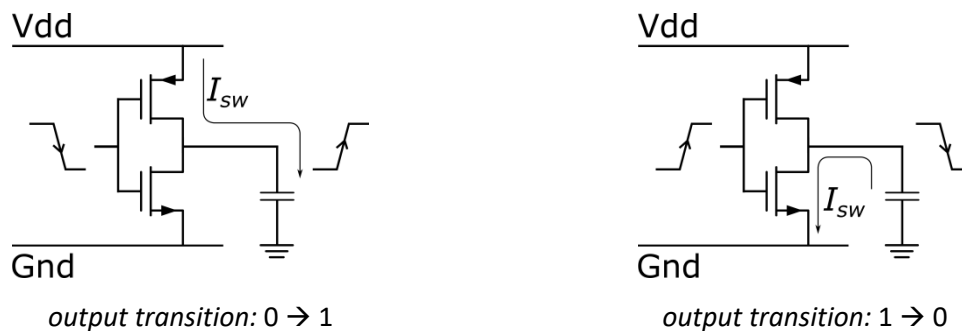


Figure 1.1 – Switching current originating from a single inverter

One of the key advantages of digital electronics compared to analog is its relatively high resistivity to noise. However, while being rather insensitive to noise themselves, digital circuits produce a large amount of noise while operating. When a digital circuit switches, i.e. changes the signal value at its output, its output capacitance gets charged or discharged by a current surge on supply or ground lines leading to this circuit, as shown for an example of an inverter in Fig. 2.1. These currents are called switching currents, and the noise originating from them is called switching noise. For more complex gates, switching currents also appear (although to a lesser extent) when an internal node of a gate is changing its voltage level. While a switching current from a single digital gate is usually rather small, a huge amount of simultaneously switching gates in an integrated digital module may cause strong switching currents. Noise generated by simultaneous switching of a larger number of digital gates is denoted as simultaneous switching noise (SSN) [18]. It is very important to

notice that SSN is a system level phenomenon – it depends on the switching activity of the aggressor, which is a system level parameter.

The first major form of SSN is electromagnetic radiation. It is caused by the flow of switching currents through bond wires, package pins and conducting lines on the PCB. This radiation can lead to various electromagnetic interference (EMI) related issues in the neighboring electrical systems. For example, electromagnetic radiation originating from one bond wire on the chip may induce voltages on another bond wire of the same chip, thus corrupting voltage levels [19]. EMI issues are also possible on chip, where electromagnetic coupling between close lines can happen, especially if these lines have a long common parallel section. Since electromagnetic radiation is directly dependent on the switching current spectrum, the switching current spectrum is used as a quantitative description of EMI. Thus, for the purpose of noise analysis, the switching current itself can be regarded as synonymous to the EMI, i.e. as the first form of SSN.

Moreover, while flowing through the power delivery network (PDN), due to on-chip, package and on-board parasitics, switching currents cause voltage fluctuations on supply and ground lines (as shown in Fig. 1.2). This is the second major form of switching noise – supply and ground bounce, also called supply and ground bounce. This form of switching noise is of importance for purely digital systems as well. Supply and ground bounce may cause delay uncertainty in digital circuits, especially in tapered buffers [20]. As pointed out in [21] and [22], the average supply and ground noise impacts the variation in the delay of a critical path, while the maximum peak of supply and ground bounce narrows the noise margins in digital circuits and can cause false switching to occur. Analog circuits are even much more sensitive to supply and ground bounce. Thus, in order to avoid supply and ground noise coupling from digital to analog circuitry in MSICs, power delivery networks for digital and analog circuitry in such systems are usually separated.

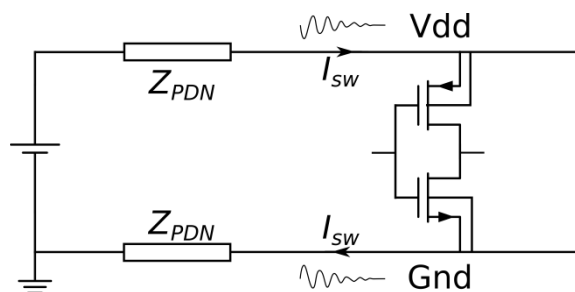


Figure 2.2 – Ground bounce being generated by the flow of the switching current I_{SW} through the power delivery network impedance Z_{PDN}

In digital integrated modules, substrate biasing is usually done by connecting the ground lines (in metal) to the substrate (p-type semiconductor) through a large number of substrate contacts (so called ptaps). Similarly, N Wells are biased from supply lines through the NWell contacts. Through the substrate contacts, ground noise gets injected directly into the substrate, thus generating the third form of switching noise – substrate noise. This type of noise is of special interest in mixed signal integrated systems, where digital noise generating circuits (usually denoted as aggressors) and analog circuits sensitive to noise (usually denoted as victims) share the same substrate. Substrate noise coupling process in MSICs consists of three stages, as shown in Fig. 2.3: substrate noise generation (i.e. injection of noise to the substrate) by a digital aggressor, noise propagation through the substrate, and substrate noise reception by an analog victim.

Direct coupling of ground bounce to the substrate through substrate contacts is not the only source of substrate noise. Voltage transitions from output and internal nodes also get capacitively coupled through the source or drain junctions of the transistors connected to these nodes. In the channel of NMOS transistors, impact ionization occurs, causing extra carriers and injecting them into the substrate as a weak substrate current. Additionally, noise gets injected into N Wells, either via direct coupling through NWell contacts from supply lines, via capacitive coupling through source or drain junctions of PMOS transistors, or via impact ionization in the channels of PMOS transistors. The noise injected into N Wells can further propagate capacitively to the substrate via NWell capacitance. Noise from long and wide metal lines can also couple capacitively to the substrate beneath them. However, for most of the middle to large scale digital systems, and for most of the packages, direct coupling from ground network is the dominant source of substrate noise. [23]

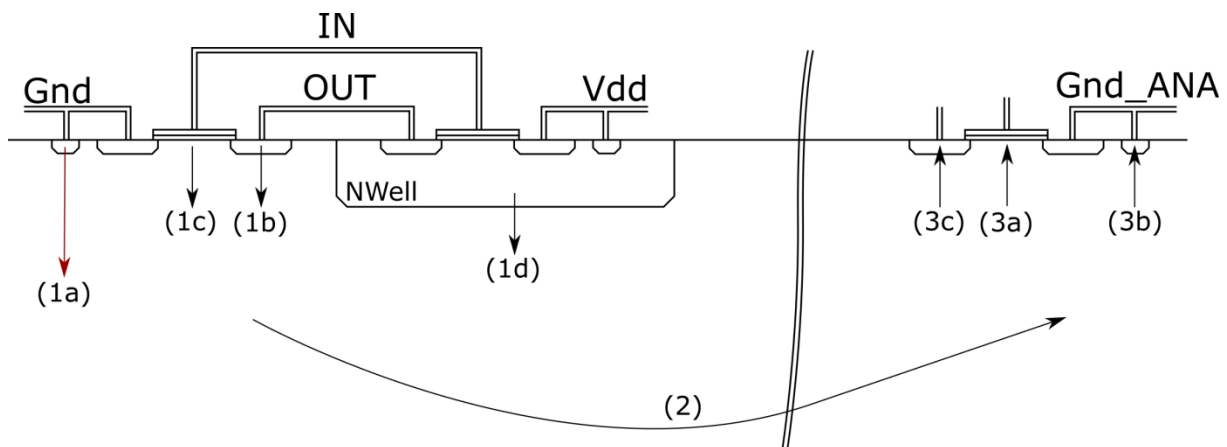


Figure 2.3 – Three stages of substrate coupling in mixed signal systems: (1) substrate noise generation (1a – direct coupling through substrate contacts, usually the dominant substrate noise source, 1b – capacitive coupling from source/drain junctions, 1c – impact ionization from the channel, 1d – capacitive coupling from Nwell), (2) substrate noise propagation, (3) substrate noise reception (3a – body effect, 3b – direct coupling to analog ground, 3c – capacitive coupling to source/drain junctions)

After being injected into the substrate, noise can propagate from digital aggressors to analog victims. Noise propagation is highly dependent on the type of the substrate. There are two main substrate types: epi-type substrates and lightly doped substrates.

Epi-type substrates, also called heavily doped substrates or low resistivity substrates (Fig. 2.4.a), consist of a heavily doped bulk, and a thin lightly doped surface layer on top of it, called epitaxial layer, shortly epi-layer. Epi-type substrates are traditionally mostly used in purely digital integrated circuits. Heavily doped bulk enables homogenous biasing of the substrate, it lowers the risk of latch-ups, and even helps making ground more stable and attenuate ground bounce. [24] However, they're not suitable for mixed signal ICs, the main reason being precisely that substrate noise propagation across the chip is made easy by low bulk resistivity (typically at the order of magnitude of 0.01 Ωcm). [23]

Lightly-doped substrates, also called high-resistivity substrates or bulk-type substrates (Fig. 2.4.b), consist of a usually homogenous lightly doped bulk. Such bulks typically possess a resistivity at the order of magnitude of tens of Ωcm [23][25], although there are even substrates with a resistivity of about 1000 Ωcm [26]. Some newer lightly doped substrates additionally have a thin PWell layer of low resistivity (order of magnitude 0.01 Ωcm) on top of a lightly doped bulk [27][28]. While

propagating through a high resistivity substrate, substrate noise effectively gets attenuated, with attenuation being dependent on the distance between the aggressor and the victim. Thanks to this property, lightly doped substrates can provide isolation by distance, which is why they are the preferred substrate type for usage in MSICs.

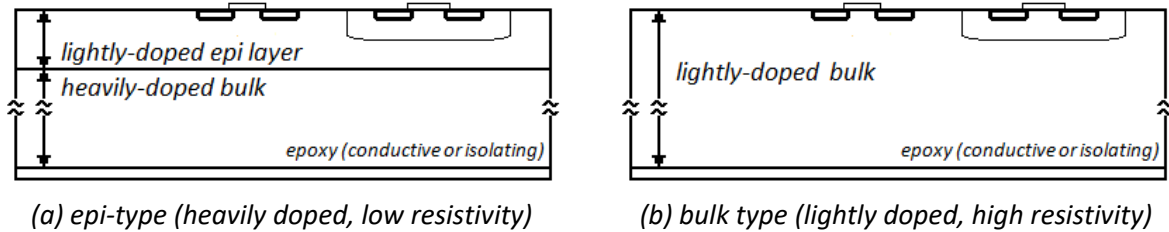


Figure 2.4 – The two substrate types

Finally, after being injected to the substrate by a digital aggressor and propagated through the substrate, substrate noise gets received by the analog circuitry. This noise reception also happens via several mechanisms. Substrate noise modulates threshold voltage of NMOS transistors via the body effect, it can also be capacitively coupled into the analog circuitry through source or drain junction capacitances, and it can get directly injected into the analog ground network through substrate contacts in analog circuitry, thus corrupting the ground level in analog circuitry.

2.2. An Overview of Existing Models for the Simultaneous Switching Noise (SSN) Analysis

In previous section, generation of simultaneous switching noise has been explained, as well as its three forms: switching current, supply/ground bounce, and substrate noise. The dominant source of substrate noise is usually the direct coupling of ground bounce through substrate contacts of a digital aggressor. The ground bounce itself is caused by the flow of switching current through ground lines of the power delivery network of the digital aggressor. Thus, in order to be able to analyze substrate noise originating from simultaneous switching in digital aggressors and the methods for its reduction, appropriate models are necessary for each of the three switching noise forms: switching current, supply/ground bounce, and substrate noise. In this section, an overview of the existing modeling approaches is presented.

Simultaneous switching noise is a system-level phenomenon. Consequently, the models for SSN have to be system level models. However, they differ in the applied level of abstraction, which strongly depends on the intended usage of a model. Models used for system level *verification* are based on very detailed physical noise models at cell level, which are then combined at system level and used in complex, time consuming simulations. On the other hand, models used for system level *analysis* and for assistance in making design decisions are not as precise, and sometimes they even work with completely abstract representations. These models, however, aim to provide just a guideline which will be used in minimizing noise, and they don't need to have a high precision. Moreover, in early phases of a design flow, a lot of data about the system being designed are still unknown, which also limits the achievable precision of these models. In this overview, a special

attention will be paid to this second class of system level models, since this is the kind of models needed for developing a system level methodology for noise reduction.

2.2.1. Modeling of Switching Current

The exact switching current waveform is dependent on several system characteristics. First of all, the types of gates composing the system and the current characteristics of each of those gates for different input combinations determine the possible contributions of each gate to the total switching current. Further, system input signals and the topology of the system determine which of those possible combinations will occur. And finally, due to parasitics in the power delivery network (PDN), switching current causes the oscillations on supply and ground lines, i.e. supply and ground bounce, which also impacts the switching current itself. However, even though there are many parameters which determine the switching current waveform, several simplified representations which can be successfully used in switching current analysis on system level have been developed.

It has been shown [29][30] that for a synchronous system two components of the switching current can be distinguished: a periodic current pulse, which represents the average current over one period, and a noise signal, which represents cycle-to-cycle variations. In frequency domain, the periodic current pulse corresponds to the discrete spectrum harmonic peaks, while the noise signal corresponds to the noise floor.

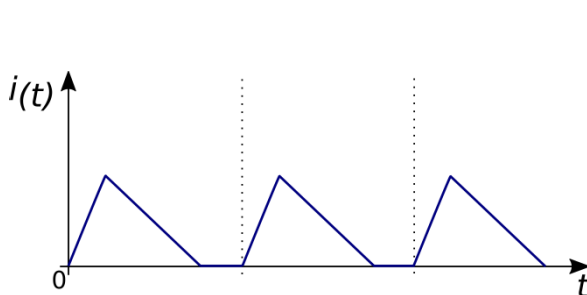


Figure 2.5 – A periodic triangular pulse as a switching current model

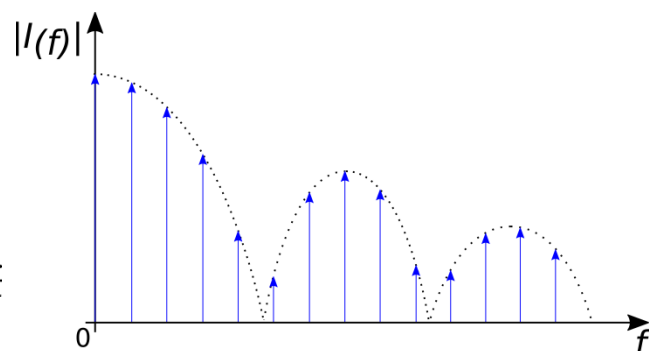


Figure 2.6 – The spectrum of a periodic triangular pulse

In [29], switching current waveform was modeled as a periodic triangular pulse, and this model was analyzed in detail. The time-domain waveform and the spectrum of such a signal are presented in Fig. 2.5 and Fig. 2.6, respectively. The deviations from the average waveforms in every cycle have been taken into account, and the impact of those deviations to the spectrum was evaluated. It was shown that, for the PSD of the current up to the first notch frequency of the spectrum envelope (i.e. for the spectrum peaks which fall inside the first lobe), the impact of the deviations can be neglected. The approach has been verified on a medium-scale circuit (40k gates).

In [30] the behavior of digital circuitry has been analyzed statistically, using Markov chain theory. It was shown that the envelope of spectral harmonic peaks stays high above the noise floor. Additionally, it was shown that harmonic peaks are quadratically proportional to the switching frequency, while the noise floor stays linearly proportional to the switching frequency. In other

words, as the switching frequency becomes higher, the discrete peaks become more dominant. After determining the switching current spectrum from this analysis, a fitting simplified waveform which should match this spectrum was determined. Two waveform models of the average current pulse have been taken into account: an exponential and a triangular model. In the exponential model, a good matching has been obtained for lower frequencies, while the triangular model showed a good matching at higher frequencies as well. In [31], the switching behavior in a combinational network is modeled as a stochastic process, and the periodic triangular pulse model has been used in deriving the expressions for power spectrum density.

In [32], it was proposed to use a triangular pulse if the peak current is more than twice bigger than the average current during one period, and a trapezoidal pulse if the peak current is less than two times the average current. The trapezoidal pulse might be a good model for large scale designs, since in these, the switching is more dense and more distributed during the clock period.

In [16], an approach similar to the one from [29] was applied to analyze GALS architectures as well. The switching current of each LSM composing a GALS system was modeled as a periodic triangular pulse, and the switching current of the entire system was modeled as a sum of switching currents of individual LSMs.

2.2.2. Modeling of Supply and Ground Bounce

Ground bounce models consist of two main parts: a model of switching circuitry, i.e. a switching current model, and a model of power distribution network (PDN). Unlike switching current, which has been often modeled in frequency domain, ground bounce has historically mostly been modeled in time domain. The most important reason is that, as pointed out in [21] and [22], the time domain metrics are important for characterizing the impact of noise on the supply and ground lines to the digital circuitry using those lines. Specifically, average noise in time domain impacts the delay variation, while peak noise in time domain lowers the noise margins in digital logic. Another reason is that most of the techniques for ground bounce reduction are physical design techniques, and such techniques are mostly analyzed in time domain. [22][33]

In [22], power and ground lines have been modeled as a series RL circuit. A capacitance C_d between supply and ground lines has also been introduced into the model, representing the sum of the decoupling capacitance and the internal capacitance of non-switching gates. Switching circuitry has been represented by a current source with periodic triangular pulse waveform.

In [34], switching current was modeled by a voltage-controlled current generator. The current source stretches the current waveform when the voltage drops, taking care that the consumption caused by the FF switching (both on rising and falling edge) is properly timed (while the consumption caused by the combinatorial logic is aligned to the consumption from the FF on rising edge). This way, the feedback from supply and ground oscillations on the switching current is also taken into account. For modeling a PDN, a lumped RLC circuit was used, including the capacitor representing the intentional decoupling capacitance. Additionally, an RC branch in parallel with the generator was included, modeling the capacitance of the switching circuitry. Errors compared to

SPICE are much smaller than in case when the current source is modeled as independent, while the simulation time is a bit faster.

In [17], power network was represented by the model shown in Fig. 2.7.a. Switching circuitry was modeled by inverters, with transistors described by short channel equations. Load capacitances were attached to the outputs of the inverters. This model was further modified to the one presented in Fig. 2.7.b. Here, the current I_g represents the sum of the currents of all the switching inverters, while the capacitor C_d models both decoupling capacitance and the load capacitance of the non-switching inverters. The model was used for comparing GALS and synchronous designs in terms of ground bounce in time domain.

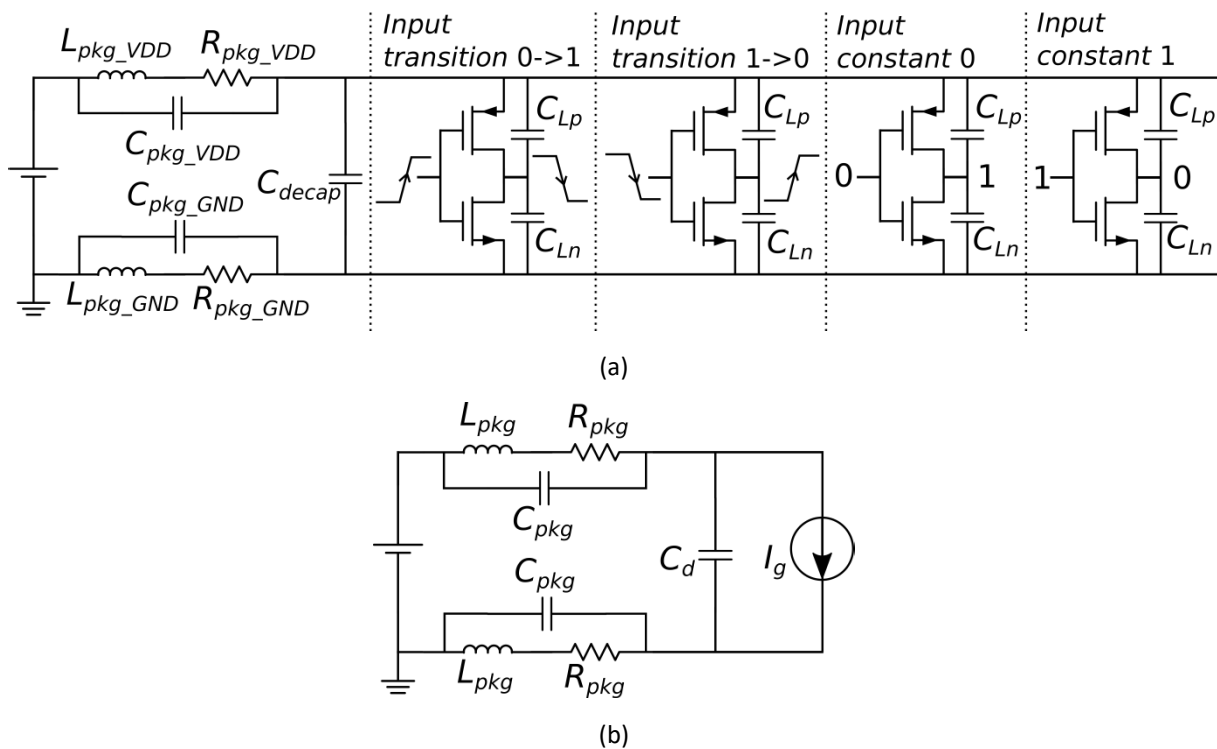


Figure 2.7 – Lumped model of a power delivery network (adapted from [17])

All the aforementioned models analyze ground bounce in time domain. In [29], ground bounce is analyzed in spectral domain. It's pointed out that ground bounce spectrum presents the product of switching current spectrum and power network transfer function. Supply and ground lines have been modeled as series RL circuits. Decoupling capacitor with its parasitic series resistance is also included in the model, connected in parallel with an additional capacitor representing the impedance of non-switching circuitry. Switching circuitry has been modeled by a current source with periodic triangular waveform. It has been shown that, due to resonance in PDN, dominant components of ground bounce spectrum are not at low, but on higher harmonics. Reference [30] also uses a similar model of power network to the one used in [29].

In each of the aforementioned models, the PDN was modeled by a lumped RLC network. This is a good approximation in case that package parasitics are dominant to on-chip parasitics. Otherwise, the on-chip PDN has to be included as well, and it has to be modeled as a distributed RLC network. Examples of such models are given in [35][36].

2.2.3. Modeling of Substrate Noise Coupling

The process of substrate noise coupling can be separated in three stages: substrate noise generation on aggressor's side (i.e. noise injection in the substrate), substrate noise propagation from an aggressor to a victim, and substrate noise reception on victim's side. A complete substrate noise coupling model would contain models for each of those three stages. Noise reception on the side of an analog victim is strongly dependent on the victim itself, and it's usually analyzed and simulated separately, by using an analog simulator such as SPICE. This analysis is performed either by including the calculated values of noise once the level of noise in the substrate at victim's position is known, or by sweeping the substrate noise level in order to determine the highest acceptable substrate noise level for the specific analog block. Thus, substrate noise coupling models generally deal with the first two stages – noise generation and noise propagation, which means that they include modeling the aggressor(s) and modeling the substrate.

The complexity of an *aggressor model* depends on the purpose of substrate coupling modeling, i.e. on whether the model is intended for substrate noise analysis in an early stage of the design cycle, or it's intended for substrate noise verification at the end of the design cycle, when the layout details are known. The complexity of the *substrate model* primarily depends on the type of the substrate – for lightly-doped substrates, the models are much more complicated than for epi-type substrates.

In most of the *substrate models*, some general assumptions excluding the faulty behavior which should not occur during regular operation are taken [23]. It is assumed that there are no irregular states in substrate, such as latch-up or directly-biased well-substrate junctions. Specific semiconductor behavior, such as substrate inversion, is not modeled. It is also assumed that the inductive coupling can be neglected, which is generally correct if the wavelength of magnetic field is much bigger than the dimensions of the chip. Taking these assumptions contributes a lot to the simplification of the analysis [23].

There are two main categories of substrate models – *numerical* and *analytical* models, with numerical models being much more precise.

Obtaining a closed-form analytical solution is possible only for some simple structures, and precise analytical models for complicated geometries do not generally exist [37]. Thus, all the existing *analytical models* are only approximate. Several such methods are described in [18]. However, they are all aimed either at epi-type substrates, or at circuits with very few elements in lightly-doped substrates. All of those models require some technology dependent fitting parameters. Additionally, they are mostly based on neglecting the interaction between the distant contacts, and focusing on coupling between the adjacent contacts instead, which makes them unsuitable for full-chip analysis.

An interesting approximate analytical model for lightly doped substrates is presented in [38], and further elaborated in [39]. The interaction between a single aggressor (which can even be a large digital circuit) and a single victim is modeled by approximating the propagation path of the coupling current with a half-ellipse, while the substrate itself is approximated to be purely resistive. The model achieves quite a high accuracy, but its applicability is limited, since it doesn't provide the capability of modeling systems containing several independent digital aggressors. Models presented in [40] and

[28] use a similar approach, but they also only include a single aggressor and a single victim surrounded by a guard ring.

Numerical models, on the other hand, can further be categorized in two main classes: those based on solving Poisson's equation, and those based on obtaining Green's function [18].

In order to apply methods based on Green's function, such as Boundary Elements Method (BEM), two assumptions have to be satisfied. The first one is that the substrate must be homogenous in horizontal directions. In other words, substrate parameters, such as conductivity or dielectrical permittivity, can change only in vertical direction. In the surface region, for a couple of microns below the substrate, there is no horizontal homogeneity due to the layout details. In full-chip analysis, however, inhomogenities in this small region can be neglected; since the substrate below this region is homogenous in horizontal direction, regardless of the substrate type, the first assumption is generally satisfied. The second assumption, on the other hand, is that the current density through each substrate port must be uniform. Ports are actually connections of the substrate to the outside world. This assumption is not satisfied if whole contacts are considered as ports, because current density is not uniform over a contact. In order to overcome this, contacts need to be separated into several panels with uniform current density, which can then be considered as ports. [41][42][18]

In methods based on solving Poisson's equation, such as Finite Difference Method (FDM), substrate is first discretized into (usually cubic) domains, which are small enough for the electrical field inside a domain to be considered homogenous [18]. Then, for each of the domains, a Poisson's equation can be written, which can further be transformed into the following equation [37][18]:

$$\sum_j \left[\frac{V_i - V_j}{R_{ij}} + C_{ij} \left(\frac{\partial V_i}{\partial t} - \frac{\partial V_j}{\partial t} \right) \right] = 0 \quad (2.1)$$

with $R_{ij} = \rho \cdot dl/dA$, $C_{ij} = \rho \cdot dA/dl$, where ρ is substrate resistivity, ϵ is substrate dielectric constant, dl is the distance between the adjacent nodes (centers of adjacent domains), and dA is the facing area between the adjacent domains.

Equation (2.1) is actually the Kirchoff's law for the circuit at the Fig. 2.8, with R_{ij} as resistive elements, and C_{ij} as capacitive elements [23][18].

This way, the substrate is represented as a large number of nodes connected with a mesh of impedances consisting of resistive and capacitive part (i.e. a large RC-mesh). For each of the branches, the resistive and the capacitive impedance parts are equal at the following frequency [23]:

$$f_T = \frac{1}{2\pi R_{ij} C_{ij}} = \frac{1}{2\pi \rho \epsilon} \quad (2.2)$$

At frequencies much lower than this one, capacitive part can be neglected, and the substrate mesh can be approximated as purely resistive.

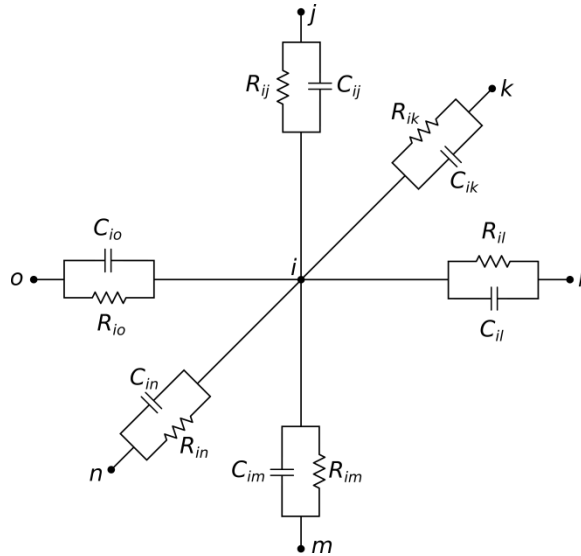


Figure 2.8 – Part of a substrate RC-mesh surrounding one node (adapted from [18])

By writing the Kirchoff's law for all the nodes within the substrate, a huge matrix is obtained. It would require a lot of time to directly solve such a matrix. However, in addition to the symmetry (which is a property of every matrix describing some linear electrical circuit), this matrix is also sparse, since the nonzero elements exists only at the positions corresponding to branches between adjacent nodes. Thanks to the symmetry of the matrix and its sparsity, some faster iterative techniques can be applied [18].

Substrate discretizations steps in FDM don't have to be equal in all substrate regions. In the regions where the magnitude and direction of electrical field change with a higher gradient (i.e. near the substrate ports), a more finely grained grid should be placed, while in the other parts of the substrate, a more coarse mesh density can be used in order to reduce the size of the mesh [43].

Each linear mesh with n ports can be reduced to an equivalent network with $n(n - 1)/2$ impedances. As the internal nodes of the substrate mesh are not of interest in analyzing the substrate coupling, it would be ideal to reduce the mesh so that only the port nodes remain. If n is the number of substrate contacts, including the backplane (if present), the total number of impedances in such a reduced mesh would be $n(n - 1)/2$. For large digital circuits with many substrate contacts, this number is still prohibitively large. Such a large network cannot be included into an analog SPICE-like simulator. Additionally, a matrix representing this impedance mesh would not be sparse anymore [44], which makes it much more difficult to solve. Thus, in order to be able to simulate larger systems, some further approximations have to be applied, to minimize the number of impedances needed in the model [23]. Which further approximations can be taken depends on the type of the substrate and on the purpose of the model.

In epi-type substrates, bulk layer can be considered as a single electrical node, which leads to greatly simplifying the substrate impedance network [45][46].

For lightly doped substrates, some other approximations have to be taken. As shown in [23] and [47], dominant source of substrate noise strongly depends on parasitic inductance of package supply pins. Only for very small package inductances, MOSFET noise (consisting of capacitively coupled noise) is dominant; power supply noise coupling is dominant otherwise. Because of that, as

pointed out in [48], all the noise sources except direct coupling from substrate contacts can be neglected. On the other hand, if the parasitic impedance of the supply rails can be neglected compared to the impedance of the package (which is the case for most packages, especially at higher frequencies), it can be assumed that the ground bounce caused by switching is almost identical at every point on the ground line. Therefore, within one power domain, all of these substrate contacts can be shorted to a single node [48]. This greatly reduces the number of impedances needed to model the substrate. Similar approach is also used in [49].

Some papers go even further in the approximation, and consider the complete substrate underneath the digital aggressor as a single node, regardless of the type of the substrate [50][51].

A more precise approach in reducing the number of substrate contacts is given in [22]. In this paper, voltage domains in substrate, i.e. areas where the substrate is biased with approximately the same voltage, are detected. Substrate contacts belonging to a specific voltage domain are further replaced by a single equivalent contact.

In [44] a placement algorithm taking into account substrate noise was proposed. In this approach, simulated annealing algorithm was used to find the optimum placement, which included evaluating substrate noise in each algorithm iteration. All the mesh nodes of the substrate surface were considered as ports, and substrate mesh was initially reduced to a mesh with n nodes, where n was the number of substrate surface nodes of the mesh. The impedance matrix representing such a mesh is not sparse, and evaluating noise by using such a matrix in each iteration would take a prohibitively long time. Thus, the number of impedances in the mesh, i.e. the number of non-zero elements of the matrix, was reduced so that only impedances lower than a certain threshold are kept in the mesh. The threshold is initially low, and it gets higher as the algorithm advances, thus increasing precision. This combines well with the simultaneous annealing algorithm, which doesn't require high precision in its initial phases. However, this method requires the layout of each component to be known, and surface elements from the layout (contacts, diffusion, nwells) are connected to the substrate grid. This can be somewhat simplified if noise frequency which is being analyzed is low enough so that the substrate can be considered purely resistive – in that case, only resistive connection to the substrate (i.e. only the positions of substrate contacts) have to be known.

As already said, *aggressor models* can have different complexity, depending on the purpose of the model. In [37], substrate was modeled in order to develop a tool for synthesizing an optimal power-delivery network for analog noise-sensitive blocks. Thus, precise level of noise generated in digital circuitry was not of interest, and noise sources were modeled as independent current sources with various PWL waveforms (triangular, trapezoidal, etc.), which inject current into the substrate.

Various aggressor models based on cell characterization were developed. In [50], each standard cell was characterized, and its „noise signature“ was determined, which was later combined with the switching events which were detected from an event-driven simulator. However, noise characterization in this paper was done by not taking into account the impact from the supply/ground noise, which is actually the dominant source of substrate noise in realistic systems. In [51], aggressor modeling is also based on cell characterization, but here the noise introduced from power-supply lines was taken into account as the dominant noise source. Similar approach is also taken in the SWAN tool [47][48].

The most comprehensive approach regarding the aggressor modeling is proposed by [15][52]. It is also based on cell characterization, but offers different levels of it. So, macromodels for the cells can be obtained from gate-level simulation, SPICE netlist, or SPICE netlist extracted from cell layout. As the macromodel complexity increases, the accuracy increases as well, however at the cost of simulation speed. Also, different substrate models are possible – the most precise is extracted from the full layout, while the more simple one is extracted as a purely resistive network by having only the information about the positions of substrate contacts. Additionally, the simplest substrate model, requiring no layout information, was proposed, but it stayed unclear whether it can be applicable to lightly-doped substrates at all.

As a contrast to the aforementioned characterization-based aggressor models, in [49] a model based on power waveform estimation was applied. Knowing the power waveform P_{vdd} , switching current waveform was calculated as $I_{sw} = P_{vdd}/V_{dd}$. This current was included as a independent current generator in the substrate coupling model, containing also of package parasitics and the substrate mesh. Substrate mesh was used by using an equal ground bounce approximation. The method was intended for early substrate noise estimation rather than for verification. However, connecting the aggressor to the substrate mesh was done by using the exact positions of contacts, obtained from the circuit layout.

2.3. An Overview of Existing Substrate Noise Reduction Methodologies

In order to counteract the issue of substrate noise coupling in MSICs, a variety of substrate noise reduction methodologies have been developed so far. In this chapter, an overview of these methodologies is given.

There are several possible ways to classify the substrate noise reduction methodologies. A common classification, as proposed in [53], is based on a noise coupling stage which is targeted by a particular substrate noise methodology. Since there are three stages of noise coupling process, three classes of substrate noise reduction methodologies can be differentiated based on this classification:

- Methodologies targeting noise generation
 - These methodologies include e.g. using low-noise digital logic families or spreading of switching activity in digital aggressor. These methodologies are especially interesting, since they can also contribute to reducing the remaining two forms of SSN – ground bounce and EMI.
- Methodologies targeting noise propagation from the aggressor to the victim
 - These methodologies include increasing the distance in lightly-doped substrate, introducing various types of isolation structures such as guard rings, deep trenches or triple-well structures, using SOI technologies, or using active noise reduction techniques.
- Methodologies targeting noise reception by the victim

These methodologies include using noise-resistant analog circuit techniques such as differential topologies, or digital correction techniques similar to those used for elimination of conversion errors in A/D converters.

Another possible classification of substrate noise reduction methodologies is based on the level of hierarchy in the MSIC at which the substrate noise coupling is counteracted. It includes the following classes:

- Physical level techniques

In physical level techniques, the elements or the properties of the technology are used in such a way that they help reducing substrate noise coupling. Examples are using p+ substrate contacts as guard rings, optimizing the floorplanning in technologies with high-resistivity substrates, or generally applying a technology less prone to substrate noise coupling (such as SOI) for fabrication of the IC.

- Circuit level techniques

In circuit level techniques, circuits out of which the aggressor or the victim are comprised are designed in such a way that the noise coupling is minimized. This includes e.g. the usage of low-noise digital families for digital aggressor design, or the usage of the topologies resistant to the substrate noise for analog victim design.

- System level techniques

Finally, in system level techniques, digital aggressor design is optimized on system level, so that the noise generation is minimized. Thus, all system-level methodologies are at the same time also methodologies targeting noise generation.

Of course, none of these two classifications is exclusive. A methodology can fall in more than one category of each of the two classifications - some methodologies can target more than one noise coupling stages, or they can combine techniques at different hierarchy levels. For example, the methodology which will be proposed in Chapter 5 of this thesis targets both noise generation and noise propagation, and it counteracts the noise both on system level and on physical level.

An overview of existing physical and circuit-level methodologies is presented in Section 2.3.1, while an overview of existing system-level methodologies for substrate noise reduction is presented in Section 2.3.2. Only the methodologies targeting noise generation and propagation are analyzed. The methodologies targeting substrate noise reception by applying substrate noise resistive topologies of analog circuits when designing the victim are out of scope of this thesis.

2.3.1. Physical- and circuit-level methodologies for substrate noise reduction

Physical-level methodologies are the most widely used type of substrate noise reduction methodologies. As proposed in [54], these methodologies can further be categorized in two main subgroups, based on the principle used for reducing the substrate noise flow from the aggressor to the victim:

- Methodologies using insulating or semi-insulating regions in the conducting silicon substrate in order to interrupt the noise propagation from the aggressor to the victim. Examples of this type include n+ type guard rings, deep trench isolation, triple-wells and silicon on insulator (SOI). Isolation by distance in lightly-doped substrates also falls into this category.
- Methodologies applying signal grounding paths in order to redirect the noise flow from the victim. The most prominent example of this type are p+ type guard rings.

Another classification, also proposed by [54], is into the methodologies with vertical shielding and the methodologies with horizontal shielding, based on the direction of substrate noise flow which is being inhibited. In this overview, however, physical methodologies were grouped based on the technological options used to implement them:

- Using only the isolating property of lightly doped substrates
- Using commonly available process options for shielding
- Using advanced or additional process options for shielding (such as silicon post-processing or through-silicon vias)
- Using the silicon on insulator (SOI) technology

Circuit-level methodologies, on the other hand, may be categorized into two main categories:

- Circuit-level methodologies applied at aggressor's side, reducing the noise injection. An example of this is applying low-noise logic families in aggressor design.
- Circuit-level methodologies applied at victim's side, reducing the noise at the point where it is received. An example of this is applying circuits for active noise reduction at victim's side.

This section provides an overview of related work regarding existing physical-level and circuit-level methodologies for substrate noise reduction.

2.3.1.1. Using the isolating property of lightly-doped substrates for floorplanning

Since the substrate noise propagation process strongly depends on the type of substrate used, the same is valid for the applicable substrate noise reduction methodologies targeting substrate noise propagation. As advised in [55], the substrate type, package and pin assignment have to be chosen as the first step, and only then it makes sense to consider the applicability of further measures. Generally, in MSICs it's preferable to use lightly doped substrates as opposed to highly doped substrates, since lightly doped substrates provide noise attenuation by distance between the aggressor and the victim [55]. For example, in [56], two technologies (0.18 μm and 0.35 μm) with lightly-doped substrates were examined, and it was shown that increasing the distance between the aggressor and the victim from 50 μm to 100 μm contributes to 5 dB reduction of substrate noise for both technologies.

The isolating property of lightly doped substrate makes it possible to impact the substrate noise at the victim's position by floorplanning, i.e. by choosing the relative position of the aggressors and the victim(s) such that the substrate noise coupling is minimized. Several noise-aware floorplanning methodologies based on this idea have been proposed so far. The reference [49] proposes extending the already existing module placement approaches in digital circuits with additional constraints concerning substrate noise. For this purpose, the substrate was modeled as an

impedance grid containing n elements, with n being the number of substrate contacts. Mathematically, this grid was initially represented by a $n \times n$ impedance matrix. This matrix can be rendered sparse by applying a threshold value for the matrix elements i.e. the grid impedances. A simulated annealing algorithm was applied for performing the placement optimization. During the execution of the algorithm the threshold increases – this way, the estimation in the initial phases of the algorithm is faster (since the matrix is sparser), while the results in the later phases of the algorithm are calculated more precisely.

While reference [49] included substrate noise constraints directly in the existing placement algorithm, the reference [57] took another approach. It started from a mixed-signal design which had already been placed and routed in a standard manner, and for which the analog subsystem had already been implemented. A dynamic rail analysis was then performed on the entire digital circuitry in order to obtain the waveforms for supply and ground noise. These waveforms were used as piecewise linear (PWL) noise sources, connected to the substrate network through the substrate biasing contacts. The substrate network was obtained by extraction using a substrate noise analysis tool. The system was then simulated in SPICE, and noise results were obtained. If the results are unsatisfactory, the procedure has to be repeated in a loop. This procedure requires substrate extraction in every optimization step, which can cause long optimization runtimes.

In the references [58][59][60][61][62] such procedure repetition was avoided by using the analytical estimations for substrate noise transfer function from digital aggressors to analog victims. Those analytical formulas were derived under an assumption that the digital block is much larger than the analog one. The optimization target of the procedure described in [58] is minimizing the substrate noise at the position of the victim, while [59][60][61] optimize both the noise and the occupied area. Additionally, the optimization algorithm in reference [62] takes care of the symmetry in the analog blocks, which contributes to the resilience of analog circuitry to noise.

References [63][64] are not performing any analytical estimation of the substrate transfer function during the optimization procedure. For all aggressor-victim pairs in the design, noise coupling was estimated before starting the optimization procedure, for a predefined distance between them, and preference lists were formed. Optimization was then performed without using any noise model – just the topological relative positions of blocks were taken into account. The optimization algorithm tries to position the blocks so that they follow the preferred orders as much as possible. Reference [65] also used a similar approach, but in addition to the noise itself, it also aimed at reducing the gradient of the noise at the place of the victim.

Note, however, that the analytical formulas in [58][59][60][61][62] have been derived for substrate noise coupling between one aggressor and one victim. This means that the impact of other blocks was not taken into account at all, which in general case is unjustified, knowing that the presence of other blocks changes the equivalent substrate impedance network, and thus also the noise transfer function from aggressor to the victim.

Additionally, all of the references [58][59][60][61][62][63][64][65] applied an assumption that the noise generated by an aggressor stays unchanged regardless of the position of the aggressor. This assumption is also not applicable in general case. As already explained in the previous sections, the dominant source of substrate noise is usually the direct coupling of ground bounce from ground lines. Ground bounce itself depends on all digital circuits sharing the same power domain,

not just on the single digital circuit observed as an aggressor. Thus, the substrate noise injected by a digital circuit depends not only on that circuit alone, but also on the other circuits sharing the same power domain. In large chips, with significant on-chip parasitics, the noise generated by an aggressor will depend on the relative position of the observed aggressor and the remaining digital circuits sharing the same common power domain. On the other hand, in smaller chips, where on-chip parasitics can be neglected compared to the package parasitics, the entire digital circuitry sharing the same power domain can be considered to have the same ground bounce, i.e. it can be approximately considered that noise is injected into the substrate uniformly across the entire power domain. [P6]

2.3.1.2. Shielding by applying the commonly available process options in lightly-doped and heavily-doped substrates

The most important types of shielding techniques by using commonly available process options in highly-doped or in lightly-doped substrates are presented in Fig. 2.9.

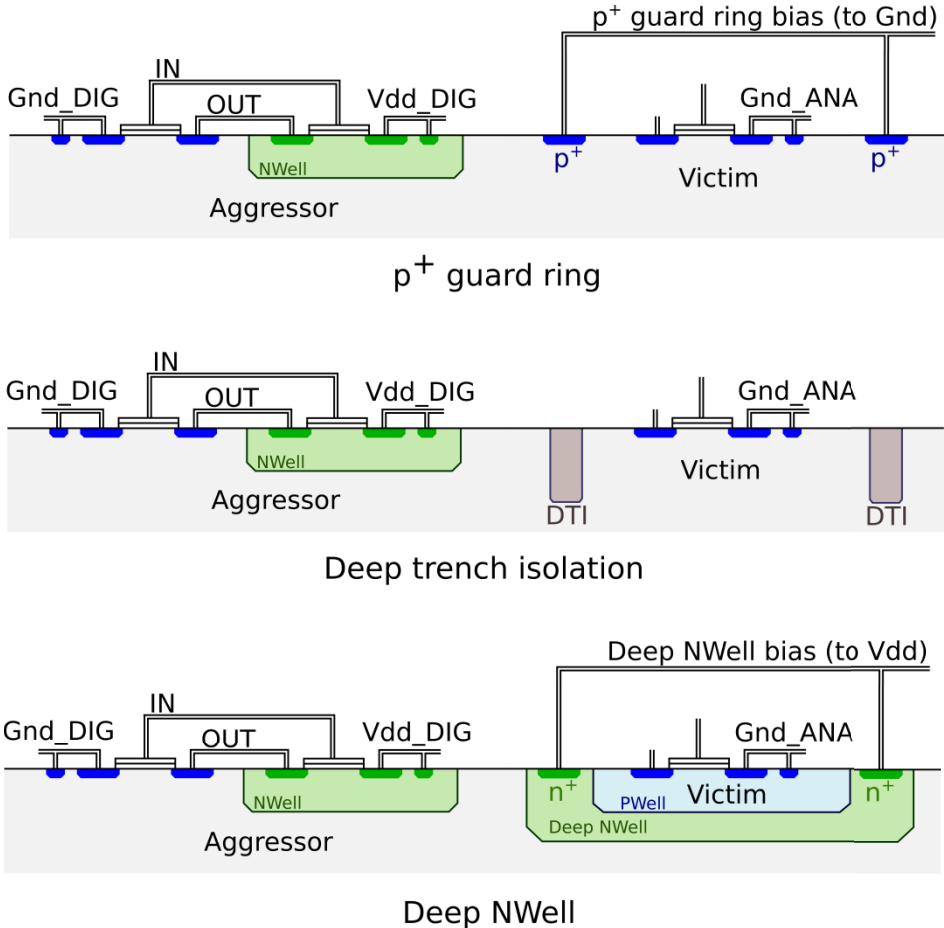


Figure 2.9 – Shielding methodologies (adapted based on [54] and [117])

A large number of papers dealing with shielding methodologies can be found in literature. In [66], several configurations have been examined, including single p⁺ guard ring, double guard ring containing an inner p⁺ ring and an outer NWell ring, and isolation bars (guard stripes) of various types (p⁺, NWell and both p⁺ and NWell together). The guard rings were placed around the

aggressor and/or the victim, while the isolation bars were placed between the aggressor and the victim. The configuration with double guard rings and NWell guard stripes turned out to be the most favorable combination. Additionally, it was concluded that having grounded p+ guard rings is beneficial compared to floating guard rings, and that both width and separation can contribute to a better isolation.

In [67], substrate noise coupling from a single switching NMOS transistor to a NMOS transistor operating in linear region was examined. In order to reduce noise, five types of guard rings were applied: p+ type biased to ground, NWell type guard ring biased to ground, NWell type guard ring biased to supply voltage, a double guard ring consisting of an internal p+ guard ring and an external NWell guard ring, both connected to the ground, and double guard ring consisting of an internal p+ guard ring and an external NWell guard ring, with p+ guard ring connected to the ground and the NWell guard ring connected to the supply voltage. Interestingly, here the simple p+ guard ring connected to the ground resulted in the lowest noise, with the two double guard ring configurations narrowly following, and the NWell configurations resulting in the highest noise. Decreasing the distance between the guard ring and the victim also proved to be beneficial for noise reduction. In [68], on the other side, it was shown that keeping the ring close to the aggressor can somewhat improve the isolation, while modifying the guard ring width was of little influence. A p+ guard ring was also applied in [69], where about 15 dB noise reduction was achieved.

While a p+ guard ring reduces the substrate noise by providing an alternative low-impedance path and thus redirecting the noise flow, deep trench isolation (DTI) blocks the direct path between an aggressor and a victim with an isolating material, thus forcing the noise flow to deeper substrate regions where it has to traverse a longer path to the victim. In [70], substrate noise coupling between two inductors was examined, and deep trench isolation (DTI) was applied in order to reduce the coupling. An attenuation of about 20 dB for frequencies higher than 3 GHz was achieved. DTI has also been successfully applied in [71] in order to break a buried p-layer which may act as a low-impedance path for the noise.

Triple wells (also called deep N Wells or n-isolation) offer a p/n junction isolation (i.e. a capacitive isolation) for the circuitry they enclose. A triple well consists of a PWell which is placed inside a larger NWell, which is placed inside the p-type substrate.

The circuitry to be protected is placed in the PWell, while the aggressor has to be kept outside of it. In [72], the isolation by a p+ type guard ring and by an isolated PWell (i.e. by a triple well) were compared, both for a highly-doped substrate and for a lightly-doped substrate. In a highly-doped substrate, a p+ guard ring did not contribute to reducing the substrate noise. A triple-well reduced substrate noise for around 20 dB at low frequencies, however the isolation significantly deteriorated at higher frequencies. In lightly-doped substrates, both p+ guard ring and triple well achieved a significant attenuation, ranging from about 40 dB at low frequencies to about 30 dB at frequencies around 10 GHz. Triple well has proven to be somewhat more efficient at frequencies lower than 10 GHz, but at higher frequencies the two isolation methodologies achieved similar attenuation.

Deep NWell isolation was also applied in [73]. A maximum attenuation of 35 dB at 100 MHz was achieved, without impacting the performance of the isolated transistor. In [74] and [75], not only the victim but also the aggressor was isolated by a so called “n+ pocket structure”, which is actually

an n+ layer which surrounds the circuit from all sides (basically a triple-well structure). An immense noise improvement of 37 dB was achieved at 1 GHz noise frequency. For comparison, a p+ guard ring applied in the same technology on the same circuits achieved a reduction of 5 dB. However, as stated in [75] and [76], this noise improvement comes with an issue of a magnetic energy loss for an inductor which is close to the pocket (in other words, a Q-factor reduction for such an inductor). Thus, such n+ pockets should be kept away from inductors in layout.

A deep NWell (i.e. triple well isolation) was compared to a p+ guard ring also in [Ishak04]. Here, the aggressor was shielded instead of the victim. The triple well exhibited a high isolation peak at about 1 GHz, where its isolation outperformed a guard ring for 45 dB. At all other frequencies, however, the difference was significantly smaller, and at frequencies above 2 GHz it was relatively stable at somewhat less than 10 dB. In [77], various shielding methodologies were compared for coupling between two contacts, with the victim being isolated. The compared structures include p+ and n+ guard rings, deep NWells and combinations of ground rings and deep NWells. The noise attenuations which were achieved varied between 8 dB and 17 dB for the 20 μm distance between the aggressor and the victim at 2.4 GHz. The attenuation for DNW was a couple of decibels better than the attenuation achieved by guard rings, while the best result was obtained by combining the DNW and the p+ guard ring. One more comparison between a p+ guard ring and a deep NWell was presented in [78]. Deep NWell turned out to perform better at low and middle frequencies, up to about 3 GHz. Above that limit, the performance is comparable to that of a p+ guard ring. A comparison to the previously published data for SOI showed that the performance of DNW is similar to that of SOI. In [79], substrate noise isolation by somewhat more than 15 dB was achieved at low frequencies (lower hundreds of MHz) by applying a deep NWell. With the increase in noise frequency, the isolation provided by a deep NWell gradually degraded.

Various isolation structures applying wells were used to improve isolation between two p+ contacts acting as an aggressor and a victim in [80]. Placing the contacts in the same substrate instead of placing them in a common PWell already contributed to 12 dB isolation. Additional 5 dB were achieved by putting a victim in a separate isolated PWell. Surrounding this well by an NWell ring further added 5 dB of isolation. Applying a deep NWell, however, turned out to be the most effective methodology, especially at lower frequencies, with about 30 dB more attenuation at 1 GHz and 10 dB more attenuation at 10 GHz.

Three isolation structures were examined in [81] at the frequencies up to 20 GHz. The first structure included a DTI and a p+ guard ring around the aggressor. The second structure had an additional 40 μm wide and 16 μm long p+ guard stripe between the aggressor and the victim. Finally, in the third structure, the p+ guard stripe was 30 μm wide and 50 μm , and the victim was isolated in a triple well. The second structure was more than 8 dB better than the first in the range from 1 GHz to 20 GHz. While the first two structures exhibited little frequency dependence, the third one (containing a triple well) was strongly frequency dependent, with isolation deteriorating as the frequency increases. At frequencies below 2 GHz, the third structure outperformed the second for more than 15 dB, but at 10 GHz this advantage was only about 3 dB. At frequencies above about 12 GHz, the third structure provides less isolation than the second one, and at 20 GHz, it is barely better than the first one. It was discussed that the frequency dependence of the isolation provided by triple wells is strongly dependent on the technology applied, i.e. on doping levels of the substrate and the well, since that determines the parasitic capacitance of the well.

In [56], various shielding methodologies were compared for two technologies (0.18 μm and 0.35 μm) with lightly-doped substrate. A p+ guard ring improves isolation by 30 dB at low frequencies for both technologies. Above 1 GHz, however, the isolation diminishes, initially at about 10 dB/dec, and then even more abruptly. While the effectivity of a p+ guard ring was almost the same in both technologies, the effectivity of the n+ guard ring strongly depends on its implementation, which is different in two technologies which were examined. In 0.35 μm technology, the n+ guard ring was implemented as a shallow diffusion, and it didn't contribute to additional noise reduction. In 0.18 μm technology, the n+ guard ring was implemented as an NWell, and it improved noise reduction for additional 5 dB. Similar to [81], the performance of the triple well was strongly dependent on frequency – while it can be significant at low frequencies (e.g. around 10 dB noise reduction at 100 MHz), at about 10 GHz the improvement by the triple well was completely lost. Adding a guard ring inside the well significantly improved the isolation – additional 20 dB of noise reduction were achieved.

As already mentioned, the most important paths of substrate noise injection to the victim are either through substrate contacts or via body effect. However, if inductors with big area are used, the capacitive coupling from the substrate to the metal layers of the inductor may become significant. In [82], two techniques for reduction of substrate noise coupling to the inductors of a LC VCO were examined: a deep NWell underneath the inductor, and a patterned ground shield consisting of metal stripes underneath the inductor. Metal stripes were used instead of a continuous metal plate in order to improve the Q-factor of the inductor. As a reference, an inductor surrounded by a p+ guard ring was used. All the shielding structures were biased to the VCO ground. The intermodulation spur of third order (IM3) at the output of the VCO, caused by the substrate noise coupling, was reduced for 6 to 8 dB with the deep NWell and by 10 to 15 dB with the patterned ground shield.

In [83], the impact of substrate noise isolation methodologies on the output of a reference voltage bandgap was examined. Applying a deep NWell (called n-isolation in the paper) contributed to about 10 % reduction in root mean square (rms) noise voltage at the bandgap output. On the other hand, applying a PWell ring even worsened the noise. However, a combination of these two brought up the best result, with about 25 % reduction in rms noise voltage at the bandgap output. Package impact was also examined, and it was concluded that packaging solutions with lower parasitics are beneficial for substrate noise isolation.

A deep NWell guard ring (DNW-GR) was introduced in [84]. As compared to a common deep NWell (DNW, also called triple-well), where the victim is inside the well, here there's no well beneath the victim. Instead, the deep NWell surrounds the victim like a guard ring. The DNW-GR implemented on a substrate with resistivity of 1 $\text{k}\Omega\text{cm}$ was compared to a standard p+ guard ring implemented on a 10 Ωcm substrate, and showed a 20 dB lower noise level. However, it has to be noticed that the improvement comes partially from a higher substrate resistivity, not only from the superiority of the shielding technique.

A p+ guard ring was compared to a NWell guard ring in [85]. The p+ guard ring provided a substantially better isolation than the NWell guard ring at low frequencies, with more than 15 dB difference at about 250 MHz. However, at higher frequencies, the isolation provided by an NWell becomes better and above 2 GHz it's only about 2 dB worse than the isolation provided by the p+

guard ring. A guard ring surrounding the victim from three sides only provides an attenuation which is only slightly smaller than the attenuation provided by a guard ring surrounding the victim from all sides, which is a bit lower than 20 dB. The attenuation provided by a single guard stripe between the aggressor and the victim is almost 7 dB worse than the attenuation provided by a complete ring, but with 13 dB still significant.

A deep NWell trench combined with a p+ guard ring was used for shielding in [86]. As opposed to a triple well or a deep NWell pocket (as it was called in this paper), a deep NWell trench doesn't enclose the victim from below, but only surrounds it as a ring. While a deep NWell pocket (i.e. a triple well) provides a considerably better isolation at lower frequencies, at about 10 GHz the isolation provided by a deep NWell trench becomes comparable, and they both provide an isolation of roughly 15 dB.

The effectivity of most of the shielding methodologies depends strongly on biasing. The results of a comparison of a guard ring and a triple well were presented in [87][88], with regards to the quality of biasing. For perfect biasing (without any parasitics), the attenuation achieved by a triple well is 13 dB better than the one achieved by a guard ring. However, with only 500 pH inductance in the biasing, the attenuation achieved by a guard ring outperforms that of a triple well by a couple of dB. The efficiency of the well depends also on the noise frequency; at high frequencies and with large inductive parasitics, it's possible for a guard ring even to increase the noise, by making the path to the ground high-impedance, and the path from one point on the chip to another through the ring a low-impedance path [87]. The effectiveness of a ring is also dependent on the ring width. As stated in [87], larger guard rings provide better isolation at lower frequencies but lose their advantage at higher frequencies.

The impact of substrate noise at the output of an LC VCO was observed in [89]. In order to reduce the noise, a p+ guard ring was applied. The dependence of the isolation on the width and biasing of the guard ring was examined. The isolation improves with the width of the ring, but not linearly – instead, it reaches saturation at some point. Increasing the impedance of the ground connection used for the biasing worsens the isolation.

In [90], the triple well isolation was examined in a broad range of frequencies, all the way up to 50 GHz. The dependence of isolation effectiveness on the parasitic bias resistance of the well was explored. Similar to other papers, this one states that better results regarding noise isolation are achieved for smaller parasitic bias resistance. However, it was also shown that the triple well impacts the performance of the isolated NMOS transistor. Floating bias (i.e. bias with a high resistance to ground) turns out to be preferable for the NMOS parameters – it results in a higher voltage gain and a higher output impedance. So, the biasing conditions preferable for better noise isolation are not at the same time preferable for NMOS parameters.

Substrate noise isolation by p+ guard rings and by a triple well, and its dependence on ring and well parameters, respectively, was investigated in [91]. Here, the application of a p+ guard ring didn't result in any noise improvement. The reason was that the biasing parasitics were too high. For a triple well, it was confirmed that the achieved isolation is dependent on the well area. Larger area results in a larger well capacitance, which results in a lower impedance (which especially comes into effect at higher frequencies) – thus, increasing the area of a triple well worsens the isolation.

Substrate noise suppression by a p+ guard ring, with victim ground and the ground ring being biased by the same ground lines, was analyzed in [92]. It was shown that the noise injection here dominantly consists of two mechanisms – coupling through the body effect from the bulk, which is dominant for lower ground resistances, and coupling from the substrate to the ground through the substrate contacts, which is dominant for higher ground resistances. The optimum reduction was achieved for ground resistance of 0.5Ω .

In [27], substrate noise coupling between two p+ contacts and between two n+ contacts was investigated. For reduction of noise coupling from p+ contact to p+ contact, two shielding techniques were examined: a p+ guard ring and a deep NWell. On the other hand, for reduction of noise coupling from n+ contact to n+ contact, a n+ guard ring was used. The two guard ring techniques (p+ and n+) performed similarly at almost entire frequency range. The noise reduction in both cases was about 30 dB at low frequencies and it decreased to about 10 dB at 40 GHz. At lower frequencies (below 1 GHz) the n+ guard ring performed slightly better than the p+ guard ring. This was explained by the fact that at low frequencies almost the entire noise flow is limited to the NWell in which the both contacts are (i.e. almost no noise reaches the substrate underneath the well); since the noise flows closer to the surface, it gets collected by the ring more easily. The deep NWell provided the best isolation at lower frequencies (about 10 dB better than the guard rings at the lowest frequencies measured, and about 5 dB better than the guard rings at 2 GHz), but the advantage gradually decreases at higher frequencies. For the frequencies above 25 GHz the isolation provided by the deep NWell falls below the isolation provided by the guard rings. The reason for this is the large coupling through the well capacitance at higher frequencies.

As it can be seen from the references listed so far, the isolation performance of the shielding techniques such as guard rings and especially triple wells can have some dependence on frequency. This dependence is determined by the technology parameters (such as doping profiles), by layout parameters (such as well areas or guard ring widths) and by parasitic impedances of the biasing structures. Thus, it's possible to use this frequency dependence to optimize the noise reduction.

In [93] and [94], forward-biased n+ guard-ring diodes were used to generate variable capacitance, which was resonated with the inductance of substrate biasing in order to form a low-impedance path to the ground. This way, noise reduction by an order of magnitude was reached. The principle from [93] and [94] was used as a basis for the work presented in [95]. Similar to [93] and [94], in [95] the p-n junction cap in the guard ring resonated with the inductive interconnect to the ground, forming a very low impedance path for the substrate noise current – thus additionally enhancing the isolation of the GR. The position of resonance can be controlled in several ways: by PWell conductivity, by separation distance between the contacts and by GR width. This way, a guard ring can be designed for a specific target frequency. The attenuation achieved at the resonant frequency was about 30 dB better than the one achieved by a conventional structure. Note, however, that the frequency range which can be targeted by this methodology is not deliberate – there are limitations due to technology parameters and feasible layout structures dimensions.

2.3.1.3. Silicon post-processing in order to enhance isolation

In the previous section, some existing shielding techniques using process steps which are usually available in serial production were presented. It is, however, also possible to enhance isolation by introducing additional process steps (such as forming porous silicon) or by silicon post-processing after the common process steps are finished.

In [96] and [97], a gap between analog and digital circuitry in heavily-doped substrate was etched from the backside. The achieved substrate attenuation was large – a 30 dB interference signal was completely removed. This, however, came at a large area cost – the gap was about 14 μm wide at the bottom of the epi-layer, and the total loss of space for the active circuitry was about 20 μm wide.

In [98] and [99], by introducing additional process steps, porous silicon was locally formed in a heavily-doped substrate between the analog and digital circuitry. The silicon was formed from the upper side, and further the wafer was thinned, in order to make the porous region as deep as the substrate. Porous silicon has an extremely high resistivity, even above 1 M Ω , and the noise reduction which was achieved this way is immense: at 2 GHz, a reduction of 70 dB was achieved, while at 8 GHz a reduction of 45 dB was achieved. The width of porous silicon region used to achieve these huge numbers, however, was 180 μm , which represents a huge loss of active area. Also, since the methodology was applied in a heavily doped substrate, its effectiveness falls drastically if there's residual silicon left below the porous silicon region because that would represent a low resistance path for noise propagation. In [100], the methodology was also applied in a lightly-doped substrate. The contribution of the porous silicon to isolation in a lightly-doped substrate, however, was not as high as in heavily-doped substrate: 14 dB at 2 GHz and 5 dB at 20 GHz. In combination with two p+ ground stripes, the achieved noise reduction was 34 dB at 2 GHz and 16 dB at 20 GHz. In [101], porous silicon trenches were selectively grown from the backside of the wafer. This selectively-grown porous silicon trench was surrounded from both sides with an NWell and the residual silicon above it was biased from upper side of the die. The application of this trench contributed to 42.8 dB substrate noise isolation at 2 GHz, however once more at a large cost in area, since the width of the trench was about 85 μm .

In [102] and [12], after finishing the silicon process and prior to packaging, a beam of penetrating protons was applied at selected locations on the substrate, thus removing free carriers and degrading carrier mobility in irradiated areas. This way, a 25 to 30 dB substrate noise reduction was achieved. There were however a lot of practical implementation issues which would make applying of this method difficult in practice: achieving mask alignment in commercial operation, possibility of radiation damage to active devices, nuclear activation cooling time and sulfur generation as a byproduct of a nuclear reaction. In [103], a region between an aggressor and a victim was irradiated by a Helium-3 beam from a cyclotron, which created a local semi-insulated region with a resistivity of over 1 k Ωcm in an otherwise low-resistivity (i.e. heavily-doped) silicon. After applying the technique on a structure already having a guard ring, an additional attenuation (compared to the measurement prior to irradiation) was about 10 dB at 2 GHz. The improvement came at a cost of increased area – the irradiated region was 50 μm wide, and an additional 15 μm margin from active area was needed in order to avoid radiation damage.

2.3.1.4. Shielding by applying through-silicon vias (TSVs) or through-silicon trenches

Backside etching, porous silicon and introduction of irradiated areas improve substrate noise isolation by increasing resistance on the noise path through the substrate. Another approach would be to shield the victim by a ground-biased Faraday-cage-like structure. A readily available way to form such structures is to use through-silicon vias (TSVs).

In [104], a Faraday-cage was formed by 10 μm wide through-wafer vias, and substrate noise attenuation of 20 dB at 1 GHz was achieved for 10 μm space between the vias. In [105], the vias were also 10 μm wide with a 10 μm space between them; the achieved substrate noise suppression was 40 dB at 1 GHz and 36 dB at 5 GHz. In [106], a 30 dB noise reduction at 10 GHz and 16 dB at 50 GHz were demonstrated for 100 μm distance between the aggressor and the victim. It was shown that the attenuation achieved doesn't depend on spacing between the vias up to a 40 μm spacing. A double cage was also applied in this paper; however, despite of a significantly larger area penalty, it contributed with only 3 dB additional noise suppression.

In [107], two isolation structures were demonstrated in a heavily-doped substrate: a Faraday cage formed by TSVs, and a porous silicon formed from the backside of the wafer. Both methodologies showed a similar performance: with TSV-based Faraday cage, noise reduction of 5 to 18 dB was achieved up to 35 GHz, while with porous silicon, the noise reduction of 5 to 20 dB up to 35 GHz was achieved.

In [108] and [109] a method which can be used for forming a grounded trench-like Faraday cage named "self-aligned wafer level integration technology – SAWLIT" was presented. In this method, separate chips were integrated on a common silicon interposer. The interposer was first etched with cavities slightly bigger than chips. Further, the interposer was metalized both from the sides and from the backside to create metal walls which form a continuous Faraday cage. As the next steps, the chips were inserted in these cavities. Finally, the handling substrate underneath the cavities was separated, and metallization and passive components were added to the interposer, thus providing communication between the chips now lying in cavities. Compared to a high-resistivity (lightly-doped) substrate, the improvement in noise separation of 20 dB for frequencies above 2 GHz and up to 25 GHz was achieved. For lower frequencies, however, the improvement becomes much smaller. On the other hand, compared to a low-resistivity (heavily-doped) substrate, the improvement was more than 20 dB for frequencies above 5 GHz and up to 25 GHz, while for lower frequencies the improvement was even larger than 30 dB.

In [110] an H-shaped trench-shaped grounded TSV, enclosing both the aggressor and the victim from three sides, was used to reduce substrate coupling. The attenuation of 30 dB at 100 MHz and 30 dB at 1 GHz was achieved.

2.3.1.5. Silicon on insulator (SOI)

Silicon on insulator (SOI) technologies have a buried oxide layer separating the silicon substrate beneath it from the shallow silicon layer above. The active devices are integrated in the

shallow layer above the oxide, which usually has depths in the order of magnitude from 0.1 μm to above 1 μm , depending on the type of the SOI technology applied [111]. Regarding the substrate noise propagation, the oxide acts as vertical shielding, which prevents the noise from flowing through the depth of the substrate and limits its flow to the surface region. Adding a trench isolation into the surface region can then completely isolate the victim and further improve substrate noise reduction. An example profile of a SOI technology is shown in Fig. 2.10.

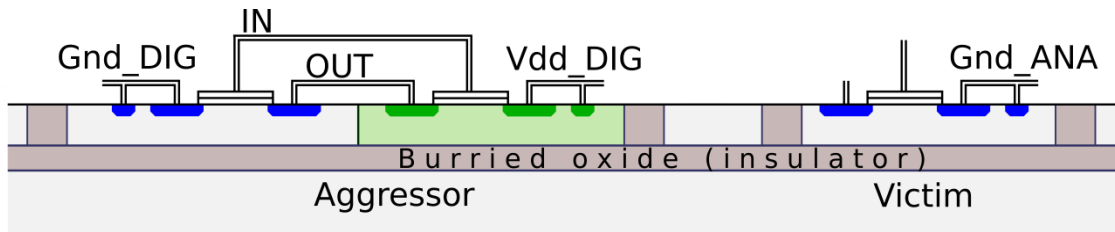


Figure 2.10 – A profile of an SOI technology (adapted based on [117])

The application of SOI technologies in order to mitigate substrate noise in mixed-signal integrated circuits has been investigated for a long time. One disadvantage of lightly-doped substrates compared to heavily-doped (i.e. epi-type) substrates is that they are much more prone to the latch-up effect. In [112], epi-substrates with various epi-thicknesses, a lightly-doped substrate and a SOI technology were compared both in terms of latch-up and noise. Additionally, junction isolation via PWell with a buried n-layer beneath it was analyzed. While epi-substrates are generally less prone to latch-up (the thinner the epi-layer, the less the possibility of latch-up), they have a poor noise performance. Bulk substrates are quite the opposite – while they provide a good noise performance due to their inherent ability to attenuate noise with distance, they are the most prone to latch-up. SOI technologies, as well as a PWell with a buried layer beneath, have both good noise performance and low susceptibility to latch-up.

Substrate noise levels in SOI and bulk substrates without any additional isolating measures were also compared in [113]. It turned out that the noise in SOI was about 10 dB lower than in bulk substrate in a frequency range from 100 MHz to 1 GHz. At higher frequencies, however, there was little difference, due to the increased capacitive coupling through the oxide. Also in [114] it was shown that SOI technology provides 5 dB improvement at 1 GHz compared to a bulk substrate when no additional isolating measures were taken. In [115], [116] and [117], some typical isolation techniques in the bulk substrates, including p+ guard rings, NWell rings and deep trench isolation, were compared to SOI. While SOI was superior to all the examined structures at lower frequencies, at higher frequencies (above 1 GHz) a p+ guard ring provided the best results. Additionally, in [118], [116] and [117], SOI was also compared to a p/n junction isolation, i.e. to the application of a triple well. Here, the triple well was combined with a p+ guard ring inside the well, surrounding the victim. The noise reduction achieved by junction isolation was 2 to 11 dB better than the one achieved by SOI in the frequency range up to 1 GHz. They were, however, both outperformed by a simple p+ guard ring at higher frequencies (about 1 GHz and above) [116].

So, while SOI performs better than bulk substrates at lower frequency, the advantage is lost at higher frequencies, around and above 1 GHz. In order to counteract this issue, [119] suggests using a special type of SOI where the substrate underneath the oxide layer has a high resistivity. This way, a significant crosstalk reduction can be achieved for frequencies up to 10 GHz. The idea was

further analyzed in [120] which also showed that, by increasing the substrate resistivity from 20 Ωcm to 200 Ωcm , the substrate noise reduction improves in the frequency range up to 10 GHz. While the improvement was small for very low frequencies, at about 1 GHz it reached its highest value of about 15 dB, and for higher frequencies it decreased, until at around 10 GHz the difference between the SOI version with higher and lower substrate resistivity completely diminished. Numerical simulations in [120] showed that by increasing the substrate resistivity beneath the oxide, the frequency region in which the noise reduction performance is better than in a bulk substrate extends towards higher frequencies. One more advantage compared to bulk substrates is that in SOI much lower doping rates (and thus higher substrate resistivities) are possible without affecting the performance of active devices, since they are in the surface silicon region with higher doping rate. Besides, the isolation by the oxide makes the lower doping rates more feasible by preventing the doping impurities from the process steps to propagate into the substrate [120]. A very high-resistivity SOI, with substrate resistivity above 1 $\text{k}\Omega\text{cm}$ was also applied in [121], and compared to a common SOI with substrate resistivity of 30-40 Ωcm . At frequencies from 100 MHz to several hundred MHz, the improvement was between 5 and 10 dB.

High-resistivity SOI technologies, however, may suffer from a surface conduction effect beneath the surface which degrades the resistivity of the underlying region of substrate [122]. The effect is caused by the transversal electrical field through the SiO_2/Si interface, which induces a conductive surface carrier channel in the substrate [122]. As a solution for this issue, the introduction of a trap-rich layer beneath the oxide was proposed in [122]. The SOI version using this solution was called eSI HR-SOI, which stands for enhanced signal integrity high-resistivity SOI. This solution was compared to a common HR-SOI by measurements in the frequency range between 10 MHz and 25 GHz. At frequencies above a couple of GHz, no difference was observed. However, while at lower frequencies the noise level measured in HR-SOI gradually flattened, in eSI HR-SOI it continued reducing at a 20 dB/dec rate. At 100 MHz, eSI HR-SOI provided a 15 dB reduction in substrate noise compared to a common HR-SOI. Additionally, the noise induced from the substrate to the drain of a victim transistor was measured and compared for the two HR-SOI versions. The largest measured peak in the noise spectrum, which was at 500 kHz, was 25 dB lower in eSI HR-SOI.

While using SOI with a higher substrate resistivity is based on the principle of reducing noise by increasing isolation, a ground-plane SOI (GPSOI) is based on providing an additional low-impedance path to the ground. In GPSOI, a conductive buried plane, locally biased to the ground, is placed between the oxide layer and the substrate beneath it. In [123], a metal-silicide with 2 Ω/sq was used as the ground-plane, and the substrate noise attenuation was improved for 20 dB compared to the one from SOI with high-resistivity substrate from [120]. In [124], by applying a locally grounded GPSOI, a factor of ten improvement was achieved compared to standard SOI. In [125][126] and [127], GPSOI was combined with a Faraday cage structure formed by the vertical trenches filled with metal. The bottom of the Faraday cage was comprised of a buried metal tungsten silicide ground plane beneath the buried oxide layer, while the cage walls were formed by vertical metal-lined n+ polysilicon-filled trenches in the active silicon layer [126]. The trenches formed this way were quite narrow (only 2 μm), so the area penalty was kept low. By comparing this configuration to a standard SOI and a GPSOI without a Faraday cage, an immense improvement was demonstrated – while GPSOI offered about 30 dB additional noise suppression compared to a standard SOI, combining it with a Faraday cage structure increased the noise suppression for further 30 dB, resulting in a total 60 dB improvement compared to a standard SOI in the frequency range up

to 10 GHz. Interestingly, however, when metal walls were replaced by a 0.5 μm deep diffused n+ guard ring biased to ground, the achieved noise suppression was even a couple of dB larger.

Two different versions of guard rings in SOI were examined in [128] – a guard ring in the surface region, and a supporting substrate guard ring (SSGR), reaching the substrate beneath the oxide. The SSGR showed the best performance at 10 GHz frequency, achieving an improvement of 11 dB compared to SOI without a guard ring and 7.5 dB compared to SOI with a surface guard ring. However, around 1 GHz frequency, the SOI with a surface guard ring performed substantially better, and SSGR barely offered any improvement compared to SOI without a guard ring. The general impact of the biasing of the substrate beneath the oxide and of the silicon above the oxide to the noise suppression achieved by SOI was analyzed by simulations in [129]. It was shown that the optimal biasing configuration is having both substrate beneath the oxide and silicon above the oxide biased – the advantage to the case when they are both floating was as high as 50 dB. The simulations, however, assumed ideal biasing. Also, in order to avoid a low-impedance path between the two sides of the oxide, two separate biasing lines should be used for regions below and above the oxide.

In common thin SOI technologies, the depth of the surface silicon region is around 0.1 μm , which impacts the forming of the process layers of the components integrated in such technologies and consequently also influences the parameters of those components [111]. This is not the case with thick SOI substrates, which have depths of more than 1 μm [111]. In [111], the noise reduction performance of thick SOI substrates with deep trench isolation (DTI) was analyzed. When thick SOI was used without DTI, increasing the distance between the aggressor and the victim from 10 μm to 100 μm contributed to substrate noise reduction of around 20 dB for frequencies up to a couple of GHz. Regardless of the distance, it was observed that the substrate noise isolation starts to diminish with frequency at a rate of about 20 dB/dec at frequencies above 5 GHz. The reason for this is that buried oxide becomes transparent at high frequencies. The same happens with the oxide used for DTI; thus, at such high frequencies, it also doesn't make a difference whether DTI is applied or not. For lower frequencies, substrate noise suppression improves when the DTI is applied, and the improvement becomes larger as the frequency decreases. Below 100 MHz this improvement rate even reaches around 40 dB/dec, and the distance between the aggressor and the victim doesn't impact the noise attenuation anymore. The application of double trenches was also analyzed, but it was found that it doesn't make a difference whether a single DTI or a double DTI is applied.

All of the articles mentioned so far in this subsection analyzed substrate noise coupling on test structures, mostly consisting of a substrate contact where the noise is injected (acting as an aggressor) and a substrate contact where the noise is detected (acting as a victim), and the additional shielding mechanisms were applied on such test structures. In [130], real circuits were used for a comparison between a 0.18 μm SiGe BiCMOS process with a bulk substrate and an SOI technology. As an aggressor, a stepped buffer was used, while a fully-differential wide-band sense amplifier with low gain was used as a victim. At the same separation between the aggressor and the victim, SOI suppressed the noise more efficiently than the bulk substrate, with some 50% to 70% lower time-domain noise peaks at the output of the sense amplifier. However, the aggressor used a relatively slow 1 MHz clock, which means that the dominant spectral noise peaks could be expected at low frequencies, at which, as reported in earlier papers, SOI performs better than bulk substrates.

In [131], an in-die through back end of the line (BEOL) metal wall structure was applied in SOI in order to reduce substrate noise coupling. This metal wall goes through the oxide and reaches the silicon substrate beneath it. It was formed in the following manner: first the trench ring around the victim was etched by a focused ion beam (FIB), and then this trench was filled with silver nano-powder in a post-CMOS process step. A common source amplifier at 1 GHz was used as a victim, and the impact of the substrate noise at the output of this amplifier was measured. By applying this methodology, a 9 dB reduction in the 3rd order intermodulation (IM3) side spurs at the output of the amplifier was achieved.

2.3.1.6. Applying low-noise logic families in digital aggressor design

As explained in Section 2.1, the substrate noise originating from digital circuitry is caused by simultaneous switching currents in digital logic. In CMOS logic, idle current is theoretically equal to zero, and switching, i.e. loading output and internal capacitances of logic gates, causes huge current peaks. These current peaks further cause ground bounce, which gets injected into the substrate through substrate contacts in digital circuitry. If these current peaks in digital logic could be avoided, substrate noise injection from the digital circuitry would be greatly reduced. Based on this idea of reducing current peaks, several low-noise logic families were developed.

In [132], a fully-differential source-coupled logic (SCL) and folded source-coupled logic (FSCL) families were introduced. In these logic families, MOS differential pairs are used to steer current from constant current sources. The idea is to keep the total current approximately constant, and thus avoid current peaks. The SCL and FSCL inverters are shown in Fig 2.11; as it can be seen, the technique is very similar to bipolar emitter-coupled logic (ECL). Compared to conventional static CMOS logic, switching current spikes in SCL and FSCL families were reduced by about two orders of magnitude [132].

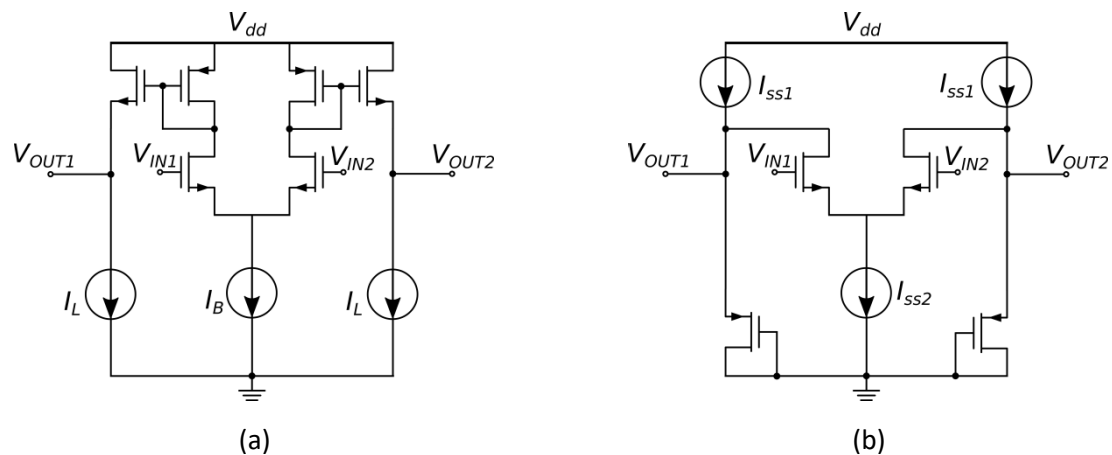


Figure 2.11 – An inverter in SCL (a) and FSCL (b) logic family (adapted from [132])

In [133], a detailed comparison between FSCL and CMOS logic families in terms of switching noise, power, delay and area was presented. The improvement in terms of switching noise was huge: noise peaks in FSCL were 30-300 times smaller than in CMOS. While FSCL had better power-delay product at maximum frequency, the average power at maximum frequency was about two times

larger than for CMOS. This is due to the fact that CMOS consumes power only when switching, while FSCL has a constant current all the time, thus also consuming power all the time. Area of FSCL was also substantially bigger than for CMOS.

While SCL and FSCL logic gates are fully-differential, contain a lot of transistors and occupy a large area, a simpler current steering logic (CSL) family, based on single-ended gates, was proposed in [134]. The version proposed in [134] was a current-mode circuit, which has a drawback that a separate branch is needed for each fanout [135]. Thus, in [135], a voltage-mode CSL logic family was proposed. Inverters in current-mode and voltage-mode CSL logic families are shown in Fig. 2.12. Regarding switching noise, performance of both CSL types was very similar to SCL and FSCL – the switching current spikes were reduced by about two orders of magnitude. As an additional advantage compared to CMOS, it was stated in [133] that CSL logic maintains maximum frequency with reduced supply voltage, which is not the case with CMOS.

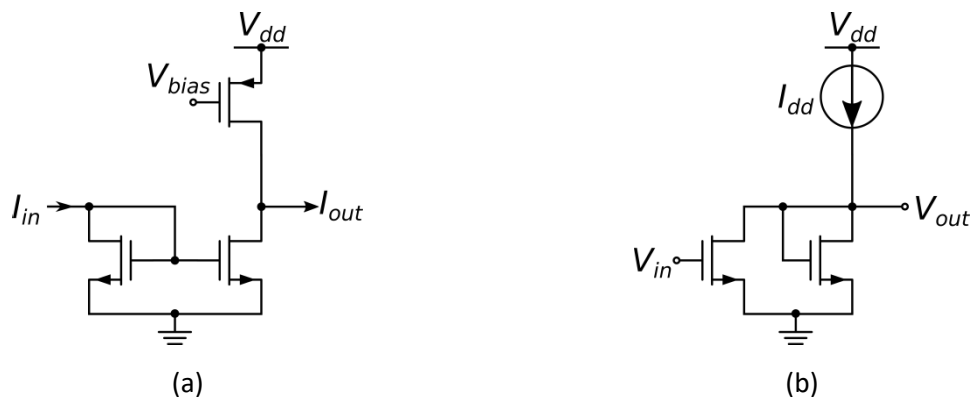


Figure 2.12 – A current-mode CSL inverter and a voltage-mode CSL inverter (adapted from [135])

In [136], FSCL and CSL families were analyzed and compared to each other. As long as the power consumption of both was kept the same, CSL showed a smaller propagation delay. Also, it was concluded that CSL is more robust to process fluctuations, is more capable to work at lower supply voltages and with lower bias currents and it also has a lower transistor count and thus smaller area. However, FSCL performed better in terms of noise: while CSL offered a factor of 10 reduction of switching noise peaks compared to CMOS, the reduction achieved by FSCL was by two orders of magnitude compared to CMOS, and thus one order of magnitude better than CSL.

Another logic family, named current-balanced logic (CBL) was proposed in [137] and [138]. A CBL gate is shown in Fig. 2.13.b. In this logic family, an NMOS pull-down network is used to generate a logic '0'. For a logic '0' at the output, the M_2 transistor from Fig 2.13.b is conducting current. This transistor also acts as a pull-up, generating logic '1' when the NMOS network is not conducting. In case of a logic '1' at the output, the M_3 transistor from Fig 2.13.b is conducting current. The purpose of this transistor is to keep the current consumption constant, which happens if M_2 and M_3 are properly matched. This way, switching current peaks can be avoided. The CBL logic family was reported to have similar degree of noise reduction compared to CSL, while the performance (noise margin, delay) was better and the area was lower for the same supply voltage and consumption current. A further version of CBL, called complementary CBL (C-CBL) was proposed in [139], containing an additional PMOS transistor in the balancing branch, as can be seen in Fig. 2.13.c.

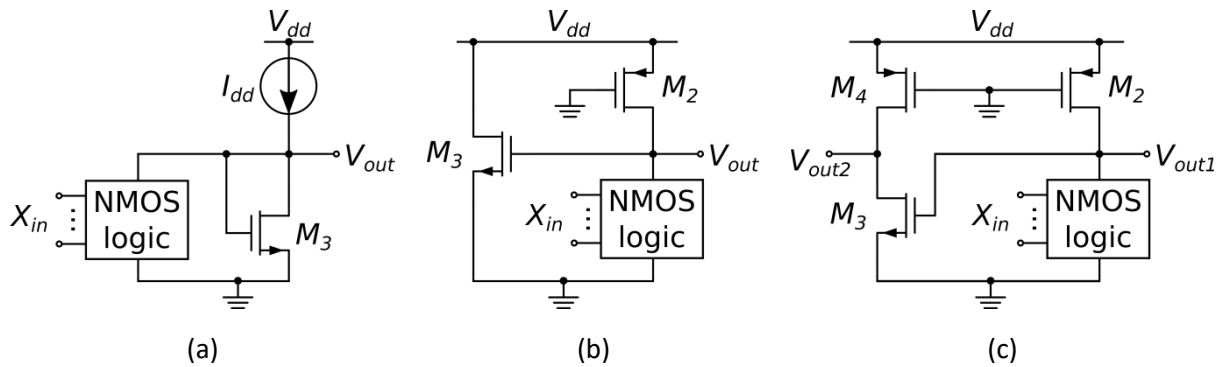


Figure 2.13 – A CSL logic gate (a), a CBL logic gate (b) and a C-CBL logic gate (c) (adapted from [138] and [140])

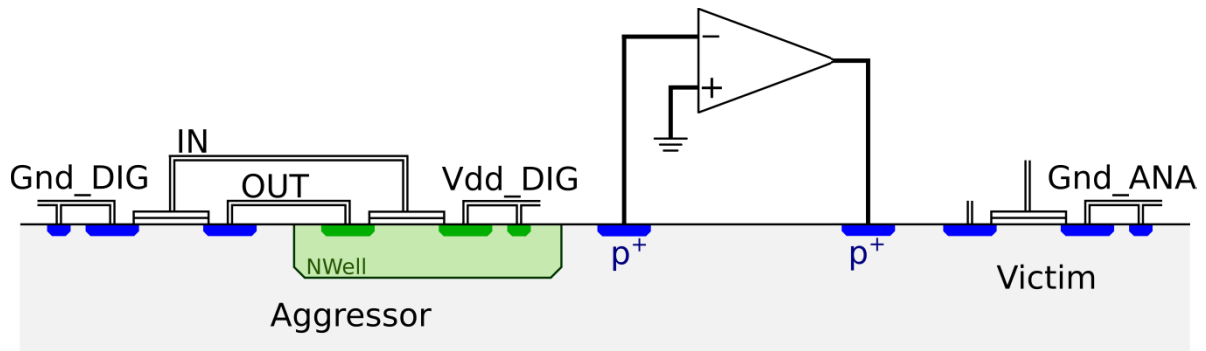
In [140] and [141], substrate noise generated by different logic families was evaluated by simulation and measurements. For small cells, the advantage offered by CSL and CBL compared to CMOS with regards to substrate noise was only marginal. This was because for a single small cell, other mechanisms of substrate noise injection become dominant instead of direct ground bounce coupling through substrate contacts. The same conclusion was applicable for larger cells in case of very low supply inductivity. However, for medium and high inductance levels, larger cells were able to cause larger ground bounce, which consequently became a dominant substrate noise source. Thus, in this case, CBL cells were effective in reducing substrate noise, while CSL cells were less effective.

Generally, despite providing substrate noise reduction, low-noise logic families have never gotten widespread acceptance, although there are still new modifications and improvements being made, like for example D-PFSCS, proposed in [142]. The reason is that CMOS still maintains the two key advantages – low power and low area, which made it the logic family of choice for the overwhelming majority of digital circuits designed and produced today.

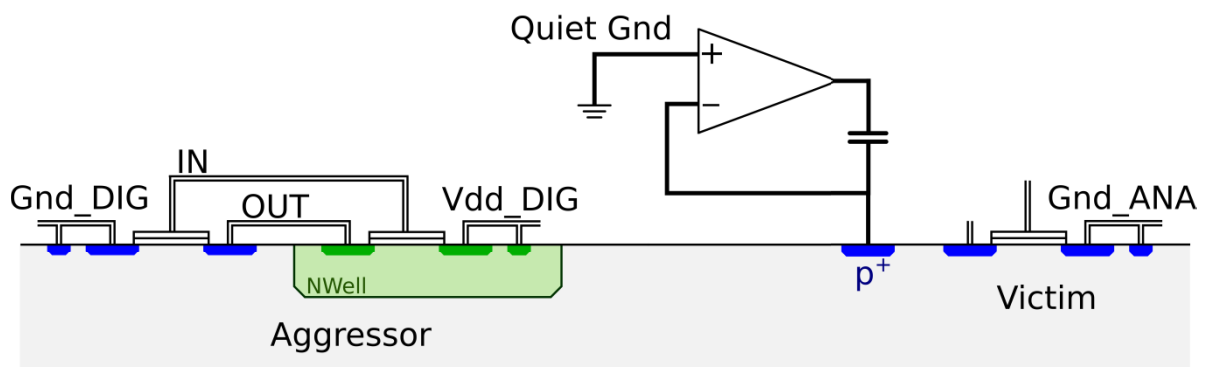
2.3.1.7. Active noise reduction

As already mentioned in the Subsection 2.3.1.3, noise isolation via p+ guard rings is based on providing a low-impedance path in order to divert noise flow. In other words, a p+ guard ring ties the surrounding region of the substrate to a quiet ground voltage, thus keeping substrate noise voltage at zero. However, the efficiency of a p+ guard ring strongly depends on the parasitic impedance of the biasing. Due to that parasitic impedance, substrate voltage under the ring will not be zero. If the parasitic impedance is too large, it can even have an opposite effect and contribute to the increase of the substrate noise.

Active substrate noise reduction, also called active guarding, alleviates this problem by actively regulating the guard ring voltage. The principle of active guarding is shown in Fig 2.14.a. As it can be seen, the amplifier is applied to sense the substrate noise, and after amplification insert the noise of the opposite sign into the substrate.



(a) active guarding



(b) active decoupling

Figure 2.14 – Active substrate noise reduction: (a) active guarding (adapted based on [143]) and (b) active decoupling (adapted based on [148])

In [143], the technique was demonstrated on a CMOS test chip in 0.8 μm technology. The noise cancelation circuit, however, was implemented by using an external operational amplifier. By applying the methodology, the amplitude of the substrate noise (injected as a sinusoidal signal) was suppressed to less than 1% of its original value (from the case when no isolating measure was applied). The suppression was achieved for frequency up to the frequency of 1 MHz – this limit is due to the finite bandwidth of the amplifier. Since the substrate noise is a low level voltage signal close to ground, it is difficult to implement an amplifier for sensing it. An AC coupling configuration, with capacitors between the feedback input of the amplifier and the substrate, as well as between the output of the amplifier and the substrate, was applied in [144]. Substrate noise suppression to less than 5% for frequencies between 100 Hz and 2 MHz was demonstrated in a 0.35 μm technology. Here, the amplifier was integrated on-chip; however, external components were used for the input and output capacitors.

In [145], an active substrate noise reduction circuit was fully integrated in a 0.8 μm SiGe HBT technology. In order to evaluate the methodology, the output spectrum of a ring-oscillator sensor was measured. Due to the substrate noise, sideband spurs can be observed in the spectrum. By applying the methodology, these spurs were reduced for around 20 dB at lower frequencies (around and below 10 MHz) and for 3 dB at around 400 MHz. The price that has to be paid was the additional area which had to be reserved for the active substrate noise reduction circuit, as well as the additional current consumption spent by this circuit.

Active substrate noise reduction based on switched-capacitor amplifiers was applied in [146] and [147] to improve the Signal to Noise plus Distortion Ratio (SNDR) of a delta-sigma modulator. The test circuit was implemented in a 0.25 μm CMOS process with a low-resistivity substrate. It has been shown that the presence of the substrate noise from a switching digital circuit worsens i.e. reduces the SNDR for 20 dB. The application of the active substrate noise shaping improved the SNDR for 10 dB in the frequency range up to 10 kHz, at a cost of a relatively small power consumption increase (additional 3 mW of power, where total consumption of two victim delta-sigma modulators was 2×18 mW).

All of the active noise reduction circuits mentioned in this subsection needed two points for connection to the substrate – a sensing point at their input and injection point at their output. For the circuits presented in [148] and [149], only a single connection point was needed. Instead of sensing the substrate noise and then generating the compensation signal with the opposite phase, this circuit connects the guard ring which is to be regulated to the negative input of an operating amplifier with negative feedback, while the positive input is connected to a quiescent ground. This way, the substrate noise voltage at the position of the guard ring is regulated to the level of a quiescent ground. This technique was called “active decoupling”, and is shown in Fig. 2.14.b. A test chip was implemented in a 0.13 μm CMOS technology, and the noise levels were compared to a case where guard ring is capacitively coupled to a quiescent ground. The improvement ranged from 21 dB at 40 MHz to 13 dB at 1 GHz.

In [150], the active decoupling was applied in a stacked 3D integrated system with through-silicon vias (TSVs). The goal was not only to reduce the noise caused by the aggressors on the same stacked die, but also the noise coupled from the TSVs and originating from other stacked dies. The system used for demonstration contained four layers, three of them digital and one mixed-signal layer. It was shown that the best results are obtained if the analog substrate decouplers are placed in the mixed-signal layer, close to the analog victim circuitry. In case that there’s a purely analog layer instead of a mixed-signal layer, it was concluded that the best combination is to have the analog layer as the bottom layer, and that the analog substrate decouplers should be placed as close as possible to the digital power TSVs in each layer.

2.3.2. System-level methodologies for substrate noise reduction

While most of the physical- and circuit-level methodologies for substrate noise reduction were developed specifically for reduction of substrate noise in mixed-signal circuits, the system-level methodologies were initially developed for reduction of other simultaneous switching noise forms – EMI caused by switching current, and ground/substrate bounce. Later they were also applied for substrate noise reduction. However, the capability to simultaneously reduce various forms of SSN remains one of their advantages over other methodologies.

All of the system-level methodologies are based on the same basic principle: The switching activity of the digital aggressor is distributed in time by modifying its system-level parameters such as clocking scheme and clock distribution. That way, switching current surges are also distributed in

time. This further causes the reduction of switching current spectrum peaks, ground bounce, and consequently also the substrate noise.

The system-level methodologies for substrate noise reduction can generally be categorized in two groups:

- Modifications of the synchronous system clock. This includes:
 - o Modifications of clock distribution network while maintaining a constant clock frequency (i.e. a set of methodologies known under a common name as “switching current shaping”)
 - o Modifications of clock frequency over time (i.e. clock modulation)
- Applying an asynchronous design approach instead of a synchronous design approach

2.3.2.1. Switching current shaping

As said in the introductory paragraphs of this section, switching current shaping represents a set of system-level SSN reduction methodologies where the clock frequency is kept constant, while the distribution of switching within the clock cycle is modified. The purpose of this modification is to spread the clocking activity as even as possible within the clock cycle, thus minimizing switching current peaks and consequently reducing all other forms of SSN. Several ways to achieve this were proposed so far: using modified clock signal with smooth edges, assigning different clock skews to different parts of the digital design, and clocking parts of digital design with opposite clock edges. The effect of switching current shaping on the switching current in time domain, as well as on the switching current spectrum, is shown in Fig. 2.15.

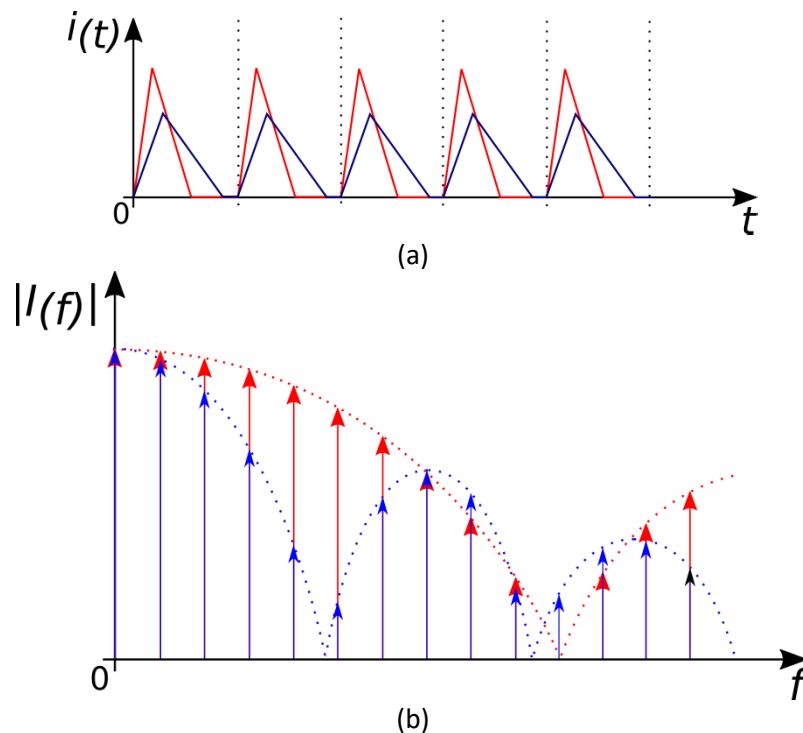


Figure 2.15 – The effect of switching current shaping (a) on the time-domain waveform of the switching current and (b) on the switching current spectrum (adapted from [29])

Historically, this noise reduction principle was first used on output drivers. For example, in [151] the output drivers were skewed in order to reduce the peak-to-peak ground bounce voltage. Later, with digital integrated circuitry growing in size, its contribution to the switching current surpassed that of the output buffers, and the first applications of this principle to the design of digital cores were proposed in [152] and [153].

In [152], clock skew adaptation was applied with a goal of reducing the peak-to-peak ground bounce voltage originating from a digital core. This was achieved by an algorithm which assigns a specific clock arrival time to each flip-flop in the core, such that the resulting current peaks are minimized, while timing constraints remain respected. In order to simplify the problem, only the current originating from clock lines and sequential elements was considered. The algorithm was a genetic algorithm, and as an input it required current waveforms for each sequential element. These waveforms were modeled by two triangular pulses – one for the rising and another for the falling clock edge; their parameters were fitted by a comparison with an analog simulation. The methodology was verified by simulations in a gate-level power simulator. On average, around 30% reduction in peak current was achieved.

In [153], the synchronous system was divided into multiple sub-clock domains with relative skew. In other words, the sequential elements were grouped into several time bins. An optimizing algorithm based on linear programming was presented. The optimization was meant to be applied after placement and global routing, but prior to detailed routing, so that delay estimations from global routing can be used in the algorithm. Sequential elements were clustered in such a way that those with similar skew phases are grouped together and driven by the same clock buffer. The time-domain waveform of switching current was observed, and its first derivative (proportional to ground bounce if ground parasitics are considered to be purely inductive) is minimized. The methodology was verified by SPICE simulations both for digital aggressors in ECL logic and for digital aggressors in CMOS logic. Of course, in order to be able to simulate the aggressor in SPICE, it had to be relatively small. A factor of two noise reduction was achieved for an ECL aggressor, as well as for RCA module as a CMOS aggressor. For a 16-bit shift register applied as a CMOS aggressor, the achieved reduction was even larger – for about one order of magnitude.

While these references dealt with the other SSN forms, the reference [154] explicitly targeted the reduction of substrate noise by shaping the switching current. Here, however, a different approach was used – instead of optimizing the skew of a common rectangular-shaped clock, the clock shape itself was modified, and a clock with smooth edges was used. This way, both the high frequency components of the clock signals (the source of capacitive coupling) as well as the current peaks in the PDN (the source of ground bounce and thus of direct coupling) were reduced. Since the system couldn't operate with such a clock if common sequential elements were used, a special kind of D flip-flops was developed for this purpose. It was shown by a theoretical analysis and by numerical simulations that the root mean square (rms) of the noise voltage coupled from the substrate to the analog ground is approximately reciprocal to the rise and fall time of the clock signal. However, only the contribution from switching of the clock net (i.e. the one clock buffer) was considered. If switching logic was taken into account as well, the dependency on the rise and fall time of the clock signal wouldn't have been so large. The theoretical and numerical evaluation from [154] was applied to fabricate a test chip, and the experimental results were presented in [155] and [156]. In [155] the substrate noise itself was measured, and the reduction achieved by applying the

methodology was between 20% and 54%. In [156], the noise coupled to the output of an analog filter as a consequence of the substrate noise was measured, and the achieved reduction was between 30% and 50%. No data on power and area overheads were provided. However, it was discussed that on one hand the application of a clock with smooth edges and the corresponding kind of DFFs would contribute to increasing the power consumption, while on the other hand the clock buffers can be made smaller, what would contribute to decreasing the power consumption.

The optimization of clock skew with a goal of reducing the substrate noise was applied in [157]. The optimization was done for a heavily-doped substrate. As a cost function of the optimizing algorithm, a product of the peak value of the supply current and its time-domain slope was used. The algorithm distributes the instances of the system into a predetermined number of clock domains. For verification, a 4-bit pseudo random noise generator was used as the digital aggressor which had to be optimized, and the methodology was verified by simulations in SPICE and a custom-made tool named SWAN. The attenuation of the first five substrate noise harmonics was evaluated for the case when the switching current corner frequency (i.e. the width of the main lobe of the switching current spectrum) is below the resonant frequency of the transfer function. The achieved harmonic attenuations were 8.5 dB, 20 dB, 18 dB, 14 dB and 13 dB, from the 1st to the 5th harmonic respectively. The total spectral power was reduced for 6 dB.

This work was further extended in [45]. Here, numerical analysis predicted 34 dB substrate noise attenuation at the resonant frequency, when the initial digital design was partitioned into 5 clock regions. A simulation without timing constraints predicted the attenuation at the resonant frequency of 6.2 dB, 15.3 dB, 23.7 dB and 24.5 dB, in case of partitioning the initial design into 2 to 5 regions, respectively. With latencies also taken into account, the attenuation in case of four regions dropped from 23.7 dB to 6 dB. Finally, a test chip in 0.35 μm CMOS technology was fabricated and the optimized design was compared to the original synchronous design. The digital circuit used in this test chip was an IQ demodulator with a nominal frequency of 42 MHz. The reduction in rms of substrate noise by a factor of two was achieved. While the dominant substrate noise harmonic peak in the original synchronous design was the 3rd harmonic with 19 dBm, the dominant peak in the optimized design was the 2nd harmonic with 13 dBm, which is 6 dBm lower. This was achieved at a cost of 3% area overhead and 4% power overheads. Additionally, a second low-noise-optimized design, which used separate substrate biasing, on-chip decoupling and a dual supply, was also fabricated. While it achieved a slightly better attenuation than the one where clock skew was optimized (with the 1st harmonics at 11 dBm as the dominant one) and even provided a 5% decrease in power consumption (thanks to applying the dual supply), it had a huge area overhead of 70%. The methodology partitioning the digital design into clock regions and assigning different clocks skews to these regions was further formalized in [158].

In [159] clock latencies were manipulated in order to target substrate noise harmonic peaks in a specific frequency band. Firstly, spectral peak reduction by SCS was theoretically analyzed by using a periodic triangular switching current model. It was shown that, for the lower harmonics, satisfying the equation

$$nf_0 < \min\left(\frac{1}{\pi t_{rm}}, \frac{1}{\pi t_{fm}}\right) \quad (2.3)$$

spectral peak attenuation by SCS is negligible. On the other hand, for higher switching current harmonics, satisfying the equation:

$$nf_0 > \max\left(\frac{1}{\pi t_{rm}}, \frac{1}{\pi t_{fm}}\right) \quad (2.4)$$

the switching current magnitude is proportional to $1/\lambda^2$, with λ being the relative width of the switching current pulse. In [160] it was additionally shown that the optimum attenuation is reached for maximum spreading of this relative width, i.e. if the relative current pulse width of the optimized system reaches $\lambda = 1$. In this case, the optimum attenuation which can be achieved is:

$$SPA_n = 40 \log\left(\frac{1}{\lambda_0}\right) \quad (2.5)$$

with λ_0 being the relative current pulse width of the initial system to be optimized.

Further in [159], the two-phased clocking was introduced and theoretically analyzed. In two-phased clocking, the design is partitioned into two clock domains, one clocked by the rising clock edge, and another clocked by the falling clock edge. For such a design, switching current had to be represented by two periodic triangular pulse waveforms (each starting at its respective clock edge). By using such a model, it was mathematically shown that for low switching current harmonics, for which:

$$nf_0 < \min\left(\min\left(\frac{1}{\pi t_{r1}}, \frac{1}{\pi t_{f1}}\right), \min\left(\frac{1}{\pi t_{r2}}, \frac{1}{\pi t_{f2}}\right)\right) \quad (2.6)$$

with t_{r1} and t_{r2} being the rising switching current edges of the two clock domains and t_{f1} and t_{f2} being the falling switching current edges of the two clock domains, the spectral peak attenuation can be expressed as:

$$SPA_n = 20 \log\left|\frac{Q_1 + Q_2}{Q_1 - Q_2}\right| \quad (2.7)$$

if n is odd, while it becomes negligible if n is even. Q_1 and Q_2 in the equation (2.7) represent the areas below the single triangular pulse in the periodic triangular pulse model (i.e. the electric charge transported in one switching cycle) in the first and the second clock domain, respectively.

In order to demonstrate both switching current shaping and two-phase clocking, a test ASIC named *SCREAMER* was implemented in a 130 nm CMOS technology and presented in [159]. It consists of four wireless baseband processors, each with a different clocking strategy: a standard synchronous design, two designs with optimized latencies (the second one with more flexibility in timing constraints, i.e. with a more aggressive optimization), and a manually partitioned design with two clock domains and two-phase clocking. The frequency band between 790 MHz and 910 MHz was targeted for optimization. The optimization was performed by using a dedicated FloorDirector™ tool by Teklatch [161]. The four macros were symmetrically placed, with a ptap substrate contact in the center of each macro, in order to sense substrate noise. The area overhead for all the three optimized macros was small: 1.45 %, 3.46 % and 4.78 % for the less aggressively optimized SCS, more aggressively optimized SCS and two-phase clocking, respectively. Power overheads were moderate for the two SCS designs, 6.52 % and 11.23 % for the less and more aggressively optimized design, respectively. For the two-phase clocking design, however, the power overheads were significantly larger, with 26.98 %. Substrate noise measurements at the fundamental frequency have shown only marginal noise reduction by the two SCS designs, while the two-phase clocked design

provided an attenuation of 9.6 dB. On the other hand, within the targeted frequency band, the two macros with SCS achieved a substrate noise reduction of the dominant in-band harmonic (at 850 MHz) by 5.9 dB and 11.1 dB for the less and more aggressively optimized design, respectively. The two-phase clocked design provided only 3.7 dB spectral peak reduction in this frequency band, although the dominant harmonic was odd. The reason was probably that the order of this harmonic is too high for the approximation (2.7) to be valid.

2.3.2.2. Clock modulation

The basic principle of substrate noise reduction by a clock modulation is to spread the switching activity of a digital circuit by frequency-modulating the synchronous clock which is fed to that digital circuit. This way, the spectral power of each switching current harmonic is spread over a wider bandwidth, resulting in a reduced spectral peak amplitude. The effect of clock modulation on the time-domain switching current waveform, as well as on the switching current spectrum, is shown in Fig. 2.16. Another term often used for this methodology is spread spectrum clock generation (SSCG), analogous to spread spectrum techniques used in communications. Sometimes, the term clock dithering is also used.

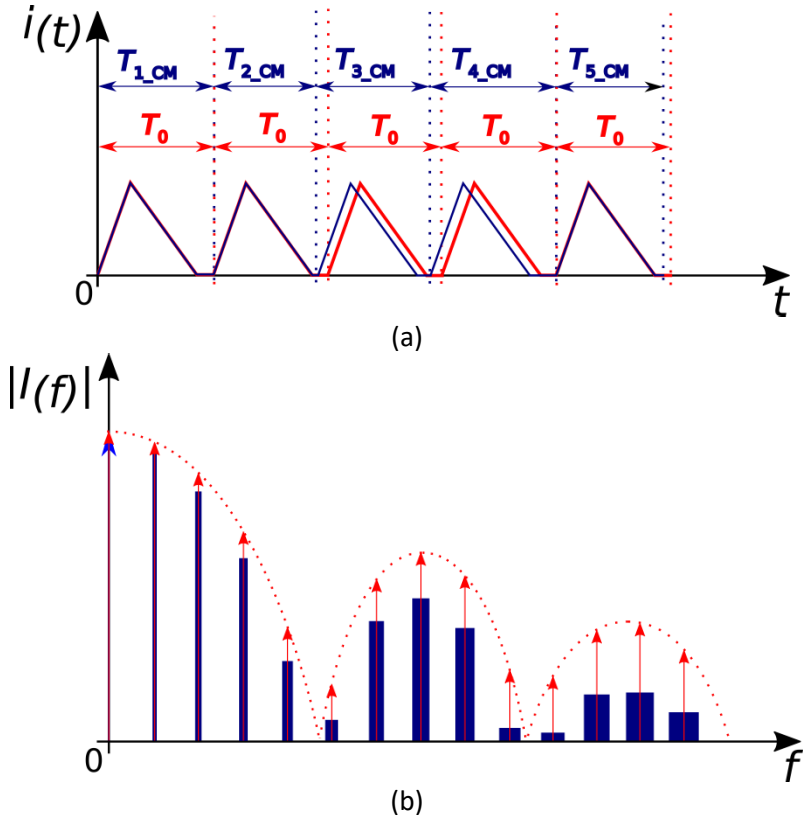


Figure 2.16 – The effect of clock frequency modulation (a) on the time-domain waveform of the switching current and (b) on the switching current spectrum (adapted from [29])

This idea was first used for reducing the EMI of the switching current at switch-mode power supply regulators (i.e. DCDC regulators), as for example in [162], [163] and [164]. The first application of clock modulation for digital integrated systems was proposed by [165]. The initially 20 MHz clock

was modulated by a unique modulating waveform, which was generated by a waveform generator. The modulating waveform had a nonlinear shape which somewhat resembled a modified (slightly curved) triangular shape. The goal was to reduce the EMI of the switching current generated by the digital core. Lower switching current harmonics were marginally attenuated (e.g. about 2 dB attenuation at the 3rd harmonic), while the attenuation as high as 13 dB was achieved when approaching 1 GHz.

A large number of further papers regarding the application of various forms of clock modulation for reducing the EMI radiation caused by the switching current of digital cores have been published so far. Various different modulation waveforms were investigated: chaotic or pseudo-random FM clock [166][167][168][169][170], triangular [171] and piecewise linear [172]. In [173], it was mathematically analyzed which continuous modulating waveform would provide a theoretically optimal attenuation – it turned out to be a waveform with a modified (curved) triangular shape. In [174], the analysis showed that discontinuous modulation waveforms can achieve an attenuation a couple of dB higher than for continuous clock modulations.

In [175], clock frequency modulation was used for reducing the ground bounce, first on its own, and then also combined with switching current shaping. A triangular shaped modulating waveform was used. The clock frequency modulation alone was tested by SPICE simulations on a pseudo-random binary sequencer (PRBS) implemented in a 0.18 μm CMOS process. The PRBS was initially clocked with a 100 MHz clock, and the resonance frequency of a transfer function from the switching current to the ground bounce was at 1.15 GHz. The largest sideband harmonic in the ground bounce spectrum, at 1 GHz, was attenuated by 8 dB. Further, the clock frequency modulation was combined with switching current shaping (by skew optimization) in order to reduce the ground bounce. The switching current shaping was performed as described in [45], with partitioning of the initial design into four clock regions with different skews assigned. This combined methodology was tested on a 40 kGate IC, also implemented in a 0.18 μm CMOS process, by simulations in a custom tool named SWAN (since the circuit was way too large for SPICE simulations to be feasible). The resonance frequency of a transfer function from the switching current to the ground bounce was at 421 MHz. At the most dominant part of the ground bounce spectrum, around the resonance frequency, a spectral peak reduction of about 16 dB was achieved.

The work from [175] was further extended in [29]. Here, a theoretical analysis of maximum achievable attenuation at the p^{th} clock harmonic achievable by clock frequency modulation was given, resulting in an attenuation of:

$$A_{\text{dB}}(p) \approx 10 \log(2\beta) \quad (2.8)$$

where β is a modulation index of the p^{th} clock harmonic:

$$\beta = \frac{\Delta F}{2f_m} \quad (2.9)$$

with f_m being the modulation frequency and ΔF being the maximum deviation of the frequency of the p^{th} clock harmonic from the master clock frequency F_{clk} (i.e. the absolute amount of clock spreading). ΔF can further be expressed as $\Delta F = p\alpha F_{\text{clk}}$, with α being a relative amount of clock spreading (relative to master clock frequency F_{clk}). Note that Eq. (2.8) is actually an approximation which can be derived from the Carson's law [176]. In order to achieve the maximum attenuation, β should be maximized. By adapting the optimum value of β , the methodology from [175] was applied

on the same test circuits again (this time, with slightly different resonance frequencies of the transfer functions - 1 GHz for the PRBS circuit and 475 MHz for a 40 kGate telecommunication circuit). By applying only clock frequency modulation on the PRBS circuit, the dominant ground bounce harmonic at 1 GHz was attenuated for 14.5 dB. By applying a combined clock frequency modulation and switching current shaping on the 40 kGate telecommunication circuit, 26 dB attenuation of the dominant ground bounce harmonic (around the resonant frequency) was achieved.

While the references mentioned so far used the clock frequency modulation for reduction of the EMI produced by the switching current or for reduction of the ground bounce, in [177] using of clock frequency modulation for substrate noise reduction was examined. Here, a pseudo-random sequence of ones and zeros was generated and used to modulate a fixed frequency clock. The clock remains unchanged if a zero is generated, and it becomes inverted if a one is generated. With this kind of clock modulation, the setup- and hold-time constraints from the original circuit remain fulfilled. As a result, however, the digital circuit driven by the pseudo-random clock has a 25 % lower throughput. The methodology was verified by Spectre simulations on a 4-bit 2-stage pipelined Kogge-Stone adder, with supply and ground parasitic inductances included. A 6 dB attenuation at the fundamental frequency (which was 25 MHz) was achieved. Interestingly, in this paper, a possibility to use GALS architecture, with different locally synchronous modules clocked by different pseudo-random clocks was mentioned, but no further analysis of this idea was provided.

In [178], the methodology from [177] was further updated by adjusting the probability distribution of pseudo-random clock generator in order to reduce the throughput penalty. Here, a pseudo-random number generator is driving a counter within a probability adjustment block. This probability adjustment block then drives a pseudo-random clock generator. The average clock frequency at the output (and thus the throughput) is dependent on the length of the counter. For a counter with the length of 5, the throughput reduction was lower than 4 %, while for a counter with the length of 2, the throughput reduction was 25 %, like in [177]. A noise generating network containing four chains of cascaded buffers was used as a test circuit, driven by a 1 GHz clock signal. The circuit was implemented in a 90 nm CMOS technology with a bulk-type substrate. The dominant harmonic in the substrate noise spectrum was the second one, at 2 GHz. The attenuation was dependent on the length of the counter in the probability adjustment block. With a counter length of 2, the dominant peak was attenuated for 7.9 dB, while with a counter length of 5, the attenuation was 3.8 dB.

2.3.2.3. Asynchronous design approach

In asynchronous design approach, no clock is needed to synchronize the data flow. Instead of a clock, an asynchronous handshake with request and acknowledge signals is used to trigger sequential elements. There are various possible realizations of such handshakes, and consequently various families of asynchronous circuits. According to [179], asynchronous circuits can be classified into following four main categories, based on the timing assumptions which need to be fulfilled for the circuit to operate correctly:

- Speed-independent (SI)

- SI circuits operate correctly assuming positive (but unknown) delays in gates and ideal zero-delay wires. Since any realistic wire would have some delay, this category can be considered as only theoretical.
- Delay-insensitive (DI)
 - DI circuits operate correctly assuming positive (but unknown) delays in both gates and wires. This is a small class of circuits – it contains only circuits composed of Muller C-elements and inverters. A Muller C-element is a sequential circuit often used in asynchronous design. Its function can be described as “if $a=b$ then $y \leftarrow a$ ”, with a and b being its inputs, and y its output.
- Quasi-delay-insensitive (QDI)
 - QDI circuits are delay insensitive with an exception of some carefully identified isochronic wire forks. A wire fork is a branching of a single original signal into two branches. A wire fork is isochronic if both branches have matched latencies.
- Self-timed (ST)
 - The correct operation of ST circuits relies on more elaborate timing assumption. A very important example of ST circuits are data paths based on bundled-data protocols. A bundled-data protocol relies on delay matching, which makes sure that the order of events at sender’s end is maintained at the receiver’s end as well.

A vast majority of asynchronous circuits fall into the latest two categories.

Asynchronous digital circuits exhibit various advantages over synchronous digital circuits. In synchronous digital design, the speed is determined by the clock frequency, which is limited by the slowest combinatorial path between two sequential elements driven by that clock. In asynchronous circuits, on the other hand, the speed of data propagation between two sequential elements is dependent only on the delay of the combinatorial circuitry between these elements, which means that asynchronous designs can theoretically work at a higher speed. Besides, as there are no clock buffers, no power is lost in the clock delivery network.

Finally, since there’s no clock in asynchronous circuits, the switching events are not simultaneous and concentrated after clock edges. Instead, they are naturally spread in time. Thanks to this property, the switching noise generated by asynchronous digital circuits can be much lower compared to their synchronous counterparts. Several ASICs demonstrating this low-noise property of the asynchronous digital circuits had been presented in literature.

A configurable self-timed digital signal processor (DSP) was presented in [180]. Together with its synchronous version, it was fabricated in a 0.6 μm technology. In the switching current spectrum, the magnitude of the dominant peak in the self-timed DSP was 1.8 times lower than in the synchronous system. The remaining peaks were much more reduced. In [181], an asynchronous embedded controller (containing, among other blocks, a 32-bit ARM-compatible asynchronous processor core) was presented. The field strength spectrum generated by its evaluation board was compared to the spectrum from a synchronous ARM board. The number of detected harmonics peaks, as well as their magnitude, were significantly lower in the asynchronous design, with the difference in the magnitude of the highest peak being over 10 dB. An asynchronous version of a low-power 80C51 microcontroller was presented in [182], and the results of the supply current spectrum measurements for both synchronous and asynchronous version were shown in [183]. The spectrum

of the asynchronous version exhibited almost no peaks, with its highest region being at a level more than 25 dB lower the highest peak of the switching current spectrum of the synchronous version.

Two simple 4-bit counters, an asynchronous and a synchronous one, were compared for peak switching current in [184]. Due to the periodicity of the input signal, the switching current spectrum of the asynchronous version also had peaks at “clock” harmonics; however, these peaks were between 4 dB and 11 dB lower than the switching current harmonic peaks of the synchronous system. In [185], a digital core of an analog-digital converter (ADC) was implemented in two versions – a synchronous and an asynchronous (self-timed) one. The maximum switching current peak produced by the asynchronous version was about 50 % lower compared to the synchronous version. In [186], a 4-taps FIR filter was implemented in 0.18 μm CMOS technology as an asynchronous circuit and compared to its synchronous counterpart in terms of switching current spectrum. The largest spectral peak in the asynchronous version was 20 % lower than in the synchronous version. A further improvement was made in [187], with the largest peak in the switching current spectrum of the asynchronous version being 9 dB lower than in the synchronous version.

A register file (which could be used e.g. as a part of a central processing unit) was used as a test circuit to compare asynchronous and synchronous design approaches in [188]. Here, the two versions weren't implemented as an ASIC, but in a Xilinx FPGA of the Spartan-3A family. A probe of a spectral analyzer was attached to the FPGA's surface, and the spectrum of the electromagnetic radiation was observed. The highest spectral peak generated by the asynchronous version was 7 dB lower than the highest spectral peak generated by the synchronous version.

The references mentioned so far in this subsection used the asynchronous design approach for reduction of switching current and the electromagnetic radiation originating from it. In [189], synchronous and asynchronous designs were compared specifically in terms of generated substrate noise. As test circuits, a pseudo-random number generator (PRNG) and a 8051 processor were used. The circuits were implemented in a heavily-doped 0.25 μm process. The asynchronous circuit versions were realized as null convention logic (NCL). Substrate noise was probed by a wideband amplifier (with unity gain bandwidth of about 1 GHz) and measured at its output. For the PRNG, 14 dB substrate noise reduction was achieved, at a cost of almost doubling the area. For the 8051 processor, substrate noise reduction was 9.5 dB, with the area increasing from 0.5 mm^2 to 0.62 mm^2 . Additionally, for the 8051 processor, the impact of substrate noise on a delta-sigma modulator integrated on the same substrate was examined. The measurements have shown that in the presence of the operating synchronous 8051 processor on the same substrate, the signal to noise ratio (SNR) of the delta-sigma modulator was degraded for 15 dB. On the other hand, in the presence of the operating asynchronous 8051 processor, no SNR degradation was observed. The work from [190] was further extended in [190]. Here, a 23 dB improvement in the case of the PRNG and a nearly 10 dB improvement in case of the 8051 processor were reported.

2.3.2.4. GALS design approach

To the best of the author's knowledge, no specific GALS-based methodology for substrate noise reduction has been developed prior to this work. However, GALS design approach was used for reduction of the remaining two forms of simultaneous switching noise – EMI and ground bounce. As these methodologies are closely related to the methodologies developed in this thesis, they are presented in this section.

GALS design approach was introduced for the first time in [191]. As already mentioned in the introductory Chapter 1, GALS architectures can be understood as a compromise between synchronous and asynchronous design styles. A GALS system consists of several locally synchronous modules (LSMs), each of them representing a separate clock domain. LSM clocks are generated independently from each other, which means that they are naturally desynchronized. In order for desynchronized LSMs to communicate between each other, they have to be connected by an asynchronous interface, as shown in the example in Fig. 1.1 in Chapter 1.

GALS systems can be classified in several ways, and several taxonomies were proposed [160][192][193]. One way to classify GALS systems is based on the clocking scheme, i.e. on the frequency and phase relation between the LSM clocks. As suggested in [160], based on this criterion there are four categories of GALS designs:

- Mesochronous clocking scheme
 - o LSMs have exactly the same clock frequency and a stable relative phase difference
- Plesiochronous clocking scheme
 - o LSMs have very close clock frequencies, with a difference up to a couple of percents
- Ratiochronous clocking scheme
 - o LSMs have different clock frequencies, but they are rationally related to each other (i.e. their ratio is a rational number)
- Asynchronous clocking scheme (sometimes also named “heterochronous clocking scheme”)
 - o LSMs have nominally different clocks, without any relation in frequency or phase assumed between them

Note that, strictly speaking, the mesochronous clocking scheme would actually represent a synchronous design with different skew inserted for different LSMs, since in order to have exactly the same clock frequency, the clock has to be generated at a common source and then skewed for different LSMs. However, the term “mesochronous clocking” is often used also if the clocks of the LSMs have only nominally the same frequency. In reality, due to process and supply voltage variations, these frequencies would have a slight offset from each other.

Another classification of GALS designs is based on the principle of synchronization used at the interface between the LSMs, i.e. on the type of wrappers used to provide the data transfer between the LSMs. As suggested in [160] and [193], based on this criterion, there are three categories of GALS designs:

- GALS systems with pausable-clock generators
 - o The clocking of LSMs is adaptive (pausable, stretchable or data-driven), so that a clock pulse is generated only when data is transferred, thus avoiding metastability.

- GALS systems with asynchronous FIFO buffers
 - o Data are being written in a circular FIFO buffer at the transmitting LSM's clock edge and read from it at the receiving LSM's clock edge. Write and read pointers address the cells to be written or read. By comparing the pointers, full or empty FIFO can be detected and signaled, thus preventing FIFO overflow or underflow. In order to properly compare the pointers originating from different clock domains, they should be represented by a code with Hamming distance of one.
- GALS systems with boundary synchronization
 - o The synchronization is done by sampling the input signal by two or more cascaded flip-flops.

Each of these three synchronization interfaces has its advantages and drawbacks. While boundary synchronizers are simple to realize and have a low area overhead, they also have a low throughput. Asynchronous FIFOs provide high throughput, but at a cost of a large area overhead. Pausible clocking may provide a good throughput at a cost of a moderate area overhead, but the design effort needed to implement it is comparably the highest. [160]

In each GALS-based switching noise reduction methodology, an initially synchronous system is converted into an equivalent GALS system. In other words, the initially synchronous system is being „galsified“. The process of galsification consists of three main steps:

- Partitioning of an initially synchronous system into LSMs, i.e. assigning the modules of an initially synchronous system to LSMs of an equivalent GALS system
- Choosing the clocking scheme, i.e. assigning the phases and frequencies of the local clocks of the LSMs
- Adding asynchronous wrappers, i.e. clock domain crossing (CDC) circuitry, in order to enable communication between the LSMs

These three steps don't necessarily have to be performed in the order mentioned above (e.g. partitioning can be performed after choosing the clocking scheme).

In narrower sense, the term „galsification“ is also used just for the step of adding the asynchronous interface to a design which has already been partitioned, and for which the clocking scheme has already been chosen. If not explicitly stated otherwise, in this work the term „galsification“ will be used in broader sense, to describe the entire process, including partitioning, clock assignment and galsification in narrower sense.

GALS design approach was applied in [194] in order to reduce spectral peaks in the switching current spectrum. A pausable clocking scheme was applied, with four-phase bundled-data handshake protocol for data transfer between the LSMs. Each LSM had its locally generated clock which could be paused or stretched. Also, each clock could be frequency modulated by a triangular frequency modulator. Additionally, an intentional skew could also be introduced via a programmable delay line within the clock. Note that, while this wouldn't make much sense for independently generated clocks, it would contribute to reducing SSN if LSM clocks were generated from the same synchronous clock, as explained in Subsection 2.3.2.1. Even if the LSM clocks were generated by from the same synchronous clock, asynchronous data transfer would have a couple of advantages. On one side, it would allow for a larger skew than in a synchronous system; also, due to smaller local clock trees, larger frequency deviation would be feasible with less overhead. The approach was verified on a 64-

point pipelined FFT processor with the Radix-2³ single-delay-feedback (SDF) architecture, implemented in a 130 nm technology. The nominal clock frequency of the initial synchronous design was 80 MHz. The initially synchronous design was partitioned into four LSMs with roughly the same average current (i.e. the same average power). The power was estimated by using Synopsys Prime Time. All the LSMs were configured to run at the same frequency, but with a skew of a quarter of a period introduced between them. Additionally, a triangular frequency modulation was applied at each clock, however with a relatively small modulation index of only $\beta = 4$. In order to compare the GALS test design with the original synchronous design, the chip could be set to work either in GALS mode, or in synchronous mode with an externally provided clock. In case of synchronous mode, asynchronous wrappers were bypassed. The measurements have shown that at the fundamental clock frequency (which is usually the dominant component in the switching current spectrum), harmonic peak generated by the GALS system was 13 dB lower than the corresponding peak generated by the synchronous system. For higher harmonics, even larger attenuation was reported.

In [17] the test design from [194] was presented again, but this time with a thorough theoretical analysis of applying GALS with a target of reduction ground bounce. An analytical expression for the time-domain magnitude of ground bounce depending on the switching activity (i.e. the ratio of switching gates to the total number of gates) was presented. It was theoretically shown that lowering the switching activity contributes to ground bounce reduction by two effects. Firstly, with fewer simultaneously switching gates, the switching current peaks in time domain will be lower. Additionally, by decreasing the number of simultaneously switching gates, the number of non-switching gates at the observed time point increases. This consequently leads to the increase of an on-chip equivalent decoupling capacitance, which is also beneficial for reducing the ground bounce peaks. According to this analysis, the optimum partitioning would require distributing the switching activity equally over the LSMs. This would be equivalent to the equally distributed average switching current, i.e. the average power consumption, which would correspond to the same partitioning as presented in [194]. A scheme to enable a robust clock phase modulation was additionally presented. This scheme makes sure that after each clock synchronization the four GALS local clocks remain evenly distributed over the clock period (i.e. with a quarter of a clock delay between them). Ground bounce measurements demonstrated a 40% reduction in peak-to-peak ground bounce in time-domain. A reduction of harmonic peaks of ground bounce was also observed. The fundamental harmonic, which was the most dominant one in the spectrum, was reduced for about 11 dB by applying the GALS design approach.

A GALS demonstrator ASIC named *Moonrake* was presented in [195] and [196]. This test ASIC represented an OFDM baseband transmitter, including a 256-point FFT module, 6 interleaver units and various other blocks. This multi-million gate design was implemented in a 40 nm CMOS process. Two separate versions were integrated on the same die – an original synchronous version, and a galsified version. The GALS version was partitioned into 6 LSMs, with pausable clocking interfaces between them. The interface was configurable for several possible frequency schemes, including nominally mesochronous and heterochronous clocking, and enabling clock frequency sweep from 25 MHz to 265 MHz. Introducing a clock phase shift, i.e. skew, was also enabled. Local clocks were generated by using ring oscillators, which generally display an increased jitter in presence of voltage variations. This, however, is actually helpful for switching noise reduction, since jitter actually represents a parasitic clock modulation which additionally reduces the noise. The partitioning itself was done so that the power and area of the LSMs is balanced as much as possible. Additionally, it

was taken care that the number of communication links between the LSMs is as reduced as possible. The measurements of the supply voltage noise spectrum were first done on a dedicated pad. The dominant spectral peak at the fundamental clock frequency was reduced for 26 dB by applying the GALS design approach. However, it was discussed that this result was most probably somewhat overly optimistic, since the sensing point on the chip is much closer to the synchronous design version. Thus, an additional measurement at another measurement point was performed, showing still a very high 19 dB spectral peak reduction in favor of GALS.

In [16] and [160], a very important theoretical analysis of an optimal GALS partitioning for spectral peak attenuation of switching current was presented. A plesiochronous clocking scheme was used. It was shown that, for partitioning of an initially synchronous design into M LSMs with plesiochronous clocking, each switching current spectral peak of the initial synchronous system was replaced by M lower switching current spectral peaks. In order to theoretically determine the optimal GALS partitioning, a triangular periodic pulse model for switching current was used. It was shown that the optimum result can be achieved by the **power-balanced plesiochronous GALS partitioning methodology**, i.e. by selecting the LSMs so that they all have equal average power:

$$(\forall_m) \left(P_m = \frac{P_{sync}}{M} = \frac{\sum_{m=1}^M P_m}{M} \right) \quad (2.10)$$

where P_m is the average power of the m^{th} LSM, P_{sync} the total power of the initial synchronous system, and M the number of LSMs. Note that an approximation has been made that the power overhead of GALS asynchronous wrappers is negligible. It was mathematically proven that for the low order switching current harmonics, which satisfy the following condition:

$$nf_0 < \min \left(\frac{1}{\pi t_{rm}}, \frac{1}{\pi t_{fm}} \right) \quad (2.11)$$

where f_0 is the initial synchronous frequency, n the order of the harmonic, t_{rm} the rise time of the triangular switching current waveform for the m^{th} LSM and t_{fm} the fall time of the triangular switching current waveform for the m^{th} LSM, the achieved spectral peak attenuation SPA_n can be represented by the following equation:

$$SPA_n = 20 \log M + O(\Lambda) \quad (2.12)$$

where $O(\Lambda)$ is a factor dependent on current profile. For $M \gg 1$, it was proven that this factor is negligible. Consequently, the spectral peak attenuation achieved by power-balanced plesiochronous GALS partitioning can be approximated as:

$$SPA_n \approx 20 \log M \quad (2.13)$$

The applicability of this approximation was confirmed by numerical evaluation in MATLAB. The power-balanced plesiochronous GALS partitioning methodology was then applied for implementing a test chip named *Lighthouse*. The design to be tested was an FMCW-RADAR processor, containing a 4096-point FFT module, and it was implemented in a 130 nm CMOS process. Two version were separately integrated on the same die – the original synchronous design, and a plesiochronous power-balanced GALS design with 5 LSMs. Each LSM was clocked by a separate ring-oscillator clock generator, and pausable clocking was used for data transfer between the LSMs. The overheads for the GALS design were small: 6.61 % in power and 4.35 % in area. The measurements of supply noise have shown that at the fundamental clock harmonic, which was the dominant component in the spectrum, an attenuation of 12.29 dB was achieved.

Note that, although the power-balanced plesiochronous GALS partitioning wasn't aimed at substrate noise reduction, it can be used for that purpose in the special case when all the analog victim modules sharing the same substrate with the digital aggressor circuitry are sensitive at low frequencies (e.g. VCOs). However, this methodology can't be used for protecting analog modules which are sensitive at higher frequencies, such as LNAs. More details on the applicability of the power-balanced plesiochronous GALS partitioning to substrate noise reduction are presented in Chapter 4. In general case, with several different analog modules sensitive in several different frequency bands, it is necessary to be able to simultaneously target substrate noise reduction at different frequency bands. The goal of the methodology presented in Chapter 4 of this thesis is to provide a GALS based approach for switching noise reduction which is capable for targeting higher frequency bands, as well as targeting several frequency bands simultaneously, thus being applicable for substrate noise reduction in general case.

2.4. Summary

In this chapter, the basics principles of simultaneous switching noise and of substrate noise coupling, as well as an overview of related work (both in modeling the substrate noise coupling and in substrate noise reduction) have been presented.

Firstly, the origin of the simultaneous switching noise, as well as its three main forms – switching current, ground bounce and substrate noise – have been presented. The sources of substrate noise have been shown, and the direct coupling of ground bounce through the substrate contacts has been identified as the dominant substrate noise source. The three phases of substrate noise coupling – injection, propagation and reception – have also been explained.

Then, the existing modeling approaches which can be used in simultaneous switching noise analysis have been presented for each of the three forms of the simultaneous switching noise. The attention has especially been paid to the simplified models which are suitable for high abstraction level analysis, such as periodic triangular pulse model used for the switching current and lumped elements model used for ground bounce. For substrate noise, it has been shown that the existing models fall into two categories – the analytical models and the numerical models. The analytical models can be useful approximations for analysis, but their applicability is limited to simple structures, mostly consisting of a single aggressor and a single victim. The numerical models, on the other hand, are precise, but too complex to be used in a simplified analysis at a high level of abstraction. Additionally, in order to apply them, some layout details should be known, which further limits their applicability to later design phases. For the analysis of substrate noise methodologies which are to be presented in this thesis, a model simple enough to be used for theoretical analysis, yet suitable to handle configurations with several deliberately placed aggressors, is needed.

Further, a detailed overview of the existing substrate noise reduction methodologies has been presented, with special attention being given to system-level methodologies. It has been stated that the most widely used substrate noise methodologies nowadays are the “standard” shielding methodologies, such as using p+ guard rings, deep trenches or triple wells. These methodologies, however, do have their drawbacks. The attenuation achieved by p+ guard rings is highly dependent

on the ring biasing, while the triple wells are effective only at lower frequencies. Silicon on insulator (SOI) technologies use a buried oxide for isolation, but similar to triple wells, they also lose their effectiveness at higher frequencies. Using some more sophisticated technology options, such as porous silicon, silicon post-processing by ion beams, or using through-silicon vias to form a Faraday-Cage-like isolation structures can provide immense noise isolation, but it usually comes at a large area penalty. Additionally, they significantly increase the complexity of the technological process, thus increasing the cost. Applying low-noise logic families instead of CMOS in order to reduce the noise significantly increases current consumption. Active noise reduction introduces additional area penalty in analog design, and its efficiency is limited by the bandwidth of the active circuit used.

System level methodologies are all based on spreading of the switching activity in a digital circuit. They generally have the advantage that, besides reducing the substrate noise, they also simultaneously reduce the remaining two forms of the simultaneous switching noise – switching current and ground bounce. Another advantage is that they can easily be combined with any physical-level substrate noise reduction methodology. Two synchronous system-level methodologies, switching current shaping (including also two-phased clocking) and clock modulation, as well as asynchronous design approach, have been used for substrate noise reduction so far. An overview of literature dealing with these methodologies has been presented. For the two synchronous system-level methodologies, the results of the theoretical analysis of their effectivity in switching noise reduction have been presented. While asynchronous designs can provide a significant substrate noise attenuation, there's no general design guideline for optimizing them for noise reduction, which makes every design a separate case-study. Additionally, they usually require large area and power penalties, as well as a huge increase in design process complexity.

GALS design approach, as a compromise between synchronous and asynchronous design styles, has also been presented. It has been used for reduction of switching current and ground bounce, where it has shown promising results. However, to the best of author's knowledge, no specific GALS-based methodology for substrate noise reduction has been developed so far. The results achieved in reducing the remaining two SSN forms by using GALS are a good incentive to investigate the possibility of substrate noise reduction by applying a GALS design approach, which has been done within this thesis.

3. The modeling approach for theoretical analysis of system level methodologies for substrate noise suppression

In order to be able to theoretically analyze system level methodologies for substrate noise suppression in lightly doped substrates, a corresponding substrate noise coupling model at high abstraction level has to be available.

As shown in the previous chapter, the substrate noise generation (i.e. ground bounce as its main source) was so far mostly analyzed in time-domain. The analysis in frequency domain was made for synchronous systems, but not for GALS systems. Substrate noise sensitivity of analog modules depends on noise frequency. Thus, substrate noise analysis in frequency domain is required, and consequently, a frequency domain analysis of ground bounce (as the main source of substrate noise) for GALS systems is necessary.

On the other hand, the existing models for substrate noise propagation can be categorized in analytical and numerical models. As shown in the previous chapter, the applicability of the analytical models is limited to simple structures, mostly consisting of a single aggressor and a single victim. The numerical models are too complex to be used in a simplified analysis at a high level of abstraction. They also require the knowledge of some layout details, which further limits their applicability to later design phases. For the analysis of substrate noise methodologies which are to be presented in this thesis, a model simple enough to be used for theoretical analysis, yet suitable to handle configurations with several deliberately placed aggressors, is needed.

Thus, as an auxiliary task of this work, the models for the first two substrate noise coupling stages – substrate noise generation and substrate noise propagation (visualized in Fig. 3.1) – are developed. The resulting models are presented and their properties are discussed within this chapter.

As a first step, a coarse numerical model of substrate noise propagation in lightly doped substrates is developed [P1]. It represents the substrate as a simple impedance network, which can consist of capacitive and resistive parts, or be purely resistive, depending on the approximation level applied. The primary goal of this model is to enable analysis and comparison of different aggressor floorplanning schemes in terms of substrate noise propagation. The model is presented in Section 3.1.

As explained in Chapter 2, for medium and large scale digital aggressors and for most of the packages, direct ground bounce coupling through substrate contacts is the dominant mechanism of noise injection into the substrate. Thus, substrate noise itself can be viewed as a consequence of supply/ground bounce, and indirectly as a consequence of the switching noise as well. In other words, substrate noise can be viewed as a consequence of the remaining two forms of the SSN. In order to analyze substrate noise generation, these two SSN forms have to be modeled in a suitable way as well. For representing the switching current, the model from [16] is used, which is a system level model applicable for the analysis of GALS systems as well. For representing the ground bounce, a frequency-domain ground bounce model is developed [P2] by using the model from [16] as an initial point. This frequency-domain ground bounce model is presented in Section 3.2.

Thus, a simplified numerical extraction model is developed. A finite difference method (FDM), a technique based on solving the Poisson's equation, has been chosen as an initial point. The substrate is discretized into small cubic domains, and a Poisson's equation is written for each of them. Then, the Poisson's equation is translated into Kirchoff's laws, and a mesh of impedances is obtained, with each discretization domain being replaced by a mesh node. Mesh impedances consist of a resistive and a capacitive component connected in parallel, as shown in the Fig. 2.8. The values of mesh elements are extracted from the specification of the IHP's SiGe:C BiCMOS technologies, which are well-suited for MSICs [25]. These technologies use a lightly-doped substrate, with 50 Ωcm resistivity. The impedances of resistive and capacitive elements in each of the mesh branches will have an equal value at a frequency of 3 GHz; which means that the resistive approximation (i.e. approximating all the mesh branches as being purely resistive) can safely hold up to at least approximately 300 MHz (where the impedance of the capacitive element is an order of magnitude higher than the impedance of the resistive element). In general, for higher frequencies, the RC-mesh approach has to be used. Note that the model obtained in this case is valid only for the specific extraction frequency (and approximately valid for the narrow band around that frequency). If there are more frequencies of interest, the procedure would have to be repeated for each of them. However, as it will be shown in Section 3.3, if the spectrum of ground bounce is taken into consideration, it is possible to use the resistive approximation also for frequencies which are somewhat higher than the border frequency mention above.

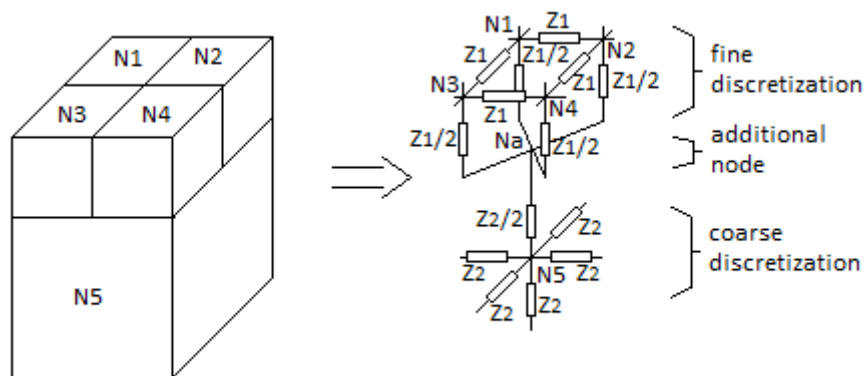


Figure 3.2 – Contacts between elements of different size – adding a node (adapted from [P1])

In order to appropriately take into account the size of substrate contacts, the discretization step on the substrate surface is equal to minimum substrate contact size. For the IHP's 250 nm SiGe:C BiCMOS technology [25], that step is 780 nm, which means that more than 1.6 million nodes are needed to discretize a substrate surface of one square millimeter [P1]. Maintaining such a discretization step throughout the internal structure of the substrate would result in an excessively large number of nodes needed to represent an entire substrate. Thus, in order to reduce the number of nodes in an initial mesh, a variable discretization step has been applied. The discretization is getting coarser for layers further away from the surface, which is justified by the fact that currents flowing away from the surface generally make larger distances, so the discretization error will have a lower impact. In order to properly interface adjacent substrate layers discretized with a different step size, additional nodes are introduced at contact surfaces between cubes of different dimensions, as shown in the example in Fig. 3.2.

Since the direct coupling through substrate contacts is the dominant substrate noise source, all the noise sources other than direct coupling through substrate contacts are neglected. Thus, the only ports of the substrate mesh to the outside world are the substrate contacts. In case that the chip has a conductive backplane, it is also connected to the substrate mesh as an additional contact. In lightly doped substrate, especially for the coupling at longer distances, equivalent substrate impedances between the substrate contacts are much larger than the impedances of substrate contacts themselves. Thanks to this, the impedances of substrate contacts (ptaps) can also be neglected.

The exact positions of substrate contacts are usually not known before the back-end design phase, in which the exact chip layout is determined. In order to be able to make noise estimations prior to this phase, an assumption about the positions of substrate contacts has to be taken, such that the extracted model doesn't differ much from the model extracted by taking into account the actual contact positions. The assumption can be made by making use of the regular structure of digital cells in an IC. All the digital cells have the same height, which is equal to the distance between the adjacent supply and ground stripes in the on-chip PDN. After placement, adjacent cell rows share the same supply/ground stripe. Substrate biasing is done by connecting the substrate contacts to ground stripes. Note that, in general case, substrate contacts don't have to be at the same distance from the cell bottom for different cells. One of the most important placement goals in modern digital circuits is the minimization of the area – thus, digital cells are usually placed with little free space between them (and if there is free space left in between, it is filled with filler cells). Thanks to this, it can be assumed that there is some regularity in horizontal positioning of substrate contacts as well.

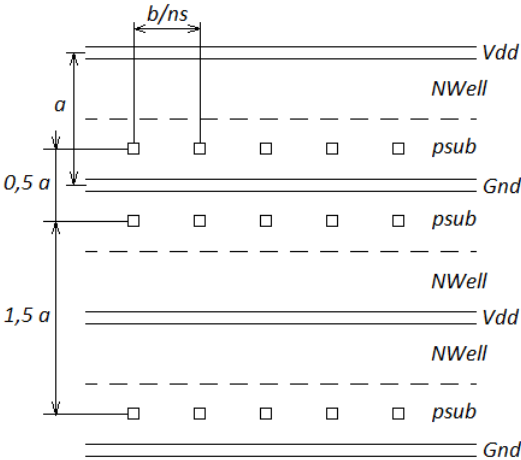


Figure 3.3 – Assumed contact positions (adapted from [P1])

Based on the assumptions mentioned above, the distribution of substrate contacts as shown in Fig. 3.3 is assumed. It is approximated that the contacts form a rectangular net. In vertical direction, each contact is positioned at a distance equal to 0.25 the cell height from the nearest ground stripe, i.e. at a distance equal to 0.75 cell heights from the nearest power stripe. This means that the vertical distance between adjacent contacts of cells sharing the common ground stripe is equal to half of the cell height, while the vertical distance between adjacent contacts of cells sharing the common power stripe is equal to 1.5 cell heights. The average horizontal distance between the substrate contacts can be determined from the area reserved for the digital module, the average cell length, and the number of substrate contacts per cell. Let the standard cell height be a , the length of

the area reserved for the digital module b , the number of cell rows reserved n_h , and the total number of cells to be placed in the module n_c . This means that the total area of the digital module is $n_h \cdot a \cdot b$, where $n_h \cdot a$ is a vertical dimension. From this, the average number of cells in one row can be calculated as $n_s = n_c / (n_h \cdot a)$. In the simplest case of one substrate contact per cell, the average horizontal distance between the adjacent contacts can be calculated as b/n_s . In case of a different average number of contacts per cell, this value should be correspondingly modified. Note that by “average number per cell”, the average over the cells used in the module is meant, not the average over the cell library.

Of course, the assumed distribution of substrate contacts will in general not match the actual distribution. While this may introduce a significant error in estimating the local fluctuations of substrate voltage within the aggressor, it shouldn't significantly impact the estimation of noise coupling to the analog victims outside the digital block (which is the actual goal of this model). The analysis of an error introduced by this assumption is presented later in this chapter.

Applying the above approximation has enabled creating the substrate mesh with its ports to the outside world defined before the exact layout details are known. The number of these ports can further be significantly reduced if an equal ground bounce approximation is applied – if ground bounce can be considered to be approximately equal for the entire digital block (which is the case if the package parasitics are dominant compared to on-chip parasitics), all the substrate contacts inside the block can be shorted.

This approximation basically reduces every digital aggressor to a single port. If a conductive backplane is present, it remains represented as a single port as well. The remaining ports belong to the analog victim, which can be assumed to have a separate biasing, in order to avoid direct noise coupling via the PDN. Note that, while the actual amount of coupling depends on the number of substrate contacts within the victim, the ratio between the amounts of coupling for different aggressor floorplans remains almost unchanged regardless the number of contacts. Since the main purpose of the model is the comparison of different aggressor floorplanning strategies, it's enough to keep just the minimum number of contacts needed to represent the victim. For the purpose of developing a general substrate noise reduction methodology, independent on the victim type and on its layout, the representation of the victim should be symmetric. Thus, as the simplest possible solution, the victim is represented by a single contact.

Of course, in case that more information on the type of the victim is available, i.e. if a methodology adjusted to a specific victim type should be developed, some more complex representation of the victim can be used. If the victim is modeled to contain more substrate contacts, and if the on-chip parasitics in the substrate biasing line of the victim are negligible to package parasitics, an equal ground bounce approximation can be applied for them as well, i.e. they can be shorted to a single node, thus reducing the victim to a single port in this case as well.

The number of ports is now reduced to one per each aggressor, one for the victim, and one for the conductive backplane (if present). However, the total number of nodes in this initial impedance network remains huge. Since nodes in the interior of the substrate are not of interest, this network should be reduced so that all the nodes except the ports are eliminated.

As a preparation for this reduction process, a modified nodal analysis (MNA) system of equations is formed, describing the initial impedance matrix of all the nodes obtained by substrate discretization. The numeration of nodes starts from one of the edges of the top layer, and the numbers increase as shown in Fig. 3.4.a. The resulting matrix is symmetric and very sparse, with sparsity pattern as shown in an example in Fig. 3.4.b.

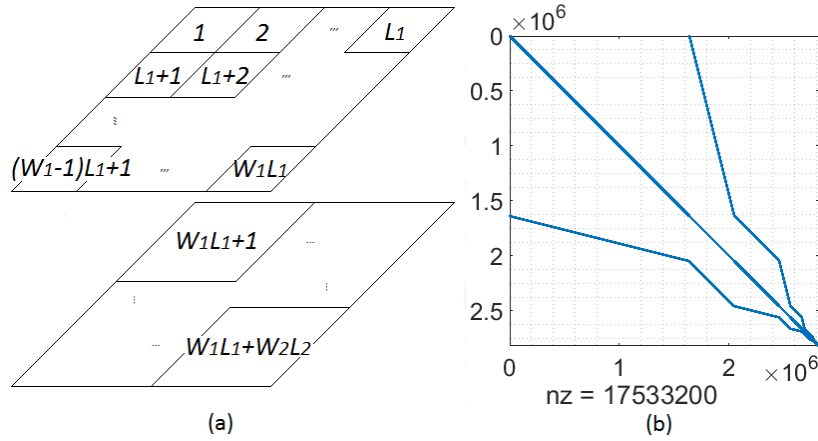


Figure 3.4 – Node numeration and sparsity pattern for the initial impedance matrix (adapted from [P1])

After forming the initial matrix, the reduction can be performed in a way similar to the one described in [18]. Let the substrate matrix have n_p ports P_1, P_2, \dots, P_{n_p} , let the port P_k be connected to a unit voltage generator, while the remaining ports are connected to the ground. In this case, the current flowing through the port $P_i, i \neq k$, will be equal to the equivalent admittance between the port P_k and the port P_i . This way, the admittances $Y_{ki}, i \neq k$, i.e. the impedances $Z_{ki}, i \neq k$ can be calculated. By repeating the procedure (connecting the unit generator to a port while the remaining ports are connected to the ground and calculating the currents through the remaining ports) for every port, all the impedances of the reduced substrate network, containing only the ports as its nodes, are determined.

In each of the iterations of the reduction process, the system of equations to be solved has the form:

$$\mathbf{M} \mathbf{V} = \mathbf{I} \quad (3.1)$$

where \mathbf{M} is the admittance matrix, \mathbf{V} is the array of node voltages, and \mathbf{I} is the array of currents flowing from the nodes. The entry in the voltage array \mathbf{V} corresponding to the port to which the unit generator is connected has a value of 1, the entries corresponding to the remaining ports have a value of 0, while the remaining entries of this array are unknowns. The elements of the current array \mathbf{I} corresponding to the port nodes have unknown values, while the elements corresponding to the remaining nodes are equal to zero. The admittance matrix \mathbf{M} can be represented as following:

$$\begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_{W_1 L_1} & \mathbf{0} \\ \mathbf{M}_{W_1 L_1}^T & \mathbf{M}_x & \\ \mathbf{0} & & \end{bmatrix} \quad (3.2)$$

where \mathbf{M}_0 is a square submatrix corresponding to surface nodes and their mutual connections, and $\mathbf{M}_{W_1L_1}$ and $\mathbf{M}_{W_1L_1}^T$ are the submatrices corresponding to connections between surface nodes and the first layer beneath. Each of the surface nodes is connected only to one of the nodes in the layer beneath, so consequently each of the nonzero elements of the submatrices $\mathbf{M}_{W_1L_1}$ and $\mathbf{M}_{W_1L_1}^T$ is equal to the admittance between a node in the surface layer and a node in the layer beneath.

Let all the columns of the matrix \mathbf{M} which are multiplied by zeroes from \mathbf{V} in (3.1) be removed, and let all the zero elements of \mathbf{V} be removed as well. Further on, let the columns of the matrix \mathbf{M} which are multiplied by ones from \mathbf{V} in (3.1) be moved to the right side of the equation, thus turning the array \mathbf{I} into a new array \mathbf{B}_{temp} , and let the elements of \mathbf{V} equal to one be removed as well. As a result, the initial matrix \mathbf{M} will be turned into a new matrix $\mathbf{M}_{red,temp}$, the initial voltage array \mathbf{V} will be turned into a new voltage array \mathbf{V}_{red} , and the equation (3.1) will be turned into (3.3), which will still hold:

$$\mathbf{M}_{red,temp} \mathbf{V}_{red} = \mathbf{B}_{temp} \quad (3.3)$$

Note that the array \mathbf{V}_{red} contains only the unknown variables. Further on, let all the rows in the system (3.3) corresponding to the port nodes (i.e. to unknown elements in \mathbf{I} , now in \mathbf{B}_{temp}) be removed. This way, the matrix $\mathbf{M}_{red,temp}$ is reduced to a new matrix \mathbf{M}_{red} , and the array \mathbf{B}_{temp} is reduced to a new array \mathbf{B} . The resulting equation is:

$$\mathbf{M}_{red} \mathbf{V}_{red} = \mathbf{B} \quad (3.4)$$

with all the elements of \mathbf{M}_{red} and \mathbf{B} known, and all the elements of \mathbf{V}_{red} being unknown variables. Note that the array \mathbf{B} could have been obtained also directly from (3.1), by moving the products of \mathbf{V} array with the elements of $\mathbf{M}_{W_1L_1}^T$ to the right side of the equation and by removing the elements of \mathbf{I} containing unknown variables.

Just like the initial matrix \mathbf{M} , the matrix \mathbf{M}_{red} is also symmetric and sparse. This enables solving (3.4) by some of the iterative methods for solving sparse systems of linear equations, readily available as built-in functions in MATLAB. An example of such functions is **bicg** for biconjugate gradients method [197].

After solving (3.4) for \mathbf{V}_{red} , all the elements of the initial voltage array \mathbf{V} are known. Let \mathbf{M}_T be a matrix consisting of the rows of \mathbf{M} corresponding to the unknown elements of the initial current array \mathbf{I} , and let \mathbf{I}_T be an array consisting of the unknown elements of \mathbf{I} . The unknown currents can now be directly calculated as:

$$\mathbf{I}_T = \mathbf{M}_T \mathbf{V} \quad (3.5)$$

Finally, as already explained, from the currents corresponding to the ports connected to the ground, the admittances of the equivalent reduced substrate network are determined. The procedure is repeated until all the necessary impedances are determined. For n_p ports, the procedure should be repeated $n_p - 1$ times. In other words, for a system with n_A aggressors, one victim and a conductive backplane, the procedure has to be repeated $n_A + 1$ times.

As an exemplary structure to demonstrate the properties of the substrate mesh reduction procedure, a 1mm x 1mm x 300 μ m die is taken. The substrate is discretized as shown in Table 3-1, with the resulting mesh containing 2184400 nodes. The first example analyzed contains two digital aggressors and one victim. Each of the aggressors consists of 100 inverters, with each of the inverters

having one substrate contact. The initial network was formed and reduced both as a purely resistive mesh and as an RC mesh at the frequency of 100 MHz; the execution times of the reduction algorithm with MATLAB on a personal computer was 1 h 19 min and 2 h 46 min, respectively. The impedance network resulting from the reduction in each of those two cases is shown in Table 3-II. By adding the voltage generators modeling the ground bounce (as it will be shown in the next section) and the package parasitics, the final model is obtained, as shown in Fig. 3.5.

The validity limit of the model obtained from the R-extraction is analyzed in Section 3.3. Note that the model obtained from the RC-extraction also has its validity limit – at very high frequencies, NWell impedance becomes comparable to bulk impedance, so the assumption about direct coupling being the only significant noise source would not be valid anymore. In this case, additional capacitances modeling the N Wells have to be introduced into the model as well.

Table 3-I – Substrate discretization [P1]

Step (μm)	0.78	1.56	3.12	6.24	12.48	24.96	49.92
Number of layers	1	1	1	3	6	4	2
Nodes per layer	1280x1280	640x640	320x320	160x160	80x80	40x40	20x20

Table 3-II – Extraction results for the exemplary structure [P1]

	Impedance ($\text{M}\Omega$)	
	Extraction from R-mesh	Extraction from RC-mesh
Z_{11}	0.7875	$0.7866 - j \cdot 0.0260$
Z_{12}	7.6649	$7.6566 - j \cdot 0.2533$
Z_{13}	2.7826	$2.7796 - j \cdot 0.0920$
Z_{22}	0.3354	$0.3350 - j \cdot 0.0111$
Z_{23}	0.5218	$0.5212 - j \cdot 0.0172$
Z_{33}	0.2133	$0.2131 - j \cdot 0.0071$

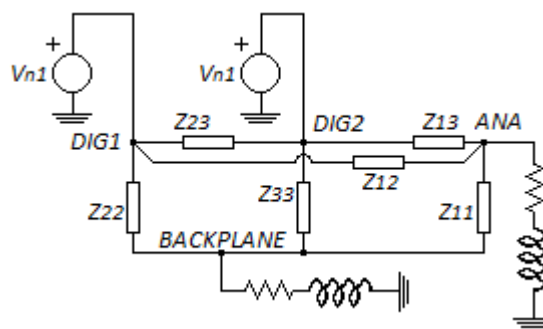


Figure 3.5 – A schematic of substrate coupling model (adapted from [P1])

The error introduced by the assumption of regularly positioned contacts is analyzed by comparing the extraction results obtained with the assumed and with modified contact distribution. Two cases of modified contact distribution are generated – case 1 (shown in Fig. 3.6.a), with contacts randomly displaced from the assumed positions, at the distance up to half the distance between the contacts, and case 2 (shown in Fig. 3.6.b), with each contact being moved for a distance equal to the maximum range from case 1, in the opposite direction for the two aggressors. Case 1 should model the realistic contact distribution, while case 2 should model the worst case error in assumption of contact positions. In both cases, the extraction is performed from the R-mesh.

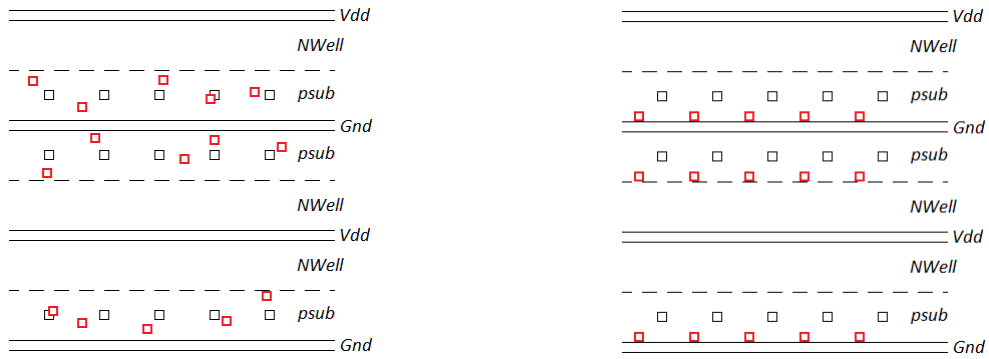


Figure 3.6 – Assumed (in black) and modified (in red) contact positions for case 1 (a) and case 2 (b)

The substrate impedances resulting from the extractions in both cases are shown in Table 3-III, together with the corresponding errors introduced by approximating the modified contact positions with the assumed regular contact distribution. As it can be seen, the maximum error introduced by this approximation was 1.5 %, which makes the approximation of regular contact positions valid for a pre-layout estimation.

Table 3-III – Results of extraction with modified contact distributions [P1]

	Case 1		Case 2	
	Resistance (M Ω)	Relative Error	Resistance (M Ω)	Relative Error
Z_{11}	0.7873	0.03 %	0.7882	0.09 %
Z_{12}	7.6424	0.3 %	7.7165	0.7 %
Z_{13}	2.7882	0.2 %	2.7655	0.6 %
Z_{22}	0.3347	0.2 %	0.3368	0.4 %
Z_{23}	0.5240	0.4 %	0.5298	1.5 %
Z_{33}	0.2138	0.2 %	0.2126	0.3 %

Finally, the dependence of extraction time on the parameters of the structure for which the substrate model is being extracted is examined.

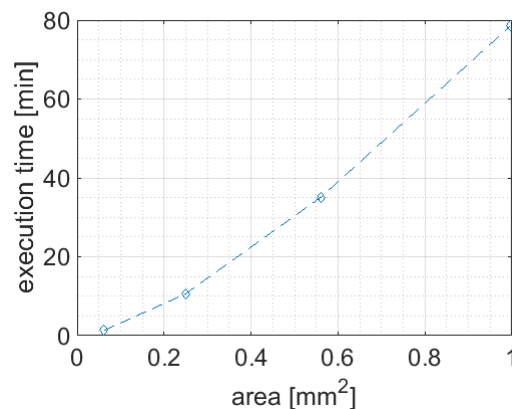


Figure 3.7 – Extraction time as a function of die area (extraction from R-mesh) (adapted from [P1])

An extraction of a system with one aggressor of constant size and one victim is repeated for various chip sizes, in order to determine the dependence of extraction time on the chip area. This dependence is shown in Fig. 3.7. Extraction time increases approximately linearly with the chip area, and for larger chips, it can become impractically long. Note that the chip area for which the model

can be extracted is also limited by the available memory resources – extracting for larger areas would require more resources. Improving the discretization strategy by applying a variable discretization step also in the horizontal direction (e.g. discretizing with a coarser step in the surface areas which are not in vicinity of contacts) might help avoiding this problem to some extent.

In order to determine the dependence of extraction time on the size of an aggressor block, the extraction has also been repeated for various aggressor sizes, for a system with one aggressor and one victim, and with a fixed die area of 0.5 mm x 0.5 mm. From the resulting graph, shown in Fig. 3.8.a, it can be seen that the dependence of the extraction time on the size of the aggressor is quite weak. Slight decrease in the extraction time with an increased size of an aggressor is due to a smaller size of the \mathbf{M}_{red} matrix for bigger aggressors. The dependence of extraction time on the number of aggressors has also been examined, by repeating the procedure for 2 to 10 equally sized aggressors, on a fixed die area of 0.5 mm x 0.5 mm. As shown in Fig. 3.8.b, the extraction time increases with the number of aggressors as expected, since for each new aggressor the extraction procedure has to be repeated once more. Note that, if the equal ground bounce approximation wasn't applied, each of the substrate contacts would have to be treated as an independent aggressor, which would result in an extremely long extraction time.

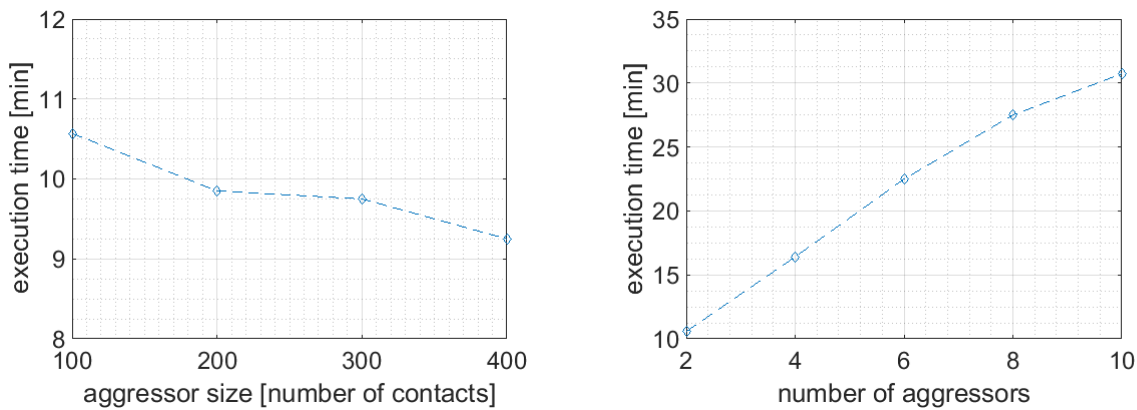


Figure 3.8 – Extraction time as a function of (a) aggressor size, and (b) number of aggressors (extraction from R-mesh) (adapted from [P1])

The output of this extraction procedure is a reduced substrate mesh, as shown in an example in Fig. 3.5. The final model is completed by adding the biasing parasitics for victim and backplane contacts, as well as the voltage generators modeling the ground bounce for digital aggressors. In the next section, modeling of ground bounce in frequency domain is explained.

3.2. A frequency domain model for ground bounce in synchronous and GALS systems

The analysis in this work is mostly limited to systems which are small enough so that package-parasitics are dominant compared to on-chip parasitics, i.e. to systems for which the equal ground bounce approximation can be applied. For such systems, there's no need for modeling the PDN as a distributed parasitics network – a lumped model of package parasitics, such as the one used in [17] for time-domain analysis of ground bounce, is sufficient. Such a lumped model is shown in the

Fig. 3.9, with R_p , L_p and C_p modeling the PDN (i.e. mostly package) parasitics, C_d representing the decoupling capacitance (including both the intentional decoupling capacitors and the capacitance of non-switching circuitry), I representing the switching current, and V_{dd} representing the steady off-chip supply voltage.

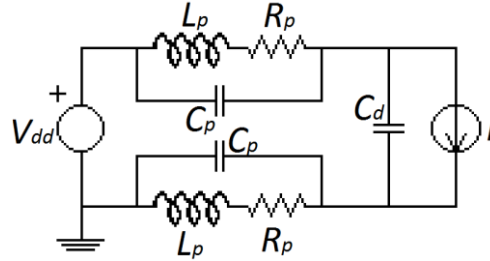


Figure 3.9 – A lumped PDN model (adapted from [P2])

The dependence of ground bounce voltage V_{gb} on switching current I can be directly determined from the schematic in Fig. 3.9:

$$H(s) = \frac{V_{gb}(s)}{I(s)} = \frac{R_p + sL_p}{1 + sR_p(C_p + 2C_d) + s^2L_p(C_p + 2C_d)} \quad (3.6)$$

The shape of the transfer function depends on the polynomial in the denominator. In case that this polynomial has two conjugate-complex zeros, i.e. in case that:

$$R_p^2(C_p + 2C_d)/(4L_p) < 1 \quad (3.7)$$

the transfer function is underdamped, with the resonant frequency:

$$\omega_r = \frac{1}{\sqrt{L_p(C_p + 2C_d)}} \quad (3.8)$$

and the resonant peak magnitude:

$$|H(j\omega_r)| = \sqrt{\frac{L_p}{C_p + 2C_d} \left(1 + \frac{L_p}{R_p^2(C_p + 2C_d)} \right)} \quad (3.9)$$

For any given package, the values R_p , C_p and L_p are fixed. Thus, whether or not the condition (3.7) is fulfilled, depends on the only remaining parameter – the decoupling capacitance C_d . Common order of magnitude ranges of package parasitics, according to specifications of some large IC producers [198][199], are $0.01 \div 1 \Omega$ for R_p , $0.1 \div 10$ nH for L_p and $0.01 \div 0.1$ pF for C_p . Based on that, the boundary value of $C_p + 2C_d$ below which the transfer function is underdamped (i.e. for which there's a resonant peak) is somewhere within the order of magnitude range of 100 pF \div 1 mF. Note that this value is much larger than the expected value for C_p , so the sum $C_p + 2C_d$ can be approximated as $2C_d$. Even in the most optimistic case, the value of C_d has to be at least 100 pF in order to completely dump the resonant peak, which would require a rather large area. For example, a MIM capacitor of this size in IHP's 250nm BiCMOS SiGe:C technology [25] would occupy 0.1 mm², which can already be too big for the applications where the available area is small. Due to this, it is assumed in this work that the PDN transfer function is underdamped, i.e. that it exhibits a resonant peak. An example of such a transfer function is shown in Fig. 3.10.

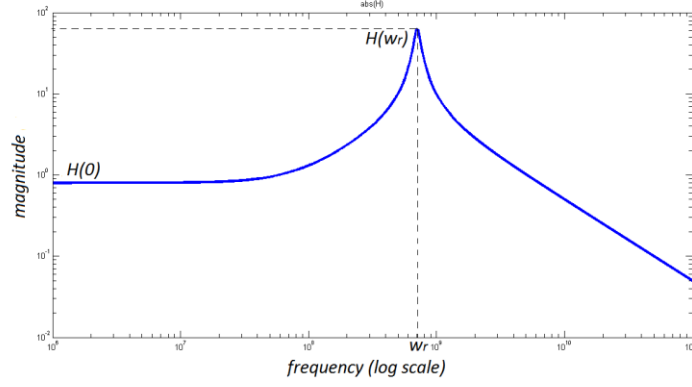


Figure 3.10 – The magnitude of an underdamped PDN transfer function

In this simple lumped model, the ground bounce spectrum is a product of the switching current spectrum and the PDN transfer function. As already mentioned in the introduction of this chapter, for modeling the switching current, the model described in [16] was used, where the average switching current of a synchronous system is represented as a periodic triangular pulse waveform. An example of one period of such a waveform is shown in Fig. 3.11.

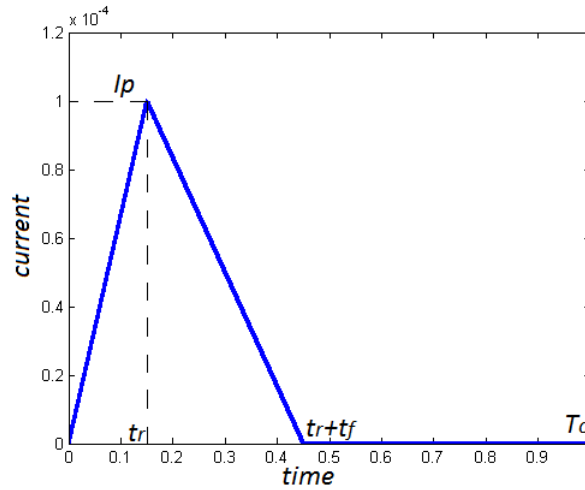


Figure 3.11 – Switching current modeled as a periodic triangular pulse (adapted from [P2])

The frequency domain representation of this signal is also provided in [16] – the n^{th} component of its Fourier series, modeling the n^{th} switching current harmonic (at the frequency nf_0), is:

$$F_n = \frac{I_p}{j2\pi n} \left(\text{sinc}(\pi n f_0 t_r) - \text{sinc}(\pi n f_0 t_f) e^{-j\pi n f_0 (t_r + t_f)} \right) e^{-j\pi n f_0 t_r} \quad (3.10)$$

where

$$\text{sinc}(x) = \sin(x)/x \quad (3.11)$$

From (3.6) and (3.10), the n^{th} harmonic of the ground bounce spectrum can be determined as:

$$F_n = \frac{R_p + j\omega L_p}{1 + j\omega R_p (C_p + 2C_d) - \omega^2 L_p (C_p + 2C_d)} \frac{I_p}{j2\pi n} \cdot \left(\text{sinc}(\pi n f_0 t_r) - \text{sinc}(\pi n f_0 t_f) e^{-j\pi n f_0 (t_r + t_f)} \right) e^{-j\pi n f_0 t_r} \quad (3.12)$$

Note that (3.10) and (3.12) represent the switching current spectrum and the ground bounce spectrum of a synchronous system, having spectral peaks at clock frequency harmonics. This can also be used for representing the switching current spectrum and the ground bounce spectrum of a single LSM within a GALS system. The switching current spectrum for an entire GALS system can then be determined as the sum of switching current spectra of all the LSMs within the system [16].

Consequently, if an initially synchronous system is galsified, so that the new GALS system consists of M LSMs, each of the switching current spectral peaks of at clock harmonics of the initial synchronous system is replaced by M spectral peaks, one per each LSM. If the LSMs are plesiochronously clocked, the new spectral peaks will be at frequencies close to the initial synchronous clock harmonic nf_0 . [16]

In a system for which the equal ground bounce approximation can be applied, the above statements are valid for the ground bounce spectrum of a GALS system as well. The simple model from Fig. 3.9 has to be modified accordingly – the current generator I has to be replaced with M current generators, each of them representing one of the LSMs of the new GALS system. The ground bounce spectrum of a GALS system is equal to the sum of ground bounce spectra of its separate LSMs; and each ground bounce spectral peak of an initial synchronous system (at the frequency nf_0 , where f_0 is the synchronous clock frequency, and n is the order of the harmonic) will turn into M ground bounce spectral peaks (at the frequencies nf_m , where f_m , $m = 1, \dots, M$, are the LSM clock frequencies, and n is the order of the harmonic) when this system is partitioned into M LSMs and galsified.

However, there's an important difference between the spectrum of ground bounce and the spectrum of switching current – while in case of switching current the dominant spectral components are the fundamental and the low harmonics, in case of ground bounce the dominant spectral components are determined by the PDN transfer function. In case of an underdamped transfer function, the dominant components of ground bounce spectrum are grouped around the resonant peak.

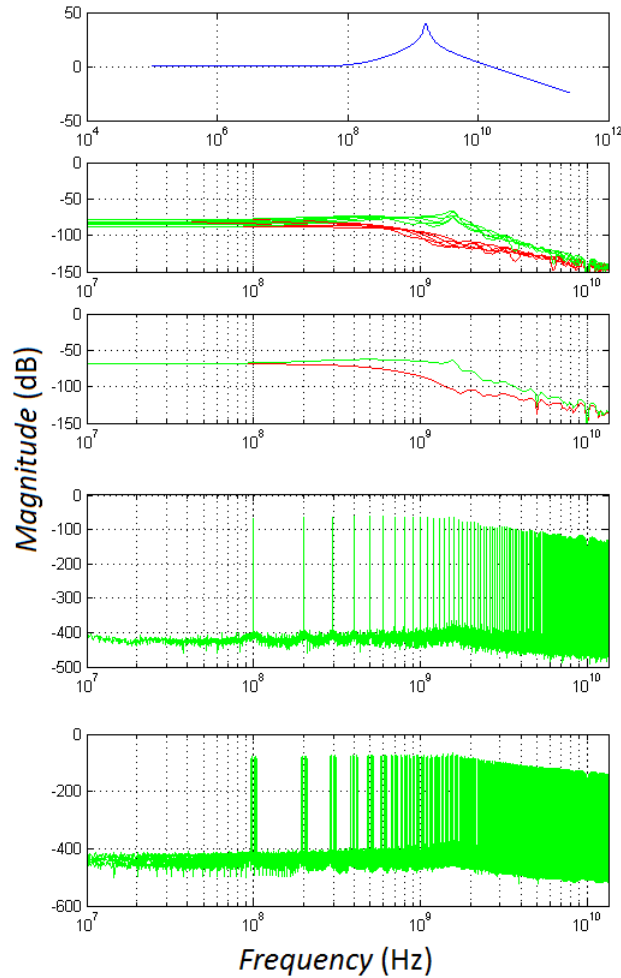


Figure 3.12 – An example of the ground bounce voltage spectrum (a) PDN transfer function (b) spectrum envelope for a spectrum of switching current (red) and ground bounce (green) for a GALS system with 5 LSMs (with envelopes drawn separately for separate LSMs) (c) spectrum envelope for a spectrum of switching current (red) and ground bounce (green) for a synchronous system (d) ground bounce spectrum for a synchronous system (e) ground bounce spectrum for a GALS system (adapted from [P2])

This is shown in an example in Fig. 3.12, both for the case of a synchronous system and for the case of a GALS system. The values of package parasitics chosen for this example are: $R_p = 1 \Omega$, $L_p = 1 \text{ nH}$ and $C_p = 0.1 \text{ pF}$, while the value of the total decoupling capacitance is $C_d = 5 \text{ pF}$. The synchronous system in this example has a clock frequency of 100 MHz, while the GALS system consists of five plesiochronously clocked LSMs, with clock frequencies close to the original synchronous clock frequency.

For the synchronous example system described above, the dominant components of the ground bounce spectrum (defined as the part of the spectrum containing 95 % of the entire spectral power) are shown in Fig. 3.13 (marked in red), for several values of clock frequency and decoupling capacitance. In all the cases, the dominant part of the spectrum can be found around and below the resonant frequency. For a larger value of C_d , the resonant frequency gets shifted toward the lower frequencies according to (3.8). Consequently, as it can be seen in Fig. 3.13, the dominant part of the spectrum also gets shifted towards the lower frequencies for a larger value of C_d .

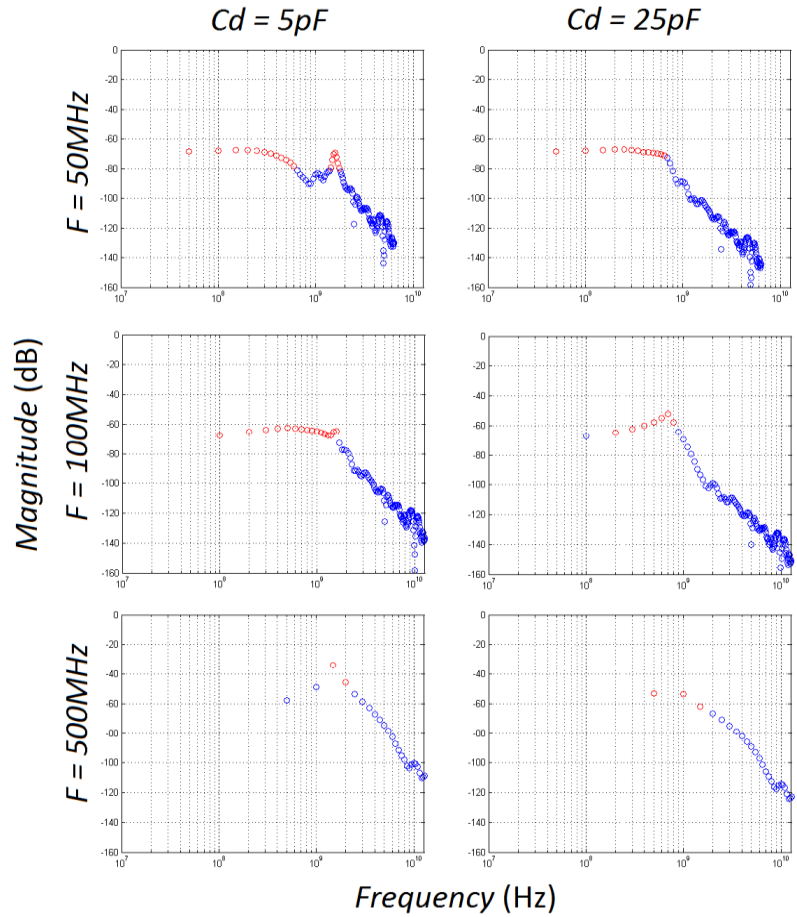


Figure 3.13 – Dominant components for the synchronous system (i.e. the part of the spectrum containing 95 % of total power) are marked in red (adapted from [P2])

In the following section, by taking into account the spectrum of ground bounce as modeled in this section, the resistive approximation in substrate noise propagation model from the previous section (i.e. approximation of the initial substrate mesh as purely resistive) is analyzed.

3.3. An analysis of applicability of resistive approximation in substrate noise propagation model

In order to estimate the error introduced by a purely resistive approximation of the substrate, a structure consisting of one aggressor and one victim is used. Also, in order to keep the analysis as simple as possible, a die without a backplane has been considered. Under these assumptions, the model from Section 3.1 reduces to a schematic shown in Fig. 3.14. Since there are only two ports, the substrate is represented by the extracted impedance Z_{sub} between the aggressor (A) and the victim (V). The substrate biasing network at the victim side is represented by the impedance Z_p , which can be modeled as a lumped RLC network.

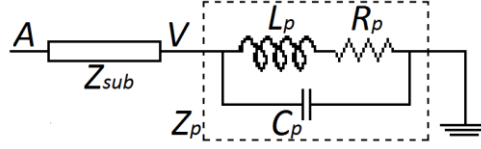


Figure 3.14 – Substrate coupling from aggressor to the victim, simplified (no conductive backside) (adapted from [P2])

From this simple model, the transfer function from aggressor to victim point can be determined, representing the dependence of the substrate noise at victim's side on the ground bounce in the ground network of the aggressor:

$$\underline{H_{AV}}(\omega) = \frac{V_V(\omega)}{V_A(\omega)} = \frac{1}{1 + \frac{Z_{sub}(\omega)}{Z_p(\omega)} e^{j\varphi_{sub}(\omega)}} \quad (3.13)$$

where $Z_{sub}(\omega)$ is the magnitude of extracted substrate impedance, while $\varphi_{sub}(\omega)$ is its phase.

By applying the resistive substrate approximation, the expression (3.13) turns into:

$$\hat{H}_{AV}(\omega) = \frac{1}{1 + \frac{R_{sub}}{Z_p(\omega)}} \quad (3.14)$$

The relative error introduced by replacing the transfer function $\underline{H_{AV}}$ with its resistive approximation \hat{H}_{AV} at frequency ω is:

$$\Delta H_{AV}(\omega) = \left| \frac{\hat{H}_{AV}(\omega) - \underline{H_{AV}}(\omega)}{\underline{H_{AV}}(\omega)} \right| \quad (3.15)$$

which can further be expressed as:

$$\Delta H_{AV}(\omega) = \left| \frac{1}{1 + \frac{R_{sub}}{Z_p(\omega)}} \right| \left| z_{sub}(\omega) e^{j\varphi_{sub}(\omega)} - 1 \right| \quad (3.16)$$

with $z_{sub}(\omega)$ being the scaled magnitude of substrate impedance:

$$z_{sub}(\omega) = \frac{Z_{sub}(\omega)}{R_{sub}} \quad (3.17)$$

As already mentioned in Chapter 2, lightly doped substrates are mostly homogenous [23]. Due to the homogeneity of resistivity and dielectric permittivity throughout the substrate, all the RC branches in the initial substrate mesh (as shown in Fig. 2.8) have a proportional impedance:

$$Z_{ij}(\omega) = \frac{R_{ij}}{1 + j\omega R_{ij} L_{ij}} = \frac{R_{ij}}{1 + j\omega \rho \varepsilon} \quad (3.18)$$

In other words, the impedances of all the RC branches have the same phase:

$$\varphi_{ij}(\omega) = \pi - \arctan(\omega \rho \varepsilon) \quad (3.19)$$

while the magnitude depends on the discretization step for the specific branch. The scaled magnitude, however, remains the same for each of the branches:

$$z_{ij}(\omega) = \frac{Z_{ij}(\omega)}{R_{ij}} = \frac{1}{1 + \omega^2 \rho^2 \varepsilon^2} \quad (3.20)$$

As a consequence, the impedance \underline{Z}_{sub} resulting from the reduction of the initial RC mesh also has the same phases and the same scaled magnitude as the initial branches:

$$\underline{Z}_{sub}(\omega) = \frac{R_{sub}}{1 + j\omega\rho\varepsilon} \quad (3.21)$$

$$z_{sub}(\omega) = \frac{1}{1 + \omega^2 \rho^2 \varepsilon^2} \quad (3.22)$$

$$\varphi_{sub}(\omega) = \pi - \arctan(\omega\rho\varepsilon) \quad (3.23)$$

From (3.22) and (3.23) it can be seen that the phase and the scaled magnitude of the equivalent substrate impedance \underline{Z}_{sub} depend only on substrate parameters ρ and ε (in other words, only on technology), not on the parameters of the aggressor or the victim, such as their size and mutual position. Consequently, the second factor from (3.16), $|z_{sub}(\omega)e^{j\varphi_{sub}(\omega)} - 1|$, is also dependent only on technology. For IHP's 250nm and 130nm BiCMOS SiGe:C technologies with 50 Ωcm resistivity substrate [25], this factor is shown in Fig. 3.15.

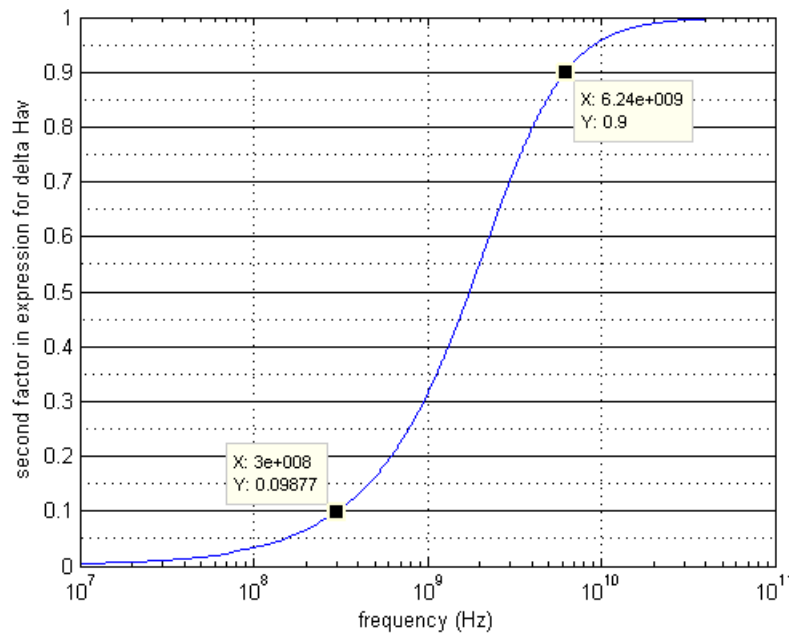


Figure 3.15 – The second factor of (3.16) for the IHP's BiCMOS SiGe:C technologies (adapted from [P2])

It can be seen from Fig. 3.15 that the second factor of (3.16) always has a value lower than 1, which means that it contributes to the reduction of error introduced by the resistive substrate approximation. The factor has a lower value at lower frequencies, which means that the suppression of error is more pronounced at lower frequencies. Specifically for the IHP's technology, the factor has a value lower than 0.1 for frequencies below 300 MHz, while its value is higher than 0.9 for frequencies above 6.24 GHz.

On the other hand, the first factor of (3.16), $\left|1/(1 + \underline{Z_p}(\omega)/R_{sub})\right|$, is dependent on the parasitic impedance of the PDN for biasing the substrate at the victim side, $\underline{Z_p}(\omega)$, and on the extracted substrate resistance R_{sub} , which is dependent on the properties of the aggressor and the victim. The impedance $\underline{Z_p}(\omega)$, its resonant frequency and its magnitude are expressed by equations (3.24), (3.25) and (3.26), respectively:

$$\underline{Z_p}(\omega) = \frac{R_p + j\omega L_p}{1 + j\omega R_p C_p - \omega^2 L_p C_p} \quad (3.24)$$

$$\omega_{rp} = \frac{1}{\sqrt{L_p C_p}} \quad (3.25)$$

$$|Z_p(j\omega_{rp})| = \sqrt{\frac{L_p}{C_p} \left(1 + \frac{L_p}{R_p^2 C_p}\right)} \quad (3.26)$$

From (3.26), and according to the ranges for the values of the package parasitics from [198] and [199], the expected order of magnitude of the resonant peak of $\underline{Z_p}(\omega)$ is within the range $1 \text{ k}\Omega \div 100 \text{ M}\Omega$. By comparing (3.25) to (3.8), and knowing that (as previously stated) $C_d \gg C_p$, it can be concluded that the resonant frequency ω_{rp} is much higher than the resonant frequency of the PDN at the aggressor's side, ω_r . Since the dominant part of the ground bounce spectrum is around and below the resonant frequency ω_r , the important components of substrate noise spectrum stay far below the resonance of $\underline{Z_p}(\omega)$. At these frequencies the magnitude of $\underline{Z_p}(\omega)$ is much lower than its resonant peak. For example, according to the ranges from [198] and [199], the expected value for the order of magnitude of $|\underline{Z_p}(\omega)|$ is within the range $10 \Omega \div 10 \text{ k}\Omega$ at the frequency $\omega = 0.1\omega_{rp}$, and within the range $0.1 \Omega \div 100 \Omega$ at the frequency $\omega = 0.01\omega_{rp}$.

The value of R_{sub} , on the other hand, depends strongly on the parameters of the aggressor and the victim, primarily on their size and mutual position. According to the extractions from the model described in Section 3.1, for a lightly doped substrate with $50 \Omega\text{cm}$ substrate resistivity, this value is usually within an order of magnitude range from $\text{k}\Omega$ to tens of $\text{M}\Omega$. Thus, at the frequencies of the dominant part of the spectrum, the first factor of the expression (3.16), $1/(1 + \underline{Z_p}(\omega)/R_{sub})$, has a value close to 1. Consequently, the expression (3.16) can be approximated with its second factor, which is only technology-dependent:

$$\Delta H_{AV}(\omega) \approx |Z_{sub}(\omega)e^{j\varphi_{sub}(\omega)} - 1| \quad (3.26)$$

Based on this, it can be concluded that the validity of the resistive substrate approximation depends on the relation between the frequencies of the dominant part of the ground bounce spectrum, and the technology-dependent error function (3.26), an example of which is shown in Fig. 3.15. If at the frequencies of the dominant part of the spectrum function (3.26) has a low value, the approximation is justified.

As already shown in the end of the previous section (see Fig. 3.13), the frequency range of the dominant part of the spectrum is determined by the decoupling capacitance. Thus, an error in total substrate noise power caused by the resistive substrate approximation will also be dependent on the decoupling capacitance, and not on the value of the extracted substrate resistance. In order to demonstrate this, substrate noise power was calculated according to the model from Fig. 3.14 for the same example synchronous system for which the ground bounce spectrum is presented in Fig. 3.13.

The assumed clock frequency is 100 MHz. The substrate noise spectrum was calculated for two values of the decoupling capacitance, $C_d = 5$ pF and $C_d = 25$ pF, and for three different values of R_{sub} . The substrate noise spectrum envelope, obtained both with applying the resistive substrate approximation (red) and without applying the resistive substrate approximation (blue), is shown in Fig. 3.16 and Fig. 3.17, and the relative error caused by the resistive substrate approximation for all the six cases is presented in Table 3-IV.

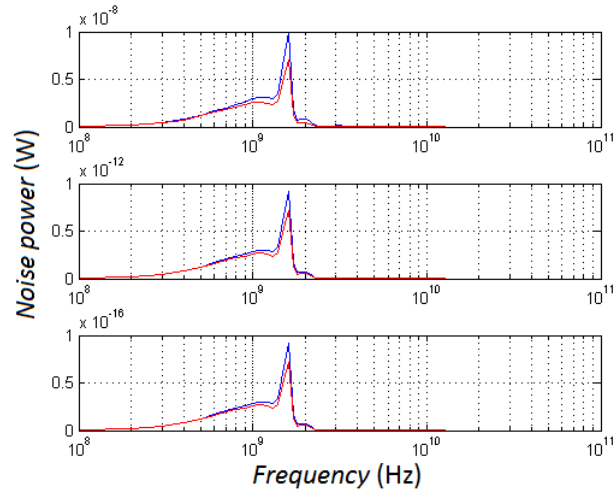


Figure 3.16 – Power spectrum of substrate noise for $C_d = 5$ pF and substrate resistivities $R_{sub} = 100 \Omega$ (a), $R_{sub} = 10 \text{ k}\Omega$ (b) and $R_{sub} = 1 \text{ M}\Omega$ (c) (adapted from [P2])

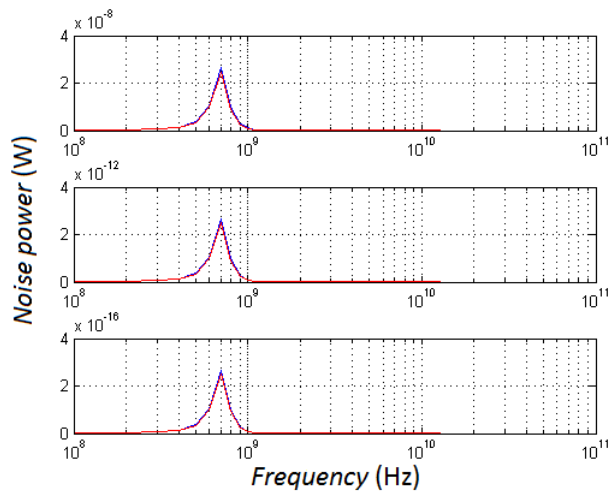


Figure 3.17 – Power spectrum of substrate noise for $C_d = 25$ pF and substrate resistivities $R_{sub} = 100 \Omega$ (a), $R_{sub} = 10 \text{ k}\Omega$ (b) and $R_{sub} = 1 \text{ M}\Omega$ (c) (adapted from [P2])

Table 3-IV – Relative error in total power of substrate noise introduced by applying resistive approximation [P2]

	$R_{sub} = 100 \Omega$	$R_{sub} = 10 \text{ k}\Omega$	$R_{sub} = 1 \text{ M}\Omega$
$C_d = 5 \text{ pF}$	24.46 %	17.54 %	17.40 %
$C_d = 25 \text{ pF}$	7.16 %	5.22 %	5.20 %

As expected, the error is almost independent on the value of R_{sub} – only for the unrealistically low value of $R_{sub} = 100 \Omega$ the error becomes slightly larger. On the other hand, the decoupling capacitance strongly impacts the error. For the larger value of C_d , the resonant frequency is at around 700 MHz and the dominant part of the ground bounce spectrum is below 1 GHz, i.e. in the frequency region where the function (3.26) from Fig. 3.15 has a value much lower than 1. Consequently, the error of the resistive substrate approximation in this case is low (below 10 %). For the lower value of C_d , the resonant frequency and consequently the dominant part of the ground bounce spectrum are above 1 GHz, i.e. in the frequency range where the the function (3.26) from Fig. 3.15 already has a nonnegligible value. Thus, the approximation error gets notably higher. However, even in this case, the error is small enough for the approximation to be used for the purpose of a coarse estimation and analysis.

3.4. Summary

In this chapter, the models developed for the purpose of theoretical analysis of system level methodologies for substrate noise suppression are presented, and some of their characteristics are analyzed.

First, a coarse model for SSN propagation through lightly doped substrates is presented. Since the analytical models are not applicable for the kind of analysis needed in this work, a coarse numerical model is developed. The model is based on the finite difference method, with variable discretization step in order to lower the size of the mesh. Noise sources other than the dominant one – the direct coupling of ground bounce through the substrate contacts – are neglected. An equal ground bounce approximation is applied, which is valid if package parasitics are dominant compared to the on-chip parasitics. In order to be able to make estimations without knowing the layout details, an assumption about regular positions of substrate contacts is made, which is the main novelty of the model. By comparing the results of the extraction with the assumed regular positions of substrate contacts and the results of the extraction with modified positions of substrate contacts, it was shown that this assumption introduces a very small additional error to the model, at the order of magnitude of only $\sim 1\%$. The model can be extracted both from an RC-mesh and from an R-mesh.

Then, a frequency-domain model for ground bounce is presented. This model is an extension of a model for switching current from [16]. Thanks to the equal ground bounce approximation, the PDN parasitics are reduced to the package parasitics, which are represented by a lumped RLC model. The decoupling capacitance is also included as a lumped element. The switching current was represented as a periodic triangular pulse current generator, similar to [16]. In case of a GALS system with M LSMs, a single current generator was replaced by M current generators, with each of them representing one LSM. It is shown that the dominant part of the ground bounce spectrum is concentrated around and under the resonant frequency of the PDN at the aggressor's side, which is strongly dependent on the decoupling capacitance.

Finally, following the conclusions about the dominant part of the ground bounce spectrum, the effects of the resistive substrate approximation in the SSN propagation model through lightly doped substrates, i.e. the effects of approximating the substrate mesh as being purely resistive, are analyzed. It was shown that the error introduced by this approximation depends on the technology

parameters and on the frequency range of the dominant part of the ground bounce spectrum, which is (for a fixed package) primarily defined by the decoupling capacitance. The technology parameters (substrate resistivity and dielectric permeability) set a frequency limit; if the dominant part of the spectrum is above this limit, the additional error introduced by the resistive substrate approximation is shown to be below 10 %. However, even for the dominant components above this limit, the error of the resistive substrate approximation can be satisfactory for the purpose of a coarse estimation.

By combining the coarse model for the SSN propagation through lightly doped substrates and the frequency domain model for ground bounce, the model for substrate noise coupling is formed. This model is further used for the theoretical analysis of substrate noise coupling during the development of GALS-based substrate noise reduction methodologies throughout this work.

4. Harmonic-balanced plesiochronous GALS partitioning methodology for substrate noise reduction

As stated in Chapter 2, no GALS-based methodology specifically targeting substrate noise reduction has been presented prior to this thesis. The power-balanced plesiochronous GALS partitioning [16], although not specifically aiming at substrate noise reduction, can be used for that purpose in the special case when all the analog victim modules sharing the same substrate with the digital aggressor circuitry are sensitive at low frequencies. However, if analog modules which are sensitive at higher frequencies are among the victims, this methodology can't be used. As it will be explained in Section 4.1 of this chapter, in general case, with several different analog modules sensitive in several different frequency bands, it is necessary to be able to simultaneously target substrate noise reduction at different frequency bands. In order to fulfill this requirement, a GALS methodology for substrate noise reduction called "harmonic-balanced plesiochronous GALS partitioning" [P5] has been developed and is presented in this chapter.

This methodology is developed without any assumptions about the substrate type, which makes it applicable both to MSICs on epi-type substrates and to MSICs on lightly doped substrates. However, an assumption is made that the digital subsystem (i.e. the aggressor) is small enough so that the package parasitics are dominant compared to the on-chip parasitics, i.e. that the equal ground bounce approximation can be applied. The specific properties of lightly doped substrates are taken into account for developing another methodology which will be presented in Chapter 5.

In Section 4.1, some general requirements, which a substrate noise reduction methodology should fulfill, are defined, by taking into account the substrate noise spectrum and the sensitivity of analog modules to the substrate noise. Based on these requirements, the theoretical foundations of a GALS-based substrate noise reduction methodology are set, including the choice of the clocking scheme and the mathematical analysis of the optimum partitioning. This is presented in Section 4.2. The methodology has an important property that, although developed for substrate noise reduction, it can also be applicable for reduction of the remaining two forms of SSN – switching current and ground bounce.

An algorithm for application of the theoretically derived partitioning scheme to a real system was further developed and numerically evaluated, as presented in Section 4.3. Further, as described in Section 4.4, the algorithm is incorporated into a GALS design flow, and applied on a real design example.

In Section 4.5, the methodology has been compared to other system-level methodologies for substrate noise reduction, and its advantages and drawbacks have been analyzed.

4.1. General requirements for a substrate noise reduction methodology in MSICs

As already mentioned, the methodology presented in this chapter is developed under an assumption that the equal ground bounce approximation can be applied for the digital subsystem. Consequently, the entire digital subsystem can be represented as a single aggressor. For the sake of simplicity, let's initially assume a system with a single victim as well. According to the model from Chapter 3 [P1], the assumed system can be represented with one of the equivalent schematics presented in Fig. 4.1, depending on the type of the substrate.

In both cases, such a system can be represented with the following equation in frequency domain:

$$V_{sub}(f) = H_{sub}(f) V_{gb}(f) = H_{sub}(f) H_{PDN}(f) I_{sw}(f) = H(f) I_{sw}(f) \quad (4.1)$$

where V_{sub} is the substrate voltage at the position of the victim, V_{gb} the ground bounce in the aggressor, I_{sw} the switching current in the aggressor, H_{sub} the substrate transfer function (from the ground bounce to the substrate voltage) and H_{PDN} the power delivery network transfer function (from the switching current to the ground bounce). Note that if an initially synchronous aggressor is galsified, and if the galsification overheads can be neglected, transfer functions H_{sub} , H_{PDN} and H stay constant regardless of the applied frequency scheme.

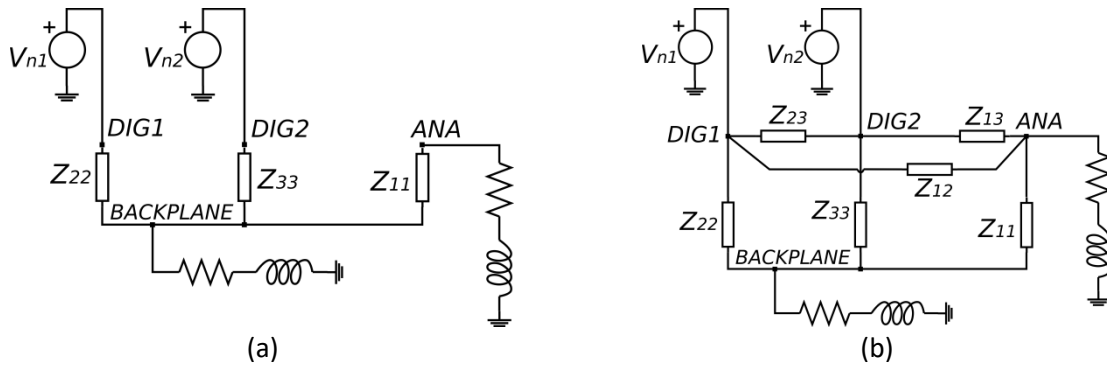


Figure 4.1 – Equivalent representation of substrate noise coupling for a system with one aggressor where the equal ground bounce approximation can be applied, in epi-type substrates (a) and in lightly doped substrates (b)

As already explained in Chapter 2, the switching current spectrum for a synchronous digital system consists of discrete peaks, corresponding to the averaged switching current waveform, and of a continuous noise floor, far below the level of the peaks, which represents the waveform fluctuations from cycle to cycle [29][30].

In a synchronous digital system there's one spectral peak per each clock harmonic. When this synchronous system is converted into a GALS system with M LSMs, each of those peaks is replaced by M lower peaks, one per each LSM [16]. As a figure of merit for switching noise reduction, a spectral peak attenuation (SPA) can be used, which is defined as the ratio between the synchronous noise peak and the highest GALS noise peak, usually expressed on the logarithmic scale [16]. For the provided switching current, SPA can be expressed as:

$$SPA_{i_{sw},n} = 20 \log \left(\frac{|I_{sw,sync}(nf_0)|}{\max\{|I_{sw,LSM_m}(nf_m)|\}} \right) \quad (4.2)$$

where f_0 is the clock frequency of the initial synchronous system, f_m is the clock frequency of the m^{th} LSM, $I_{sw,sync}(nf_0)$ is the n^{th} harmonic of switching current of the synchronous system, and $I_{sw,LSM_m}(nf_m)$ is the n^{th} harmonics of the switching current of the m^{th} LSM of the GALS system.

Since the transfer functions H_{sub} and H_{PDN} are continuous, the peaks from the switching current spectrum propagate to the ground bounce spectrum and the substrate noise spectrum, where they appear at the same frequencies as in the switching current spectrum. Consequently, just like the switching current spectrum, the ground bounce spectrum and the substrate noise spectrum also consist of discrete peaks and a continuous noise floor. However, the magnitudes of these spectral peaks are determined by the transfer functions, as shown on a simplified example of a single-resonance transfer function in Fig. 4.2, and the frequency bands containing their dominant parts will be different compared to the switching current.

In the switching current spectrum, the dominant component carrying the most power is the fundamental one. Thus, if the goal of the SSN reduction methodology is the reduction of the first form of the SSN – the switching current – the noise reduction methodology should be focused on suppressing the peak at the fundamental clock frequency. This was the case with the methodology presented in [16].

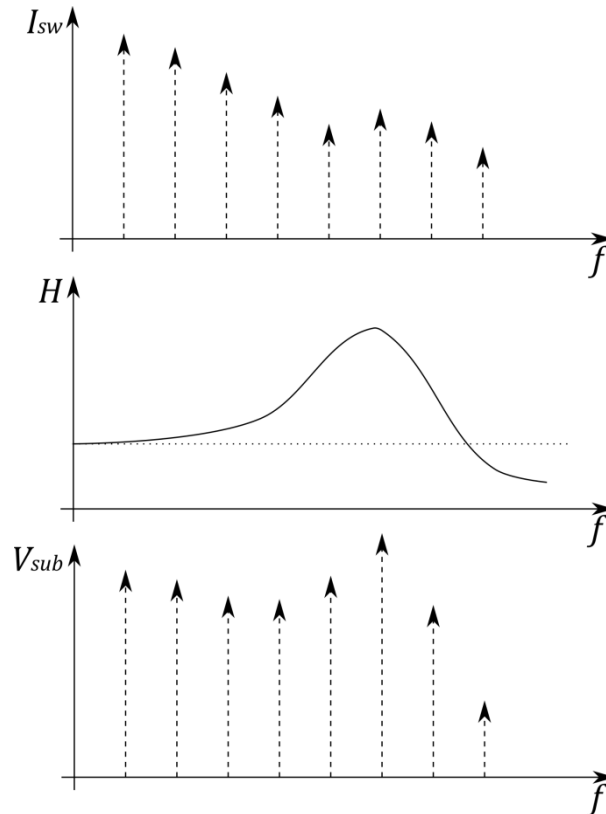


Figure 4.2 – Switching current and substrate noise have harmonic peaks at the same frequencies, but the dominant part of the spectrum is different due to PDN and substrate transfer function H . Note that this is a simplified example – H normally has several resonant peaks.

For the ground bounce spectrum, however, the dominant part of the spectrum is concentrated around the resonant frequency, as explained in Chapter 2. It is a consequence of the

transfer function $H_{PDN} = V_{gb}/I$ being underdamped, i.e. having a resonant peak. Thus, when reducing the ground bounce spectrum, the frequency band around the resonant frequency should be targeted for reduction.

Similarly, the transfer function from the switching current to the substrate noise, $H = H_{sub}H_{PDN} = V_{sub}/I$ is usually a passband function, often containing several resonant frequencies, as discussed in [30]. The poles and zeros of this function are defined by the package parasitics and by the equivalent substrate impedances, as shown in the schematics in Fig. 4.1. As a consequence, similar to the ground bounce spectrum, the dominant part of the substrate noise spectrum can be concentrated around the resonant frequencies of the transfer function H .

It can be concluded that substrate noise spectrum can have a different dominant frequency compared to the switching current spectrum. But that's not the only difference relevant for defining the noise reduction methodology. Additionally, and more importantly, the frequency band, which has to be targeted for spectral peak attenuation of substrate noise, will depend on the type of analog modules which are to be integrated on the chip.

For low noise amplifiers, which are linear modules, only the substrate noise falling into the operation frequency band can propagate from the substrate to the output, as explained in detail in Chapter 1. As LNAs which are applied in RF receivers operate at high frequencies, the substrate noise harmonics of higher order are the ones dominantly impacting their operation. In case of a narrowband LNA, the issue with the substrate noise can be avoided relatively easily by frequency planning, i.e. by taking care during the system specification that the spectral peaks of switching current (and consequently, of substrate noise) don't fall into the operational frequency band of the LNA. In other words, if the impact of substrate noise is to be avoided, the operational band of the LNA should not contain any of the clock harmonics. However, the wider the operating frequency band of the LNA, the more difficult it becomes to apply this approach, and in case of wideband amplifiers, avoiding substrate noise problems by frequency planning becomes infeasible.

On the other hand, for voltage controlled oscillators, which are nonlinear modules, low frequency noise components can propagate to the output through nonlinear intermodulation with the VCO oscillating frequency, and appear at the output of the VCO as nonlinear intermodulation spur tones, as explained in Chapter 1. As a result, the part of the substrate noise spectrum critically impacting the operation of VCOs is at lower frequencies. The most critical spectrum component in this case is at the fundamental clock frequency.

MSICs often contain various different types of analog modules, which can be sensitive to substrate noise in different (and sometimes separate) frequency bands. In such cases, different frequency bands of switching current spectrum have to be targeted for spectral peak reduction. If the critical frequency band is wide, so that it contains more substrate noise harmonic peaks, the targeted substrate noise harmonic peaks should be those being the dominant ones within the critical frequency band.

Consequently, in order to be able to effectively reduce the substrate noise in MSICs, a possibility to target several different frequency bands of substrate noise spectrum (including higher frequencies as well) is usually required. This is an important difference compared to the case where the reduction of switching current is targeted. In that case, the spectrum peak reduction of the

fundamental and low frequency harmonics were the most important. Note that this band-targeting doesn't mean that noise may only be reduced in the targeted frequency band (or bands), but rather that noise reduction should be optimized within the targeted frequency band (or bands), while in the remaining part of the spectrum noise reduction isn't required, and if it happens it doesn't need to be optimal. Depending on the specific MSIC, i.e. on the type of analog victims which it contains, the actual part of the switching current spectrum, which has to be reduced, may contain one or more frequency bands. On the other hand, sometimes the targeted part of the spectrum can simply be just the fundamental. Examples of this case are if the chip only contains a VCO, or if the only victims besides a VCO are LNAs with operating frequency bands narrow enough so that frequency planning is possible.

In the following section, a GALS-based methodology for SSN reduction is presented, providing a band-targeting option and thus also the possibility to be optimized for substrate noise reduction. This includes the choice of the most suitable clocking scheme, defining the optimal partitioning strategy through a mathematical analysis, and discussing the impact of some approximations taken during the process of formulating the methodology. As already mentioned, no assumptions about substrate type are taken, but it is assumed that package parasitics are dominant to on-chip parasitics, which makes the methodology applicable for small systems.

4.2. The theoretical foundations of a GALS-based methodology for substrate noise reduction

4.2.1. The choice of a GALS clocking scheme

Prior to determining the optimal GALS partitioning scheme, a suitable GALS clocking scheme has to be chosen, such that the substrate noise reduction methodology based on it is reliable and generally applicable. Among the various possible GALS clocking schemes, plesiochronous clocking is chosen; the main reasons for this choice are explained in this section.

Since the starting point for galsification is a synchronous system with one clock frequency, the frequencies of the LSMs generally can't be significantly increased compared to the initial synchronous frequency. Besides the critical path, in digital logic there are usually more paths with nearly critical lengths, which are distributed among the different modules of the system, and consequently also among the different LSMs once the design is galsified. Due to such path length distribution, the LSMs can usually have only a slightly higher clock frequency compared to the synchronous system. Assigning a significantly higher frequency is possible only if the LSM has much shorter critical paths compared to the critical path of the entire system, but that would be a specific case study rather than a generally applicable solution.

On the other hand, there is no timing-determined lower limit to the clock frequencies of LSMs. The lower limit, however, is set by the performance of the system. In general, when the initially synchronous system is transformed into a GALS one, the new GALS system can at most be as fast as its slowest component. Assigning a substantially lower frequency to some of the LSMs would

thus lead to performance loss of the entire system. Due to this, GALS clocking schemes requiring a large difference between the frequencies of the individual LSMs are frequently unsuitable due to the performance drop they would cause.

In a plesiochronous clocking scheme, the differences of LSM frequencies to the initial synchronous frequency are of the order of 1% or lower. This means that the processing capability, i.e. the speed of each LSM is maintained to the largest extent, and consequently, the speed of the system is maintained at the largest extent as well. For a particular system where it's possible to lower the frequency of some module without having a huge detrimental impact on the system performance, some other frequency schemes might be applied. However, once more, that would not be a systematic, generally applicable solution, but rather a specific case study. In general, the performance will be the least affected by galsification if the plesiochronous frequency scheme is applied.

Another important advantage of plesiochronous clocking is that it enables the best predictability of substrate noise reduction. Substrate noise in frequency domain (for a system where the package parasitics are dominant compared to the on-chip parasitics) can be regarded as a product of switching current spectrum and the PDN and substrate transfer function $H = H_{PDN}H_{Sub}$, as shown in equation (4.1). Thus, the spectral peak attenuation of the n^{th} substrate noise harmonic achieved by galification can be expressed as:

$$SPA_{Sub,n} = 20 \log \left(\frac{|V_{Sub,sync}(nf_0)|}{\max\{|V_{Sub,LSM_m}(nf_m)|\}} \right) = 20 \log \left(\frac{|H(nf_0)I_{sw,sync}(nf_0)|}{\max\{|H(nf_m)I_{sw,LSM_m}(nf_m)|\}} \right) \quad (4.3)$$

The transfer function H has different values at different frequencies. If some other frequency scheme was used, i.e. if the frequencies of the corresponding GALS and synchronous spectral peaks weren't close to each other, the value of $SPA_{Sub,n}$ would strongly depend on the transfer function $H = H_{PDN}H_{Sub}$, which is very difficult to estimate precisely, especially in an early phase of the design process.

However, the transfer function H is continuous, which means that if the frequencies are so close that their difference is negligible, then the differences in transfer function values at those frequencies are also negligible. Since the peaks in the switching current spectrum of a GALS plesiochronous system are at frequencies which are very close to the frequencies of the peaks in the switching current spectrum of the original synchronous system, the transfer function value for the corresponding harmonics in the two systems is practically the same, i.e.:

$$(\forall_{m \in \{1,2,\dots,M\}})(H(nf_0) \approx H(nf_m) = H(n(1 + \Delta_m)f_0)) \quad (4.4)$$

Consequently, for the plesiochronous scheme, the attenuation of substrate noise harmonic peak is approximately the same as the attenuation of the corresponding switching current peak:

$$\begin{aligned} SPA_{Sub,n} &= 20 \log \left(\frac{|H(nf_0)I_{sw,sync}(nf_0)|}{\max\{|H(nf_m)I_{sw,LSM_m}(nf_m)|\}} \right) \\ &\approx 20 \log \left(\frac{|H(nf_0)| |I_{sw,sync}(nf_0)|}{|H(nf_0)| \max\{|I_{sw,LSM_m}(nf_m)|\}} \right) \\ &= 20 \log \left(\frac{|I_{sw,sync}(nf_0)|}{\max\{|I_{sw,LSM_m}(nf_m)|\}} \right) = SPA_{i_{sw},n} \end{aligned} \quad (4.5)$$

As a consequence, instead of directly optimizing the spectrum peak attenuation of substrate noise, the spectrum peak attenuation of switching current in the corresponding frequency bands can be optimized. This is very convenient and technically much more feasible, since the switching current is much easier to estimate than the substrate noise.

In the next section, a partitioning of an initially synchronous system into a plesiochronous GALs system is mathematically analyzed, and an optimal partitioning strategy for substrate noise reduction is determined.

4.2.2. Mathematical analysis of the optimal plesiochronous GALs partitioning for substrate noise reduction

As an initial point, let's observe a synchronous system, with a clock frequency f_0 and a clock period T_0 . The waveform of the switching current produced by this system can be separated into two components – the average switching current waveform over various clock periods (theoretically over all of the clock periods), $i(t)$, and the cycle to cycle deviation from the average switching current waveform, $\Delta i(t)$ [29]. In spectral domain, these components correspond to the harmonic peaks and the noise floor, respectively. The averaged waveform $i(t)$ is a periodical signal with a period T_0 , which means that its spectrum can be represented as a Fourier series calculated with the fundamental frequency equal to the clock frequency f_0 .

Now let's assume that the initial system is partitioned into M partitions, with each of the partitions still maintaining the same clock frequency f_0 . Let the average switching current waveforms of these partitions be $i_1(t), i_2(t), \dots, i_M(t)$. For the total average switching current waveform of the system it can now be written:

$$i(t) = \sum_{m=1}^M i_m(t) \quad (4.6)$$

As the waveforms $i_1(t), i_2(t), \dots, i_M(t)$ are also periodic with a period T_0 , their spectra can be represented as Fourier series as well, calculated with the same fundamental frequency f_0 as the Fourier series representing the spectrum of $i(t)$. Let the n^{th} component of the Fourier series of the waveform $i(t)$, calculated on the synchronous clock frequency f_0 , be $F_n\{i\}$, and let the n^{th} component of the Fourier series of the waveform $i_m(t)$, corresponding to the m^{th} partition, and calculated on the same fundamental frequency f_0 , be $F_n\{i_m\}$. Since both $F_n\{i\}$ and each of the $F_n\{i_m\}$, $m \in 1, 2, \dots, M$, are calculated with the same fundamental frequency, the linearity of the Fourier series is applicable, i.e. the Fourier series of the sum of signals is equal to the sum of Fourier series. Thanks to this property, it can be written:

$$F_n\{i\} = F_n\left\{\sum_{m=1}^M i_m\right\} = \sum_{m=1}^M F_n\{i_m\} \quad (4.7)$$

Note that $F_n\{i\}$, as well as each of the $F_n\{i_m\}$, are complex numbers.

Now, let's assume that the partitions are clocked with independent clocks having mutually different frequencies f_1, f_2, \dots, f_M . In other words, let's assume that the system is galsified and that

the partitions are actually LSMs of a GALS system. Fourier series components $F_n\{i_m\}$, $m \in 1, 2, \dots, M$, are now calculated with different fundamental frequencies, and they represent the M spectral peaks of the switching current of the GALS system, corresponding to the n^{th} harmonic.

Due to setting new frequencies for clocking the LSMs, their switching current waveforms $i_m(t)$, $m \in 1, 2, \dots, M$, will change compared to the synchronous case, when they all were clocked with a synchronous clock having a frequency f_0 . This change would typically consist of extending the initial waveform with an interval in which the current is close to zero, as shown in Fig. 4.3. As a consequence, the new values of the Fourier series components $F_n\{i_m\}$ will differ from their values calculated in the synchronous case. Thus, the equation (4.7), with $F_n\{i\}$ corresponding to the switching current waveform of the initial synchronous system, and $F_n\{i_m\}$, $m \in 1, 2, \dots, M$, corresponding to the switching current waveforms of the LSMs in the new GALS system, will not be valid anymore.

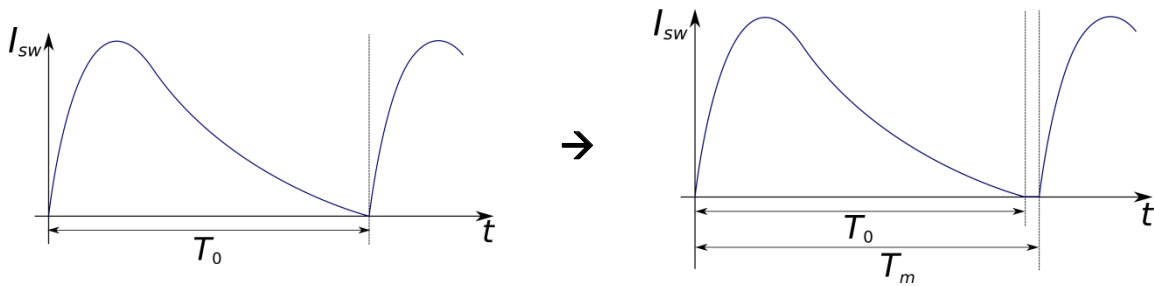


Figure 4.3 – A change of a switching current waveform of an LSM after applying a slightly lower clock frequency compared to the synchronous case

However, if the GALS system is plesiochronous, i.e. if the clock frequencies of different partitions are very close to each other and to the synchronous frequency, the switching current waveforms $i_1(t)$, $i_2(t)$, ..., $i_M(t)$, will change due to galsification only slightly. Consequently, their corresponding Fourier series components $F_n\{i_m\}$, $m \in 1, 2, \dots, M$, will remain approximately the same as they were for the synchronous system. Thanks to this, it can be considered that the equation (4.7) is still approximately valid after the initially synchronous system is converted into a plesiochronous GALS system. This very important approximation will be called the plesiochronous approximation further throughout this thesis, and will be analyzed in detail in Section 4.2.4.

It is important not to forget that, while in the synchronous system the spectral components $F_n\{i_m\}$ actually do physically sum up to form a single spectral component $F_n\{i\}$, in a plesiochronous GALS system they form M separate spectral peaks close to each other. Adding the spectral components at different frequencies wouldn't make sense physically. The plesiochronous approximation only states that if the complex numbers representing the n^{th} switching current harmonics of the LSMs of a plesiochronous GALS system are summed up, the result would be approximately equal to the complex number representing the n^{th} switching current harmonic of the equivalent synchronous system.

Similar to (4.6), for substrate voltage waveforms it can be written:

$$v_{sub}(t) = \sum_{m=1}^M v_{subm}(t) \quad (4.8)$$

where $v_{sub}(t)$ is the substrate noise voltage at victim's location for the initial synchronous system, and $v_{subm}(t)$ the contribution of the m^{th} partition to the substrate noise voltage at victim's location (i.e. the substrate noise voltage at victim's location if only the m^{th} partition is active). If the system is synchronous, the linearity of the Fourier series is valid, so it can be written:

$$F_n\{v_{sub}\} = F_n\left\{\sum_{m=1}^M v_{subm}(t)\right\} = \sum_{m=1}^M F_n\{v_{subm}\} \quad (4.9)$$

As already written in the equation (4.1) in the previous section, substrate noise in frequency domain can be represented as a product of the switching current spectrum $I_{sw}(f)$ and the PDN and substrate transfer function $H(f) = H_{PDN}(f)H_{sub}(f) = V_{sub}(f)/I_{sw}(f)$. If the average switching current waveform and the average substrate noise waveform are observed, their spectra can be represented as Fourier series, both for the synchronous system and for each of the LSMs of the GALS system:

$$\begin{aligned} I_{sw,sync}(nf_0) &= F_n\{i\} \\ I_{sw,LSM_m}(nf_m) &= F_n\{i_m\} \quad , \quad m \in 1, 2, \dots, M \\ V_{sub,sync}(nf_0) &= F_n\{v_{sub}\} \\ V_{sub,LSM_m}(nf_m) &= F_n\{v_{subm}\} \quad , \quad m \in 1, 2, \dots, M \end{aligned} \quad (4.10)$$

Consequently, if a simplified notation is applied, so that:

$$\begin{aligned} H(nf_0) &= H_n \\ H(nf_m) &= H_{nm} \quad , \quad m \in 1, 2, \dots, M \end{aligned} \quad (4.11)$$

based on (4.1) it can be written that:

$$\begin{aligned} F_n\{v_{sub}\} &= H_n F_n\{i\} \\ F_n\{v_{subm}\} &= H_{nm} F_n\{i_m\} \end{aligned} \quad (4.12)$$

Since the chosen clocking scheme of the new GALS system is plesiochronous, as explained in the previous subsection, it can be considered that:

$$H_{n1} \approx H_{n2} \approx \dots \approx H_{nM} \approx H_n \quad (4.13)$$

Thus, using the (4.12), (4.13) and (4.7) in a plesiochronous approximation, the following equation is obtained:

$$F_n\{v_{sub}\} = H_n F_n\{i\} = H_n \sum_{m=1}^M F_n\{i_m\} = \sum_{m=1}^M H_n F_n\{i_m\} \approx \sum_{m=1}^M H_{nm} F_n\{i_m\} = \sum_{m=1}^M F_n\{v_{subm}\} \quad (4.14)$$

which is actually identical to the equation (4.9). This means that the equation (4.9), besides being valid for the partitions of a synchronous system, is also approximately valid for the LSMs of a plesiochronous GALS system (in a plesiochronous approximation). In other words, this way it's been shown that the plesiochronous approximation is valid not only for the switching current, but also for the substrate noise.

According to the definition of the spectral peak attenuation of the n^{th} harmonic of substrate noise voltage (4.3) and using (4.10), it can be written that:

$$SPA_{sub,n} = 20 \log \frac{|F_n\{v_{sub}\}|}{\max\{|F_n\{v_{subm}\}|\}} \quad (4.15)$$

which, by using (4.12) can further be expressed as:

$$\begin{aligned}
SPA_{sub,n} &= 20 \log \frac{|H_n||F_n\{i\}|}{\max\{|H_{nm}||F_n\{i_m\}|\}} \approx 20 \log \frac{|H_n||F_n\{i\}|}{|H_n| \max\{|F_n\{i_m\}|\}} \\
&= 20 \log \frac{|F_n\{i\}|}{\max\{|F_n\{i_m\}|\}} = SPA_{i_{sw},n}
\end{aligned} \tag{4.16}$$

This, by the way, corresponds to the conclusion already shown in the equation (4.5) in the previous subsection – for the plesiochronous GALS clocking scheme, the attenuation of substrate noise harmonic is the same as the attenuation of the corresponding switching current harmonic. This is the predictability feature, which is, as explained in Section 4.2.1, one of the main reasons to select the plesiochronous scheme.

By using the plesiochronous approximation (4.7), the equation (4.16) can further be expressed as:

$$SPA_{sub,n} = 20 \log \frac{|\sum_{m=1}^M F_n\{i_m\}|}{\max\{|F_n\{i_m\}|\}} \tag{4.17}$$

The goal of a substrate noise reduction methodology is to maximize the SPA for the substrate noise harmonics in the critical frequency band. For the time being, let's assume that only one substrate noise harmonic is critical. The procedure in case of targeting more harmonics will be explained in the subsequent sections.

As it can be seen from the equation (4.17), maximization of the SPA for the n^{th} substrate noise harmonic is equivalent to the maximization of the ratio $|\sum_{m=1}^M F_n\{i_m\}|/\max\{|F_n\{i_m\}|\}$. This actually represents a ratio between the magnitude of a sum of M complex numbers and the largest among the magnitudes of the elements of this sum. Such a ratio is maximized if all of the sum elements are identical, both in magnitude and phase:

$$F_n\{i_1\} = F_n\{i_2\} = \dots = F_n\{i_M\} = \frac{F_n\{i\}}{M} \tag{4.18}$$

In other words, the phases of each of the sum elements $F_n\{i_m\}$, $M = 1, 2, \dots, M$, should be equal to the phase of the sum $F_n\{i\}$:

$$\varphi(F_n\{i_1\}) = \varphi(F_n\{i_2\}) = \dots = \varphi(F_n\{i_M\}) = \varphi(F_n\{i\}) \tag{4.19}$$

and the magnitudes of all of the sum elements should be equal:

$$|F_n\{i_1\}| = |F_n\{i_2\}| = \dots = |F_n\{i_M\}| = \frac{|F_n\{i\}|}{M} \tag{4.20}$$

By applying (4.20) in the equation (4.16), the maximum theoretically possible value for $SPA_{sub,n}$ is obtained:

$$\max\{SPA_{sub,n}\} = 20 \log (M) \tag{4.21}$$

Note that, in order to obtain this result, no assumption about the shape of the switching current waveforms in the time domain was taken. Thus, it is valid regardless of the switching current waveform shape. However, the deduction is based on the plesiochronous approximation for the switching current; the impact of this approximation is analyzed in the Section 4.2.4.

Since the optimum result is obtained when the corresponding switching current harmonics in all the LSMs of a GALS system are equal (i.e. balanced), this partitioning methodology is named “harmonic-balanced plesiochronous GALS partitioning methodology” [P5]. Note that, despite having

a similar name, this methodology is not related to the harmonic balance method as a procedure for determining the steady-state response of a nonlinear system.

It is interesting to notice that the maximum spectral peak attenuation from (4.21) is the same as the maximum spectral peak attenuation for the switching current low harmonics provided by the power balanced plesiochronous GALS partitioning methodology [16][160]. The relation between these two GALS partitioning methodologies will be analyzed in Section 4.6.

So far in this section, the spectral peak attenuation of a single spectral peak by the harmonic-balanced plesiochronous GALS partitioning methodology has been analyzed. Now, the attenuation of the spectral power of substrate noise will be analyzed as well. Since the continuous noise floor is lying far below the level of the harmonic peaks, only harmonic power will be taken into account.

First, let's observe the switching current spectrum. Let the magnitude of the n^{th} switching current harmonic peak in an initial synchronous system be $A_n = F_n\{i\}$, which means that the power of this harmonic can be expressed as cA_n^2 , with c being a multiplicative constant. If this system is replaced by a plesiochronous GALS system with M LSMs and perfect harmonic balancing at the n^{th} harmonic, the magnitude of each of the corresponding GALS harmonic peaks will be A_n/M , its spectral power $c(A_n/M)^2$ while the total power of GALS harmonic peaks corresponding to the n^{th} harmonic will be $Mc(A_n/M)^2 = cA_n^2/M$. Harmonic peak attenuation at the n^{th} harmonic is:

$$SPA_{I_{sw},n} = 20 \log \left(\frac{A_n}{A_n/M} \right) = 20 \log (M) \quad (4.22)$$

while the spectral power attenuation for the narrow band around the n^{th} harmonic is:

$$PA_{I_{sw},n} = 10 \log \left(\frac{A_n^2}{A_n^2/M} \right) = 10 \log (M) \quad (4.23)$$

Now, let's analyze the substrate voltage spectrum. The derivation is very similar to the one for switching current spectrum. Let H_n be the value of the PDN and substrate transfer function (from switching current to the substrate voltage at the position of the victim). The n^{th} harmonic peak of substrate noise is then $H_n A_n$, and its power is $c(H_n A_n)^2$. When the initial plesiochronous system is transformed into a plesiochronous GALS system with M LSMs and perfect harmonic balancing at the n^{th} harmonic, the n^{th} harmonic peak from the initial synchronous system would be replaced by M peaks, each having the magnitude $H_n A_n/M$, and the total power in a narrow band around the n^{th} harmonic is $Mc(H_n A_n/M)^2 = c(H_n A_n)^2/M$. Note that a plesiochronous approximation is applied here – since harmonic peaks of the plesiochronous GALS system corresponding to the same synchronous harmonic peak are very close to each other, for each of them the value of the PDN and substrate transfer function will be approximately the same (for n^{th} synchronous harmonic, equal to H_n). Spectral peak attenuation at the n^{th} harmonic is, as already shown in (4.21):

$$SPA_{V_{sub},n} = 20 \log \left(\frac{H_n A_n}{H_n A_n/M} \right) = 20 \log (M) \quad (4.24)$$

and the spectral power attenuation in a narrow band around the n^{th} harmonic is:

$$PA_{V_{sub},n} = 10 \log \left(\frac{(H_n A_n)^2}{(H_n A_n)^2/M} \right) = 10 \log (M) \quad (4.25)$$

This means that, besides providing spectral peak attenuation of substrate noise, harmonic-balanced plesiochronous GALS partitioning also provides the substrate noise spectral power

attenuation as well. This is an advantage over another system-level substrate noise reduction methodology – clock modulation, which replaces each harmonic spectral peak with a lower but wider one, and thus attenuates the spectral peak, without reducing its spectral power.

4.2.3. Avoiding the harmonic overlapping

At higher frequencies, it's possible that GALS clock harmonic of the order $n - 1$ of one LSM overlaps with the GALS clock harmonic of the order n of another LSM. Such overlapping can cause unwanted high noise peaks at higher frequencies. Examples of this are shown in Fig. 4.4 and Fig. 4.5.

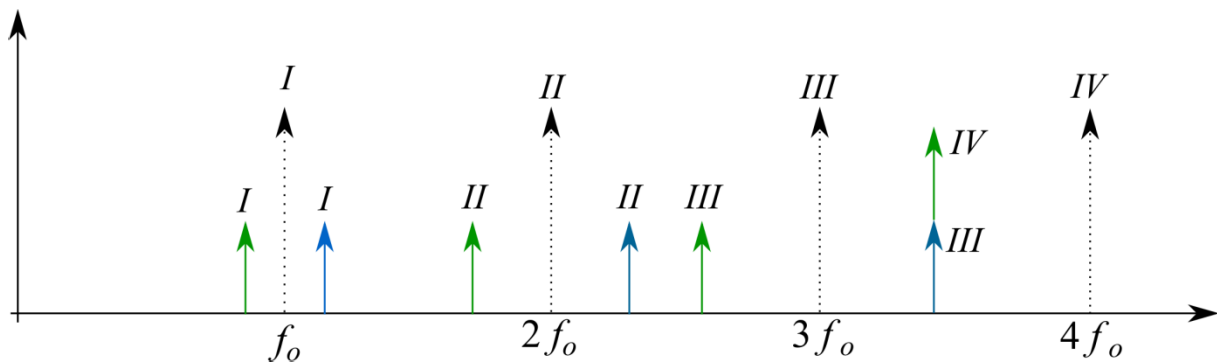


Figure 4.4 – An example of harmonic overlapping, with one of the LSMs having a clock frequency higher than the initial clock frequency, and another LSM having a clock frequency lower than the initial clock frequency. Black dotted – harmonics of the original synchronous system, Green – harmonics of the LSM with the lowest clock frequency, Blue – harmonics of the LSM with the highest clock frequency

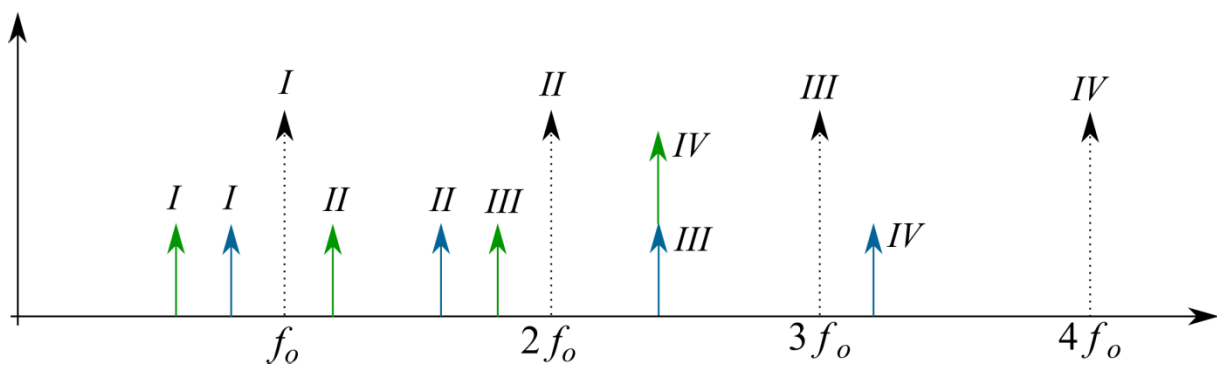


Figure 4.5 – An example of harmonic overlapping, with both LSMs having a clock frequency lower than the initial clock frequency. Black dotted – harmonics of the original synchronous system, Green – harmonics of the LSM with the lowest clock frequency, Blue – harmonics of the LSM with the highest clock frequency

The condition for the overlapping not to happen (the “overlapping safe” condition) is:

$$n f_{GALSmax} \leq (n + 1) f_{GALSmin} \quad (4.26)$$

The frequencies $f_{GALSmax}$ and $f_{GALSmin}$ are maximum and minimum LSM frequencies (at fundamental), respectively. For a plesiochronous scheme:

$$f_{GALSmax} = (1 + \Delta_{max})f_0, \quad f_{GALSmin} = (1 + \Delta_{min})f_0 \quad (4.27)$$

where f_0 is the original synchronous clock frequency, and $|\Delta_{max}| \ll 1$, $|\Delta_{min}| \ll 1$. Combining the equations (4.26) and (4.27), the “overlapping safe” condition becomes:

$$n(1 + \Delta_{max})f_0 \leq (n + 1)(1 + \Delta_{min})f_0 \quad (4.28)$$

From (4.28), the maximum order of the “overlapping safe” harmonic is obtained:

$$n \leq \frac{1 + \Delta_{min}}{\Delta_{max} - \Delta_{min}} \quad (4.29)$$

As it can be noticed from the expression (4.29), for higher Δ_{min} and for smaller difference between Δ_{max} and Δ_{min} , the “overlapping safe zone” is wider, i.e. the harmonics at which this problem might occur are higher. Since Δ_{min} usually can't be above 0, the limit will be higher if Δ_{max} and Δ_{min} can be kept as close to each other as possible, i.e. if the plesiochronous clocking is such that the difference between the LSM clock frequencies is as small as possible.

The limit from (4.29), however, is too conservative. It's possible to avoid overlapping even at frequencies above the limit from (4.29) by frequency planning, i.e. to make sure that the ratio of any pair of the LSM clock frequencies is not a rational number. This, however, might not be safe enough, since the LSM clocks might have somewhat higher phase noise (especially if they're implemented as ring oscillators), which means that even having two peaks close to each other would result in amplifying the peaks. In that case, not just completely overlapping harmonics but also the harmonics which are too close to each other should be avoided. This results in the following condition:

$$(\forall_{n_1, n_2})(|n_1(1 + \Delta_1)f_0 - n_2(1 + \Delta_2)f_0|) \geq f_{\Delta} \quad (4.30)$$

where f_{Δ} is the minimum allowable distance between two frequency peaks, which is determined based on the phase noise of the clock. If this condition is fulfilled, overlapping is avoided, and it doesn't pose any upper limit for the highest harmonic order which can be attenuated. In practice, the condition (4.30) has to be satisfied only for n_1 and n_2 within the frequency band of interest, which makes the condition easier to fulfill.

4.2.4. The impact of the plesiochronous approximation

In the Section 4.2.2, the theoretical foundations of the harmonic balanced plesiochronous GALS partitioning methodology for substrate noise reduction have been presented. The methodology is developed by applying the plesiochronous approximation, i.e. by considering that the harmonics of the switching current spectrum for each of the LSMs remain largely unchanged if the synchronous frequency f_0 is replaced by a plesiochronous frequency $f_m = (1 + \Delta_m)f_0$, $|\Delta_m| \ll 1$. In this subsection, the plesiochronous approximation itself and its impact on the result of the methodology are analyzed.

For this purpose, a periodic triangular pulse model for the switching current waveform is used, as shown in Fig. 4.6. As already mentioned in Chapter 2, this is the same model as the one which was used in [16], and before that in [48]. In [48] it was shown that the model is well suitable for small to medium scale (~40 k-gates) digital systems.

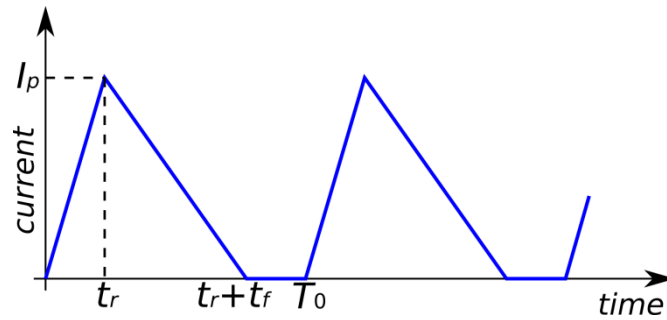


Figure 4.6 – Periodic triangular pulse as a model for switching current (adapted from [P5])

As the first step, the approximation is analyzed by observing a single partition. In Fig. 4.7.a the spectrum envelope of that partition's switching current is shown, calculated in four ways. The red line corresponds to the case when the Fourier series elements are calculated with the synchronous clock frequency as the fundamental, while the magenta, green and blue lines correspond to the case when the Fourier series elements are calculated with the plesiochronous frequency as the fundamental, having an offset Δ_m of 1%, 3% and 5%, respectively. In Fig. 4.7.b-d the difference of the harmonic peaks calculated without an offset and the harmonic peaks calculated with the respective offset is shown, expressed in dB.

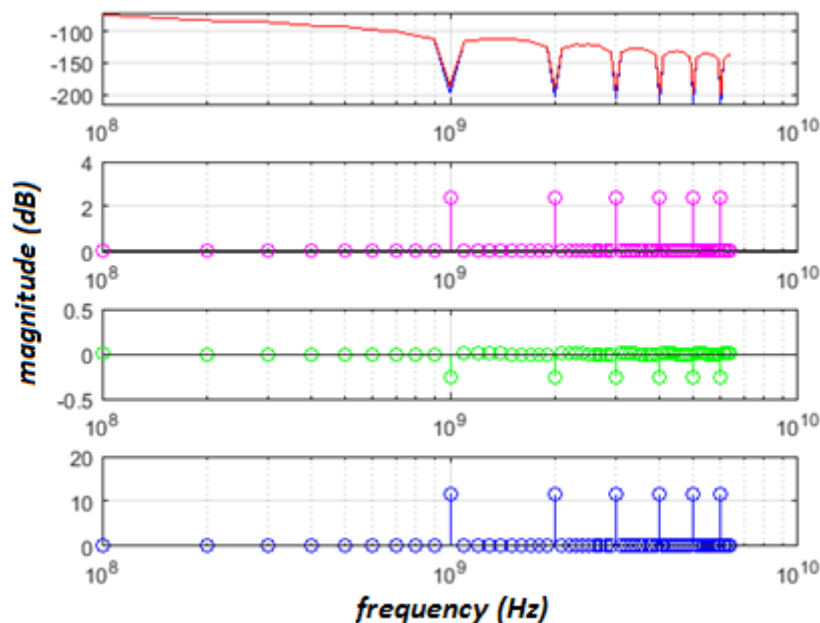


Figure 4.7 – Spectrum envelope for a single block, calculated with and without neglecting the frequency offset (a), and error introduced by this approximation (b-d)
 red – synchronous frequency, magenta – 1% offset, green – 3% offset, blue – 5% offset
 (adapted from [P5])

As it can be noticed from Fig. 4.7, at most of the frequencies the error introduced by the plesiochronous approximation is negligible. The only exceptions are the “notches”, i.e. the parts of the spectrum where the envelope has a sharp minimum. This is actually expected – since the spectrum magnitude changes the most around the notch frequencies, that part of the spectrum is the most sensitive to the frequency drift when calculating the Fourier series. However, since the magnitude of the switching current spectrum at notch frequencies is substantially lower compared to the remaining part of the spectrum, spectral peaks in this region would not have any detrimental effect. Thus, it can be concluded that the error introduced by the plesiochronous approximation

impacts only the part of the spectrum where there are no spectral peaks the reduction of which would be required.

Further, in order to evaluate the impact of this error on the results of the methodology itself, a GALS system is observed, resulting from a theoretically perfect harmonic balancing of an initial synchronous system at all frequencies. The GALS system has 5 plesiochronously clocked LSMs, having clock frequency offsets of -4%, -2%, 1%, 3% and 5%. In Fig. 4.8.a, the spectrum envelope of the switching current of an initial synchronous system is shown. Spectral peak attenuation is then calculated numerically in MATLAB, once by applying the plesiochronous approximation (i.e. by calculating the Fourier series with the synchronous frequency as the fundamental), and once by not applying the plesiochronous approximation (i.e. by calculating the Fourier series with taking into account the frequency offsets of the LSMs). The results are shown in Fig. 4.8.b and Fig. 4.8.c, respectively.

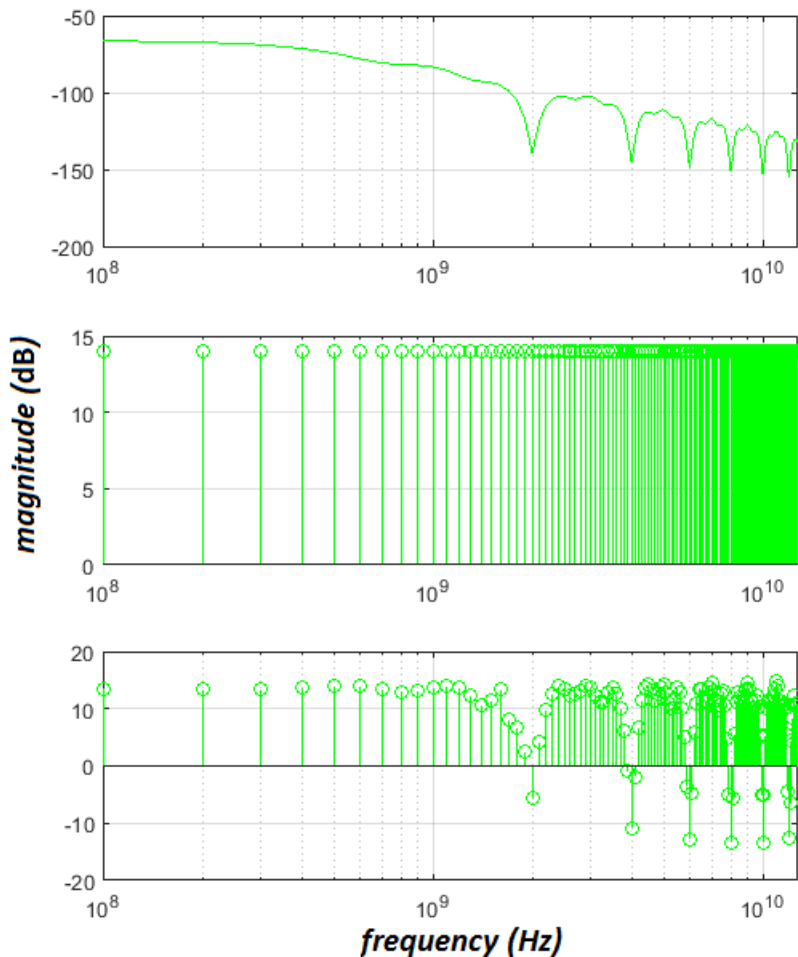


Figure 4.8 – Spectrum envelope of switching current for a synchronous system (a), and spectrum peak attenuation with plesiochronous harmonic balanced 5 LSM GALS partitioning, calculated with neglecting the frequency offset (b) and without neglecting the frequency offset (c) (adapted from [P5])

When the plesiochronous approximation is applied, the numerically obtained SPA is 13.98 dB at all frequencies, which, as expected, corresponds to the theoretically predicted attenuation of $20\log(M)$. On the other hand, when the frequency offset is also taken into account

for the numerical evaluation, the results vary from the theoretical prediction. The variations are somewhat more expressed at higher frequencies, which is expected, since for the higher order harmonics the absolute value of the frequency offset gets bigger. For most of the spectral peaks, however, the attenuation stays close to the theoretically predicted value.

The only frequency regions where the attenuation is seriously deteriorated are around the “notch” frequencies, i.e. the frequencies where the spectrum envelope, as shown in Fig. 4.8.a, reaches sharp minima. At these frequencies, the attenuation even becomes negative, i.e. it turns into amplification of spectral peaks. This, however, doesn’t have any detrimental effects – even when amplified, the peaks at and around the “notch” frequencies stay substantially lower than the rest of the spectrum.

Thus, it can be concluded that, for the relevant parts of the switching current spectrum, the spectral peak attenuation obtained by the harmonic balanced plesiochronous GALS partitioning is close to the optimum value, calculated by using the plesiochronous approximation. In other words, the plesiochronous approximation significantly impacts the result of the harmonic balanced plesiochronous GALS partitioning only in the part of the spectrum where the attenuation is not needed anyway.

So far in this section, a theoretically perfect harmonic balanced plesiochronous GALS partitioning was assumed, where the initial synchronous system is partitioned into LSMs so that all of the LSMs are perfectly balanced at each of the harmonics which should be attenuated. However, achieving a perfectly balanced partitioning is usually impossible in practice. Each digital system consists of a finite number of digital blocks, usually of different size and switching current profile, and due to this finite granularity, these blocks can’t be combined into partitions in such a way that a perfect harmonic balancing is achieved. This is especially the case if attenuation is required in a frequency band containing several harmonics or in several frequency bands – simultaneously achieving perfect harmonic balancing at each of the relevant harmonic frequencies is even less practically feasible.

As a consequence, the following question arises: How to practically perform a harmonic balanced plesiochronous GALS partitioning for a realistic system, consisting of a finite number of digital blocks, so that the partitioning result is as close as possible to the perfect harmonic balancing at frequencies of interest. To achieve this, an algorithm for practical harmonic-based plesiochronous GALS partitioning is developed. This algorithm is presented in the following section, and the partitioning methodology based on it is numerically evaluated.

4.3. An algorithm for harmonic-balanced GALS partitioning, and numerical evaluation of the methodology in MATLAB

4.3.1. An algorithm for harmonic-balanced GALS partitioning

A GALS partitioning procedure has an initial synchronous system as its starting point. Let the initial synchronous system which is to be partitioned and galsified consist of N blocks. Blocks are parts of the design which should be considered unpartitionable. While it is theoretically possible to consider each sequential element as a separate block, it is preferable to define blocks as hierarchical structures, such that the number of interface signals between a block and the rest of the system is kept relatively small. Such an approach would make galsification easier, because it would avoid having extensively large interfaces between the LSMs.

Let each of these blocks have its own switching current profile, i.e. an average switching current waveform. Let these profiles be $i_{b1}(t)$, $i_{b2}(t)$, ..., $i_{bN}(t)$. In practice, these profiles can be obtained by simulating the blocks within the system, using a representative test case which corresponds to the intended usage. For each of the profiles, a switching current spectrum can further be obtained by applying the Fourier series. For example, the n^{th} harmonic of the switching current of the block with the index b_k and the corresponding current profile $i_{bk}(t)$ would be $F_n\{i_{bk}\}$. Note that $F_n\{i_{bk}\}$ is a complex number.

Let the switching current profile of the entire initial synchronous system be $i(t)$. The system current profile is the sum of the block profiles:

$$i(t) = \sum_{k=1}^N i_{bk}(t) \quad (4.31)$$

Consequently, each harmonic of the spectrum of the synchronous system is equal to the sum of the corresponding harmonics of block spectra:

$$F_n\{i\} = \sum_{k=1}^N F_n\{i_{bk}\} \quad (4.32)$$

When the initial synchronous system is transformed into a GALS system, each of the blocks is assigned to one of the partitions of the new GALS system, i.e. to one of the LSMs. Let the new GALS system consist of M LSMs, denoted as $LSM_1, LSM_2, \dots, LSM_M$. Let p_m be a set of indices of all the blocks which are assigned to the partition LSM_m . Let the resulting current profile of this LSM be i_m . Similar to (4.31), it can be written that:

$$i_m = \sum_{b_k \in p_m} i_{bk} \quad (4.33)$$

Also, similar to (4.32), each harmonic peak in the switching current spectrum of the partition LSM_m represents the sum of the corresponding harmonic peaks of the switching current spectra of blocks which are contained within that LSM:

$$F_n\{i_m\} = \sum_{k \in p_m} F_n(i_{bk}) \quad (4.34)$$

The perfect harmonic-balanced plesiochronous GALS partitioning for this system would be achieved if for each partition LSM_m the following condition is valid at the targeted harmonic n :

$$F_n\{i_m\} = \frac{F_n\{i\}}{M} \quad (4.35)$$

In case that more harmonics are targeted, for the perfect harmonic-balanced plesiochronous GALS partitioning the condition (4.35) would have to be achieved at each of the targeted harmonics.

The goal of a practical partitioning methodology should thus be to combine the blocks $b_1, b_2, \dots, b_k, \dots, b_N$ in partitions $LSM_1, LSM_2, \dots, LSM_M$ so that the harmonic peaks of switching currents for each of the partitions are as close as possible to the value from equation (4.35). The possible criteria of determining how close a particular solution is to the perfect partitioning from (4.35) will be discussed later in this subsection – for the time being, we'll just assume that a cost function can be defined, such that its value gets closer to zero as the selected partitioning gets closer to the perfect harmonic-balancing at frequencies of interest.

An obvious way to find the optimum partitioning would be to explore all of the possible combinations of blocks, and select the one with the minimum value of the cost function. However, a solution space may be too large for such an extensive search – for N blocks which are to be grouped into M partitions, the number of possible combinations is $M^N/M! - (M-1)^N/(M-1)!$. This can be a very large number for a system containing a large number of blocks. Thus, it would be preferable to apply some heuristic method for finding the optimum partitioning. In this work, the simulated annealing algorithm [200] is applied, as shown in Fig. 4.9.

As an input to the optimization procedure, current profiles of each of the blocks should be provided, and the spectra are calculated for each of the blocks. Then, a random initial partitioning is chosen, with each of the N blocks being randomly assigned to one of the M partitions. For this initial partitioning, a cost function is calculated, determining how good the partitioning is.

After this initial procedure, the algorithm enters a loop. Each loop iteration starts with choosing a random solution which is “neighboring” to the current solution (in the first iteration, to the initial solution). Two solutions are “neighboring” if they differ only in the value of one parameter, in this case only in an assignment of one block. In other words, one block is randomly selected, and the partition to which it is assigned is randomly changed. After this rearrangement, the cost function is calculated for the new, “neighboring” partitioning as well, and its value is compared to the value of the current partitioning. If the value of the cost function is lower after the rearrangement (i.e. if the new partitioning is better than the current one), the new partitioning is set as the current partitioning for the next algorithm iteration. In case that the rearrangement has resulted in a higher value of the cost function, the new partitioning can also be set as the current partitioning for the next iteration, but only with a certain acceptance probability a_p . When the algorithm starts looping, this acceptance probability is high, and it gets reduced by a constant multiplicative factor with every iteration. In other words, the algorithm intentionally provides a possibility of choosing a worse partitioning over a better one; however, the probability of such a choice is high only at the beginning of the algorithm execution, and choosing a worse partitioning over a better one becomes less probable as the algorithm progresses. This occasional choice of a worse solution in an initial phase of the algorithm is a crucial feature of simulated annealing, since it prevents the algorithm from getting stuck at a local optimum.

The algorithm is ended when the cost function converges to the final solution. Note that the algorithm is random, which has a consequence that, in case that there are several optima with close values of the cost function, it can deliver different results for different runs. However, all of the delivered results would have close final values of the cost function, i.e. they would be approximately equally close to the perfect harmonic balancing.

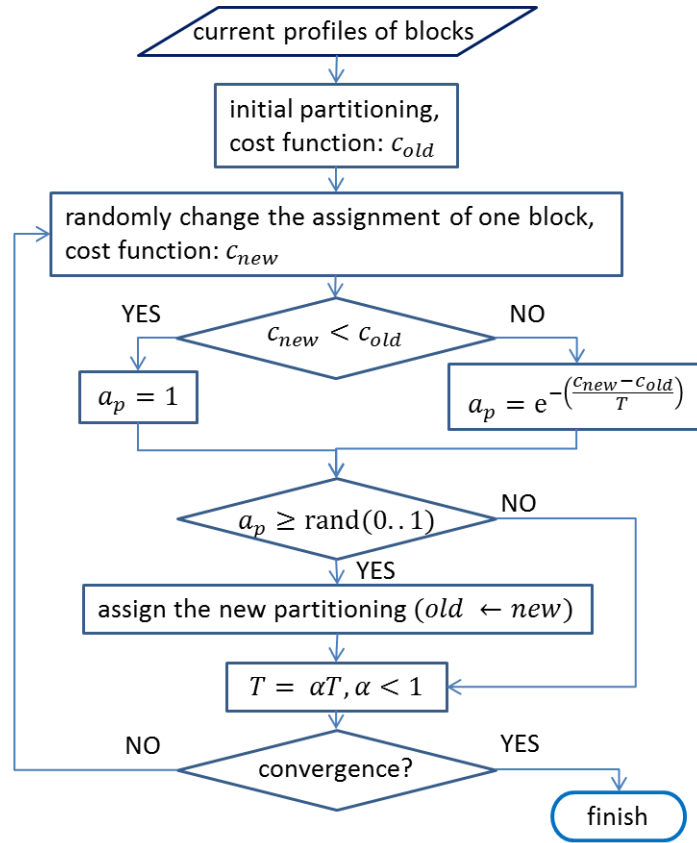


Figure 4.9 – A simulated annealing algorithm for performing the harmonic-balanced plesiochronous GALS partitioning

Now, let's discuss the possible ways to select a cost function, since the proper choice of a cost function is crucial for the algorithm to deliver an optimal result. As already stated, there are several ways to select a cost function.

Since the spectrum harmonics of the initial synchronous system $F_n\{i\}$ are known at the beginning of the algorithm, the spectrum which a theoretically perfectly balanced partitioning would have can be calculated from (4.35). Note that both $F_n\{i_m\}$ and $F_n\{i\}/M$ are complex numbers, which means that they should match as much as possible both in magnitude and phase, i.e. both in real and imaginary part. Thus, for each of the partitions LSM_m , a criterion for how good the matching is at the harmonic n can be defined as:

$$c_{1mn} = \left(\text{Re}\{F_n\{i_m\}\} - \text{Re}\left\{\frac{F_n\{i\}}{M}\right\} \right)^2 + \left(\text{Im}\{F_n\{i_m\}\} - \text{Im}\left\{\frac{F_n\{i\}}{M}\right\} \right)^2 = \left| F_n\{i_m\} - \frac{F_n\{i\}}{M} \right|^2 \quad (4.36)$$

Accordingly, the cost function can be formed by summing the matching criteria (4.36) for all the LSMs and for all the harmonics n in the frequency band of interest:

$$c_1 = \sum_m \sum_n \left| F_n\{i_m\} - \frac{F_n\{i\}}{M} \right|^2 \quad (4.37)$$

Since the cost function c_1 is based on the direct comparison to the theoretically perfect harmonic balancing, using this cost function provides the optimal harmonic peak attenuation both for the switching current harmonics and for the substrate noise harmonics in the frequency range of interest.

In case that the harmonic-balanced plesiochronous GALS partitioning is performed not for the purpose of substrate noise reduction, but for the purpose of the EMI reduction in a certain band, i.e. for the purpose of reducing the switching current harmonic peaks in a certain band, a cost function could be chosen so that it corresponds to the lowest spectral power of GALS switching current harmonics in this frequency band. As already mentioned in 4.2.2, the power of each harmonic is proportional to the square of its magnitude. Using this, the following cost function can be defined:

$$c_2 = \sum_m \sum_n |F_n\{i_m\}|^2 \quad (4.38)$$

This cost function is to be used only when the switching current harmonics themselves are targeted for reduction.

In cases when there are several analog modules in the chip, being sensitive in different frequency bands, and having different requirements for substrate noise reduction, it might be needed to favor the noise reduction in one of these bands over the noise reduction in the others. This can be achieved by introducing weight coefficients, which would give more importance to attenuation of more critical frequency components. Including the weight components in the cost functions c_1 and c_2 results in the following two new cost functions, respectively:

$$c_3 = \sum_m \sum_n w_n \left| F_n\{i_m\} - \frac{F_n\{i\}}{M} \right|^2 \quad (4.39)$$

$$c_4 = \sum_m \sum_n w_n |F_n\{i_m\}|^2 \quad (4.40)$$

Cost functions c_1 and c_2 can also be understood as the special cases of the cost functions c_3 and c_4 , where weight coefficients $w_n = 1$ are assigned to the harmonics within the frequency band of interest, while weight coefficients $w_n = 0$ are assigned to the harmonics outside of this frequency bands (i.e. to the harmonics, the reduction of which isn't required).

Note that all the cost functions defined in this section only consider the noise attenuation performance of the resulting GALS system. They don't take into account the interface cost. However, it would be possible to extend the cost functions defined here to consider the interface cost as well. One possibility to achieve this would be to include a multiplicative weight factor to an existing cost function, such that it increases with the number of interface signals.

4.3.2. Numerical evaluation of the methodology in MATLAB

The practical methodology described in the previous subsection has been evaluated by a set of numerical simulations in MATLAB. The simulations have been performed on a simplified model, as described in Chapter 3, where switching current profiles of digital blocks have been modeled by periodic triangular pulses. The pulses in this test case have random rise and fall times in ranges $0 \div 20\%$ and $0 \div 80\%$ of clock period, respectively, and a random current peak in the range $0 \div 1$ mA. There are 40 such blocks within the system, and they are to be distributed in 5 LSMs. Synchronous clock frequency is 10 MHz, while GALS clocking is plesiochronous, with offsets to the synchronous clock frequency being -4% , -2% , 1% , 3% and 5% . For each of the block profiles, spectrum harmonics have been calculated up to the 128^{th} harmonic. According to (4.21), for a theoretically perfect harmonic-balanced partitioning, a harmonic peak attenuation of 13.98 dB would be achieved.

The optimization procedure from the previous subsection has first been performed with cost function c_1 being applied, without any weight coefficients. This means that the optimization is applied on all of the available calculated harmonics (in this case, the first 128 of them). The result is shown in Fig. 4.10. In Fig. 4.10.a, the red dotted line represents the spectrum envelope of the initial synchronous system. If a “perfect partitioning” was possible, the GALS system which would result from it would have an envelope corresponding to the full red line. The optimization procedure, as described in the previous subsection, has resulted in a GALS system with a spectrum envelope shown as the full green line in Fig. 4.10.a. Spectral peak attenuation which has been achieved is shown in Fig. 4.10.b.

As it can be seen in the figure, an attenuation which is very close to the theoretical optimum of 13.98 dB has been achieved for the lower harmonics. As the harmonics get higher, the attenuation gets lower, and the achieved envelope (green) gradually deviates from the theoretically perfect one (red). As expected, the attenuation gets deteriorated the most in the regions around the “notches” (marked with arrows in Fig. 4.8.a). However, this deterioration around the “notches” is not as expressed as it could be expected from the discussion in subsection 4.2.4. This is thanks to the fact that here, although the current profile for each of the blocks is triangular, the resulting profile of the entire initial synchronous system, as well as the resulting profiles of the LSMs after galsification, are not triangular, but the sum of random triangular waveforms, as shown in Fig. 4.11. The notches of such waveforms are not as sharp as those in Fig. 4.8, and thus the deviation caused by the imperfections of the harmonic-balanced partitioning at “notch”-points aren’t as expressed.

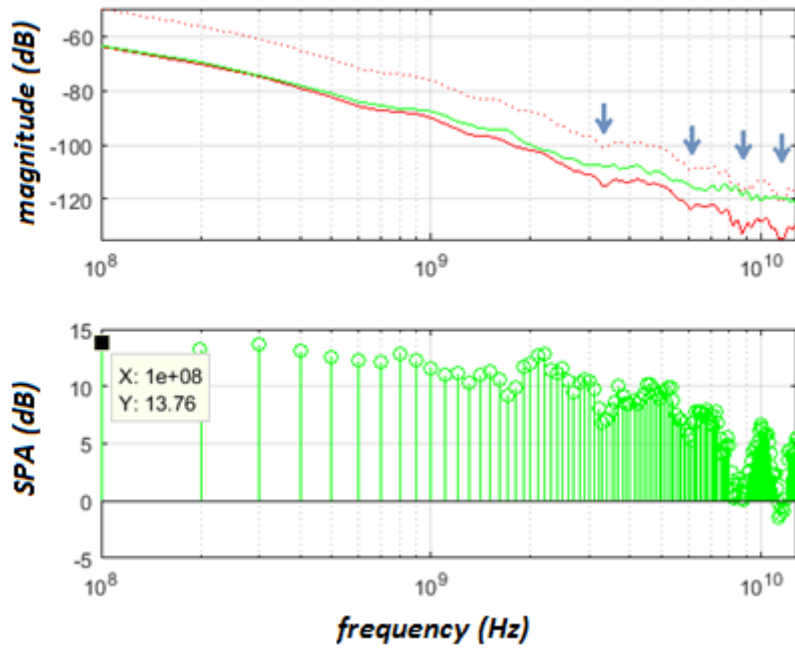


Figure 4.10 – (a) Spectrum envelopes of switching current for synchronous system (red dotted), for the ideal harmonic balanced partitioning (red) and for the partitioning obtained by optimization procedure with cost function c_1 on all harmonics (green); (b) spectral peak attenuations obtained (adapted from [P5])

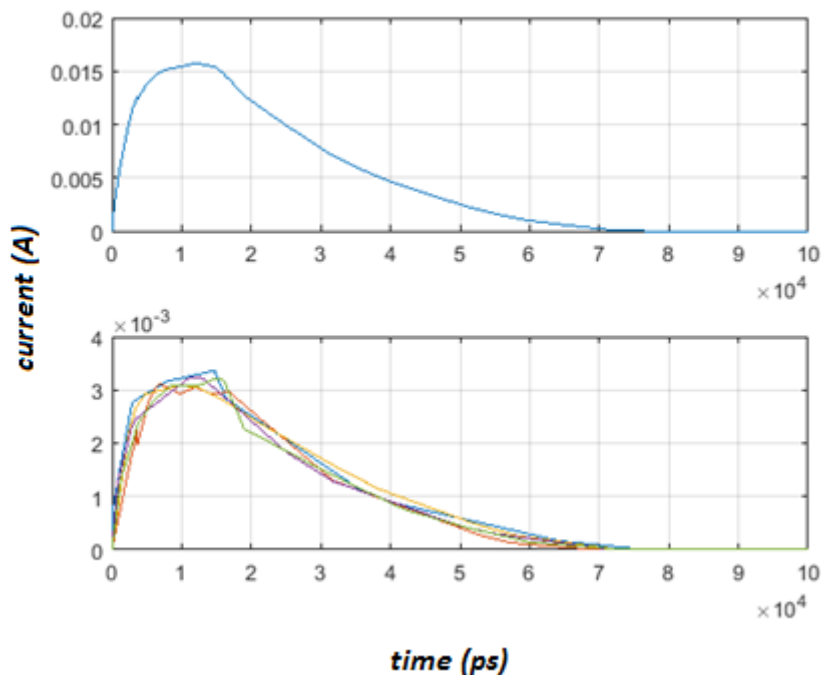


Figure 4.11 – Switching current waveform for the synchronous system (a) and for each of the GALS partitions (b) for the case from Fig. 4.10 (adapted from [P5])

In the second test, presented in Fig. 4.12, by the usage of weight coefficients the optimization has been focused on the frequency band between 1.5 GHz and 2.5 GHz. The weight

coefficients were chosen in the simplest band-selective way – weight coefficients for in-band harmonics have been set to 1, while the weight coefficients for the remaining harmonics have been set to 0. This way, it has been achieved that the attenuation in this frequency band is the closest to the theoretical optimum of 13.98 dB, although the harmonics in this part of the spectrum don't belong to the low harmonics. At this part of the spectrum, the envelope of the resulting GALS spectrum approaches the most to the envelope of the spectrum of “perfect partitioning”. In the rest of the spectrum, the attenuation is also achieved (except in a couple of high-frequency “notches”), but it's lower than in the optimized frequency band.

The results of the third test are presented in Fig. 4.13. Here, the frequency band between 5 GHz and 5.5 GHz has been optimized, containing even higher harmonics than the one in the previous test case. While the achieved attenuation in the selected frequency band is significantly higher than in the surrounding area, it's somewhat lower than in the previous test case. Thus, it can be concluded that the optimizing methodology from Section 4.3.1 delivers better results if the targeted optimization band is at lower frequencies.

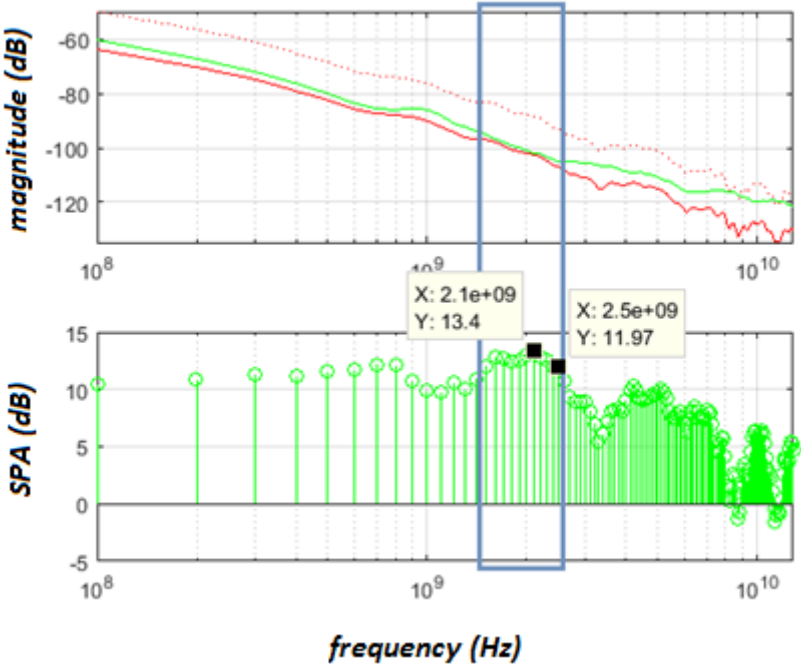


Figure 4.12 – Spectrum envelopes of switching current (a) for synchronous system (red dotted), for the ideal harmonic balanced partitioning (red) and for the partitioning obtained by optimization procedure with cost function c_1 on harmonics in frequency band [1.5 GHz, 2.5 GHz] (green); and spectral peak attenuations obtained (b). The markers are showing maximum and minimum attenuation achieved within the targeted band.

(adapted from [P5])

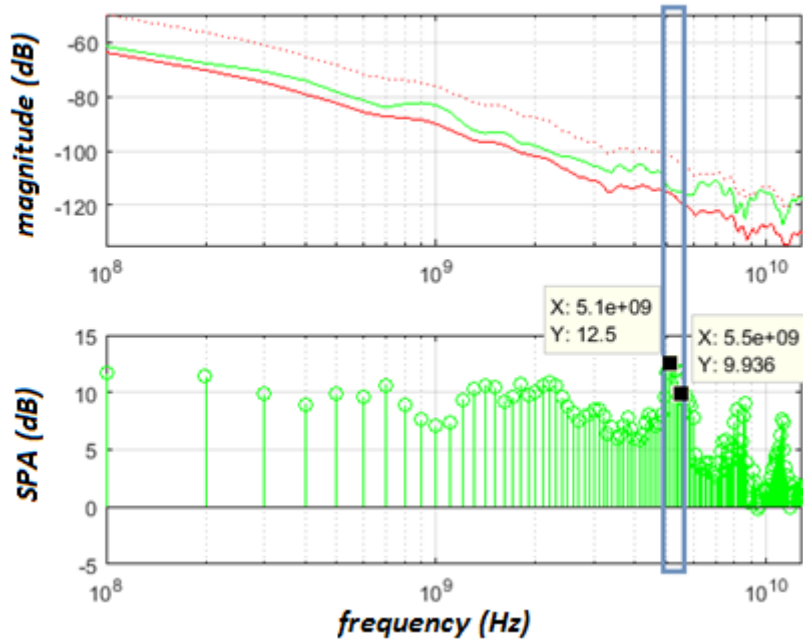


Figure 4.13 – Spectrum envelopes of switching current (a) for synchronous system (red dotted), for the ideal harmonic balanced partitioning (red) and for the partitioning obtained by optimization procedure with cost function c_1 on harmonics in frequency band [5 GHz, 5.5 GHz] (green); and spectral peak attenuations obtained (b). The markers are showing maximum and minimum attenuation achieved within the targeted band. (adapted from [P5])

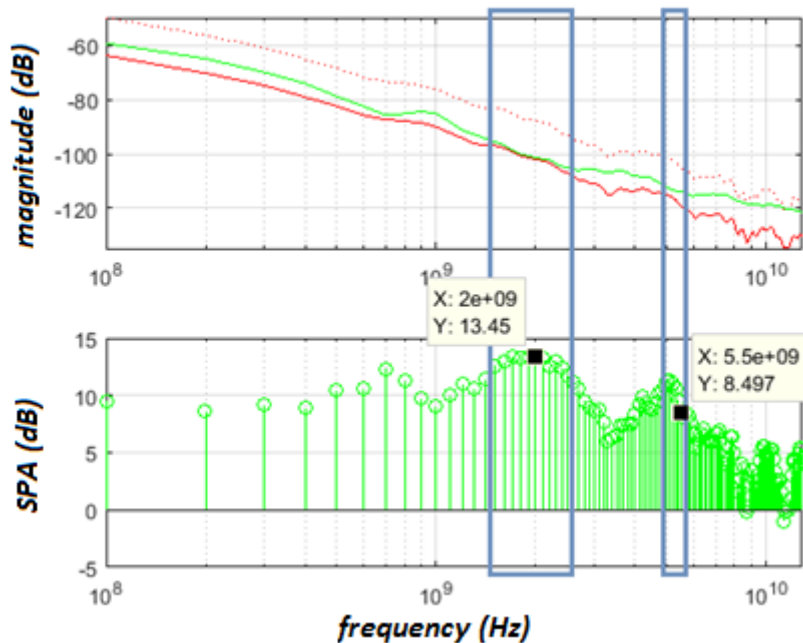


Figure 4.14 – Spectrum envelopes of switching current (a) for synchronous system (red dotted), for the ideal harmonic balanced partitioning (red) and for the partitioning obtained by optimization procedure with cost function c_1 on harmonics in frequency bands [1.5 GHz, 2.5 GHz] and [5 GHz, 5.5 GHz] (green); and spectral peak attenuations obtained (b). The markers are showing maximum and minimum attenuation achieved within the targeted bands. (adapted from [P5])

Finally, in the fourth test case, two frequency bands, 1.5 GHz ÷ 2.5 GHz and 5 GHz ÷ 5.5 GHz, have been simultaneously optimized. The result is presented in Fig. 4.14. As it can be noticed by comparing Fig. 4.14 to Fig. 4.12 and Fig. 4.13, the attenuation is slightly lower compared to the cases when each band is optimized without optimizing the other one. This is a consequence of a larger number of harmonics that have to be optimized – the more harmonics are to be optimized, the more difficult it gets to approach to the optimum solution. Additionally, similarly as from comparing the previous two test cases to each other, it can be noticed that better results are generally obtained at lower frequencies, where the switching current harmonics to be attenuated are higher.

4.3.3. Substrate noise spectral power attenuation for frequency bands containing more harmonics

In section 4.2.2 it has been shown that, besides providing the optimum harmonic peak attenuation, harmonic-balanced plesiochronous GALS partitioning also provides in-band power attenuation. More specifically, for switching current harmonics, it has been shown that harmonic peak attenuation of a targeted harmonic and spectral power attenuation in a narrow frequency band around the targeted harmonic are, respectively:

$$SPA_{I_{sw},n} = 20 \log \left(\frac{A_n}{A_n/M} \right) = 20 \log (M) \quad (4.41)$$

and:

$$PA_{I_{sw},n} = 10 \log \left(\frac{A_n^2}{A_n^2/M} \right) = 10 \log (M) \quad (4.42)$$

where A_n is the n^{th} harmonic spectral peak of the switching current, and M the number of GALS partitions, i.e. LSMs. Similarly, for substrate noise harmonics, harmonic peak attenuation of a targeted harmonic and spectral power attenuation in a narrow frequency band around the targeted harmonic are, respectively:

$$SPA_{V_{sub},n} = 20 \log \left(\frac{H_n A_n}{H_n A_n/M} \right) = 20 \log (M) \quad (4.43)$$

and:

$$PA_{V_{sub},n} = 10 \log \left(\frac{(H_n A_n)^2}{(H_n A_n)^2/M} \right) = 10 \log (M) \quad (4.44)$$

where H_n is the value of the PDN and substrate transfer function at the frequency of the n^{th} harmonic.

Note that, as a consequence of the plesiochronous approximation and the equal ground bounce approximation (which is valid for smaller systems), $SPA_{I_{sw},n} = SPA_{V_{sub},n}$ and $PA_{I_{sw},n} = PA_{V_{sub},n}$. In case of a single harmonic, both spectral peak attenuation and spectral power attenuation in a narrow band are the optimum ones which can be reached by GALS partitioning. In this chapter it will be analyzed whether harmonic balancing also provides optimal spectral power attenuation in the targeted frequency band.

First, let's observe the switching current spectrum. For a frequency band containing more harmonics, total in-band power for the initial synchronous design is:

$$P_{band, sync} = \sum_{n \in band} c A_n^2 = c \sum_{n \in band} A_n^2 \quad (4.45)$$

where c is a multiplicative constant. On the other hand, for the plesiochronous GALS system with M LSMs and perfect harmonic balancing at all in-band harmonics:

$$P_{I_{sw}, band, GALS} = \sum_{n \in band} M c \left(\frac{A_n}{M} \right)^2 = c \sum_{n \in band} \frac{A_n^2}{M} = c \frac{1}{M} \sum_{n \in band} A_n^2 \quad (4.46)$$

From (4.45) and (4.46), the total in-band harmonic power attenuation for the switching current is:

$$PA_{I_{sw}, band} = 10 \log \left(\frac{P_{I_{sw}, band, sync}}{P_{I_{sw}, band, GALS}} \right) = 10 \log \left(\frac{\sum_{n \in band} A_n^2}{\frac{1}{M} \sum_{n \in band} A_n^2} \right) = 10 \log(M) \quad (4.47)$$

However, harmonic peaks usually can't be perfectly balanced, so in practice, the partitioning which is the closest to the ideal harmonic balancing in the desired frequency band is found by an optimizing algorithm, which minimizes a cost function, as explained in Section 4.3.1. Now, let's assume an imperfect (i.e. a realistic, practically achievable) harmonic balancing. Let the magnitudes of the n^{th} switching current harmonics for the LSMs be $A_n/k_{n,1}$, $A_n/k_{n,2}$, ..., $A_n/k_{n,M}$, where $\sum_{i=1}^M (1/k_{n,i}) \rightarrow 1$, $\min\{k_{n,i}\} \leq M$ and $(\forall_i)(k_{n,i} \rightarrow M)$. Spectrum peak attenuation for the n^{th} harmonic is:

$$SPA_{I_{sw}, n} = 20 \log(\min\{k_{n,i}\}) \quad (4.48)$$

Harmonic power for the n^{th} harmonic of the GALS system is:

$$P_{I_{sw}, n, GALS} = c \sum_{i=1}^M \left(\frac{A_n}{k_{n,i}} \right)^2, \quad \min_{i \in [1, \dots, M]} \{k_{n,i}\} \leq M \quad (4.49)$$

Consequently, power attenuation for the narrow band around the n^{th} switching current harmonic is:

$$PA_{I_{sw}, n} = 10 \log \left(\frac{A_n^2}{\sum_{i=1}^M \left(\frac{A_n}{k_{n,i}} \right)^2} \right) = 10 \log \left(\frac{A_n^2}{A_n^2 \sum_{i=1}^M \left(\frac{1}{k_{n,i}^2} \right)} \right) = 10 \log \left(\sum_{i=1}^M \left(\frac{1}{k_{n,i}^2} \right) \right) \quad (4.50)$$

Note that if $(\forall_i)(k_{n,i} \rightarrow M)$ expression (4.50) converges to (4.42).

In case of a frequency band containing more harmonics, in-band harmonic power for the GALS system is:

$$P_{I_{sw}, band, GALS} = \sum_{n \in band} \left(c A_n^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right) = c \sum_{n \in band} \left(A_n^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right), \quad \min_{i \in [1, \dots, M]} \{k_{n,i}\} \leq M \quad (4.51)$$

and the total in-band harmonic power attenuation for the switching current is:

$$PA_{I_{sw}, band} = 10 \log \left(\frac{P_{I_{sw}, band, sync}}{P_{I_{sw}, band, GALS}} \right) = 10 \log \left(\frac{\sum_{n \in band} A_n^2}{\sum_{n \in band} \left(A_n^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right)} \right) \quad (4.52)$$

Note that if $(\forall_i)(k_{n,i} \rightarrow M)$ expression (4.52) converges to (4.47). This is actually the task of the optimization algorithm – it's designed so that the expression (4.52) converges to (4.47) to the largest possible extent.

Now, let's analyze the substrate voltage spectrum. The derivation is very similar to the one for the switching current spectrum. For a frequency band containing more harmonics, total in-band substrate noise spectral power for the initial synchronous design is:

$$P_{V_{sub,band,sync}} = \sum_{n \in band} c(H_n A_n)^2 = c \sum_{n \in band} (H_n A_n)^2 \quad (4.53)$$

while for the plesiochronous GALS system with M LSMs and perfect harmonic balancing at all in-band harmonics:

$$P_{V_{sub,band,GALS}} = \sum_{n \in band} M c \left(\frac{H_n A_n}{M} \right)^2 = c \sum_{n \in band} \frac{(H_n A_n)^2}{M} = c \frac{1}{M} \sum_{n \in band} (H_n A_n)^2 \quad (4.54)$$

From (4.53) and (4.54), the total in-band harmonic power attenuation for the switching current is:

$$PA_{V_{sub,band}} = 10 \log \left(\frac{P_{V_{sub,band,sync}}}{P_{V_{sub,band,GALS}}} \right) = 10 \log \left(\frac{\sum_{n \in band} (H_n A_n)^2}{\frac{1}{M} \sum_{n \in band} (H_n A_n)^2} \right) = 10 \log (M) \quad (4.55)$$

It can be noticed that, in the ideal case, spectral power attenuation is equal for substrate noise and for switching current. This is valid not just for a single harmonic, as already previously stated, but also for a frequency band containing more harmonics. From (4.55) and (4.47) it follows that:

$$PA_{V_{sub,band}} = PA_{I_{sw,band}} = 10 \log (M) \quad (4.56)$$

Now let's analyze the spectral power attenuation obtained in practice, when using an optimization algorithm. Let the magnitudes of the n^{th} switching current harmonics for the LSMs be $H_n A_n / k_{n,1}, H_n A_n / k_{n,2}, \dots, H_n A_n / k_{n,M}$, where $\sum_{i=1}^M (1/k_{n,i}) \rightarrow 1$, $\min \{k_{n,i}\} \leq M$ and $(\forall_i)(k_{n,i} \rightarrow M)$. Spectrum peak attenuation for the n^{th} harmonic is:

$$SPA_{V_{sub,n}} = 20 \log(\min\{k_{n,i}\}) \quad (4.57)$$

which is completely the same expression as (4.48). This confirms the very important previously mentioned statement that for a plesiochronous GALS system, spectral peak attenuation for any of the substrate noise harmonic peaks is the same as spectral peak attenuation for the corresponding switching current harmonic peak, regardless of whether the perfect balancing is achieved or not, i.e.:

$$SPA_{V_{sub,n}} = SPA_{I_{sw,n}} \quad (4.58)$$

Substrate noise harmonic power corresponding to the n^{th} harmonic of the GALS system is:

$$P_{V_{sub,band,GALS}} = \sum_{i=1}^M c \left(\frac{H_n A_n}{k_{n,i}} \right)^2 = c \sum_{i=1}^M \left(\frac{H_n A_n}{k_{n,i}} \right)^2, \quad \min_{i \in [1, \dots, M]} \{k_{n,i}\} \leq M \quad (4.59)$$

Consequently, power attenuation for the narrow band around the n^{th} substrate noise harmonic is:

$$\begin{aligned}
PA_{V_{sub},n} &= 10 \log \left(\frac{(H_n A_n)^2}{\sum_{i=1}^M \left(\frac{H_n A_n}{k_{n,i}} \right)^2} \right) = 10 \log \left(\frac{(H_n A_n)^2}{(H_n A_n)^2 \sum_{i=1}^M \left(\frac{1}{k_{n,i}^2} \right)} \right) \\
&= 10 \log \left(\sum_{i=1}^M \left(\frac{1}{k_{n,i}^2} \right) \right)
\end{aligned} \tag{4.60}$$

Once more, this is the same expression as the one obtained for the switching current:

$$PA_{V_{sub},n} = PA_{I_{sw},n} \tag{4.61}$$

which means that for a single harmonic, spectral power attenuation for substrate voltage is equal to spectral power attenuation for switching current, regardless of whether the perfect balancing is achieved or not. Similarly like for switching current, if $(\forall_i)(k_{n,i} \rightarrow M)$ expression (4.60) converges to (4.44).

Finally, in case of a frequency band containing more harmonics, in-band harmonic power for the GALS system is:

$$P_{V_{sub},band,GALS} = c \sum_{n \in band} \left((H_n A_n)^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right), \quad \min_{i \in [1, \dots, M]} \{k_{n,i}\} \leq M \tag{4.62}$$

and the total in-band harmonic power attenuation for the substrate noise is:

$$PA_{V_{sub},band} = 10 \log \left(\frac{P_{V_{sub},band,sync}}{P_{V_{sub},band,GALS}} \right) = 10 \log \left(\frac{\sum_{n \in band} (H_n A_n)^2}{\sum_{n \in band} \left((H_n A_n)^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right)} \right) \tag{4.63}$$

After comparing (4.63) with (4.52), an important conclusion can be drawn – for a frequency band containing more harmonics, in general case, the total in-band harmonic power attenuation which is practically achieved by an optimizing algorithm doesn't need to be equal for substrate noise and for switching current:

$$PA_{V_{sub},band} \neq PA_{I_{sw},band} \tag{4.64}$$

However, from $(\forall_n)(\sum_{i=1}^M (1/k_{n,i}) \rightarrow 1)$ it follows that $(\forall_n)(\sum_{i=1}^M (1/k_{n,i}^2) \leq 1)$, which means that simultaneously both:

$$\frac{\sum_{n \in band} A_n^2}{\sum_{n \in band} \left(A_n^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right)} \geq 1 \tag{4.65}$$

and

$$\frac{\sum_{n \in band} (H_n A_n)^2}{\sum_{n \in band} \left((H_n A_n)^2 \sum_{i=1}^M \frac{1}{k_{n,i}^2} \right)} \geq 1 \tag{4.66}$$

are valid. In other words, if in-band spectral power attenuation is achieved for the switching current, it is achieved for the substrate noise as well. Also, for $(\forall_i)(k_{n,i} \rightarrow M)$, both expressions (4.52) and (4.63) converge to (4.56), which means that if a certain set of coefficients $k_{n,i}$ is such that the attenuation of the in-band spectral power attenuation for the switching current converges to the

“perfect partitioning” result, the in-band spectral power attenuation for the substrate noise will converge to the “perfect partitioning” result as well.

The difference from (4.64) is visible only if the achieved result, although the optimum one for the targeted system, remains far from the result of the theoretical “perfect partitioning”. In this case, if a certain set of $k_{n,i}$ coefficients leads to the optimal in-band spectral power attenuation for the switching current (i.e. to the in-band spectral power attenuation for the switching current which is as close as possible to the result of a theoretically perfect partitioning), it may lead to a suboptimal spectral power attenuation for the substrate noise (i.e. some other partitioning might achieve better results regarding the in-band substrate noise spectral power reduction). A possible way to overcome this would include using weight coefficients based on a coarse estimation of PDN and substrate transfer function. This might be a topic for the future work.

4.4. A design flow for the GALS-based substrate noise reduction methodology

4.4.1. Incorporation of the methodology into a GALS design flow, and an application on a real design example

In Section 4.3, a practical methodology for performing harmonic-balanced plesiochronous GALS partitioning has been introduced. The next step, which is presented in this section, is embedding this methodology into a design flow and evaluating it on a realistic design example.

The methodology has been implemented and evaluated in the EMIAS CAD tool [201], which was developed in order to support the user in designing low noise digital circuits. The tool takes either an RTL source code or directly the hierarchical netlist of the design as an input. If only the RTL source code is available, the tool generates scripts for trial synthesis for obtaining the hierarchical netlists and design properties. Beside physical information about the design, these design properties also include the power waveforms, obtained from vector-based power analysis, which is performed by external tools, such as Synopsys PrimeTime. The power waveforms are obtained for each of the blocks contained within the system. As power waveforms are approximately proportional to switching current waveforms, they can be used for calculating the switching current spectral harmonic peaks for each of the blocks. Harmonic peaks calculated this way are then used as the input for the GALS partitioning algorithm, as described in Section 4.3.1. The algorithm is completely embedded into EMIAS. Besides balancing the harmonics, EMIAS additionally takes care about the connectivity between the blocks when performing the partitioning. In each iteration EMIAS calculates the power waveforms of the resulting partitions and the power waveform of the resulting GALS system, as well as the corresponding spectra. The procedure is repeated until the suitable partitioning is found. As a final step, the partitioned design has to be galsified by integrating asynchronous wrappers as interfaces between the LSMs. This step has to be performed manually, but it’s planned to automate it in future releases of the tool. The entire low-noise GALS-based design flow with EMIAS tool is presented in Fig. 4.15.

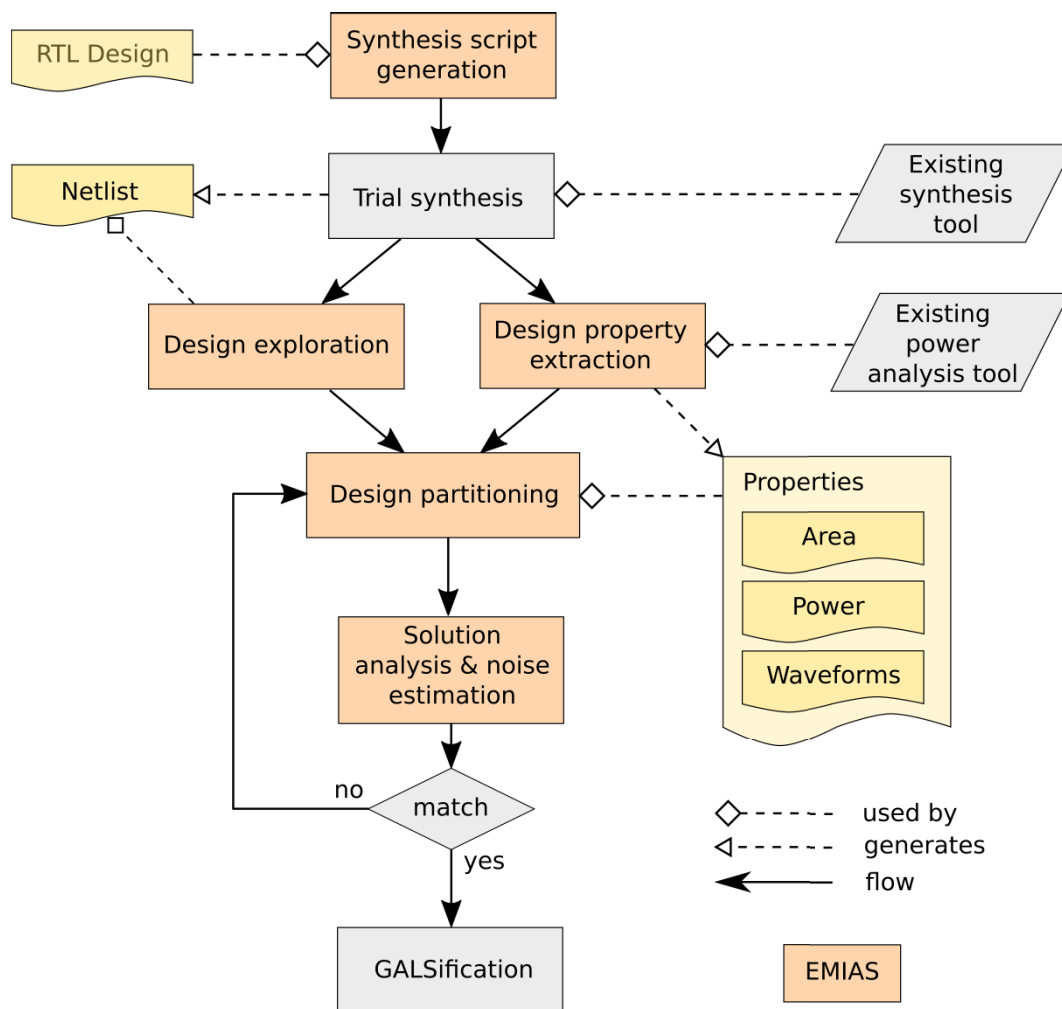


Figure 4.15 – Low noise GALS design flow methodology using EMIAS CAD tool (adapted from [P5])

The design example which was used for evaluation is a wireless sensor node consisting of a LEON2 32-bit microprocessor and three accelerator cores for cryptographic operations: SHA-1, AES and ECC. The AMBA bus system is connecting the components. Die area occupied by the design is 2.2 mm² in a 130 nm technology. The clock frequency of the design is 50 MHz, and it consumes 41.4 mW of power. The hierarchical netlist has been provided, and it has been used as the initial point in the flow.

At the beginning of the optimization procedure, 28 blocks have been identified within the hierarchical netlist. The intended number of partitions was five, and they were clocked plesiochronously at frequencies with offsets of -2.454 %, -1.254 %, 0 %, 1.25 % and 2.5 %. The optimization has been set simultaneously for two frequency bands – one containing only the first harmonic (i.e. 50 MHz), and another one containing the harmonics corresponding to the GSM-850 band (i.e. 800 MHz, 850 MHz and 900 MHz).

The results of the optimization procedure are presented in Table 4-I, and the spectra of synchronous and GALS systems are shown in Fig. 4.16 and Fig. 4.17, respectively.

Table 4-I - Harmonic peaks and harmonic power attenuations in dB (adapted from [P5])

Harmonic number		1	16	17	18
Harmonic frequency for the synchronous system (MHz)		50	800	850	900
Harmonic peak (dB)	Synchronous system (50 MHz)	42.26	41.98	20.16	41.68
	LSM0 (51.28 MHz)	-9.42	-7.31	-9.60	-3.83
	LSM1 (50.63 MHz)	37.74	34.90	17.72	32.62
	LSM2 (50.00 MHz)	25.08	34.57	9.20	34.16
	LSM3 (49.38 MHz)	10.68	14.29	-5.47	13.81
	LSM4 (48.78 MHz)	-36.21	-6.82	-18.87	-3.29
	GALS system (the highest peak among the peaks from LSM0-4)	37.74	34.90	17.72	34.16
Harmonic peak attenuation (dB)		4.52	7.08	2.44	7.52

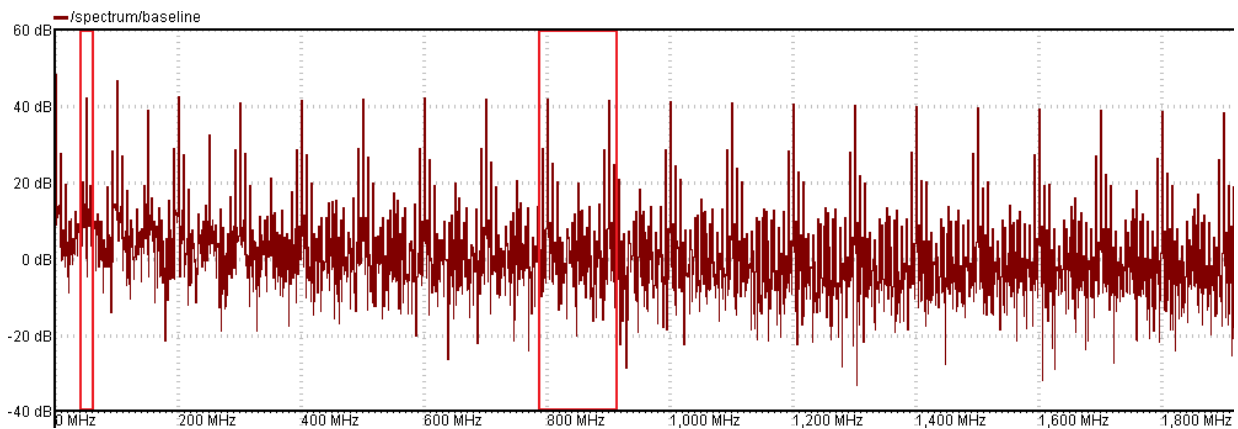


Figure 4.16 – Current spectrum for synchronous realization of the example design; bands to be optimized marked with red rectangles (adapted from [P5])

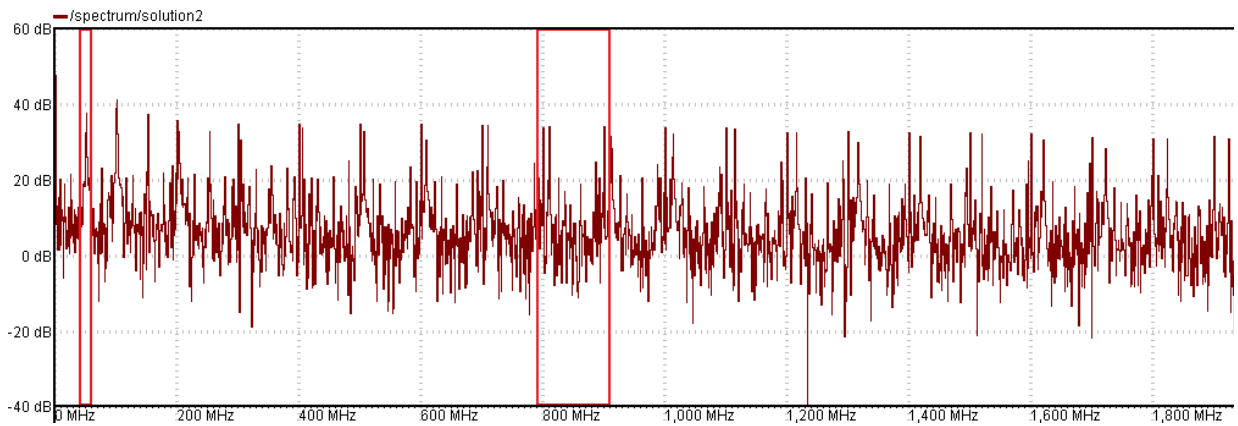


Figure 4.17 – Current spectrum for GALS realization of the example design; optimized bands marked with red rectangles (adapted from [P5])

Due to a small number of blocks, the low granularity of the design, large diversity in power among blocks, and a limited number of allowed combinations between them, the perfect harmonic balancing could not be achieved. Since some too large blocks couldn't be further partitioned, the harmonics couldn't be balanced well enough, as can be seen from the table. For each of the targeted frequencies it can be noticed that successful harmonic balancing has been achieved only for a couple LSMs per harmonic, not for all of them, as the methodology goal would be. Thus, the attenuations

obtained are lower than expected, and range between 2.44 dB and 7.52 dB. The attenuation is the smallest for the components which have the lowest power in the initial, synchronous system, which can be expected, as explained in Sections 4.2 and 4.3.

The impact of granularity as a limiting factor for the spectral peak attenuation which can be achieved by harmonic-balanced plesiochronous GALS partitioning methodology will be theoretically explained in the next section.

4.4.2. The impact of granularity

According to the equation (4.21), as M (the number of LSMs) increases, the spectral peak attenuation increases as well. However, there is an upper limit for choosing M , determined by the granularity of the design.

Granularity refers to the hierarchical structure of the design. If the design consists of many small hierarchical units, it's said that it has a high granularity. Hierarchical structure often puts a limit to the GALS partitioning. If a partitioning is made so that a natural hierarchical block is separated in two LSMs, it can result in a huge number of domain crossings between the LSMs, which may cause a large area and power overheads, which would also impact the noise performance. Additionally, if some hierarchical block appears as a predefined module (an IP core), this block can't be partitioned any further. If one block can't be partitioned any further, partitioning of the remaining blocks into smaller segments will not have an impact to the noise, because balancing with the largest block will not be possible, and this block will determine the highest spectral peak of the GALS design.

This is exactly what happened in the design example presented in Section 4.4.1. When EMIAS defines the blocks of the designs, it examines the hierarchy and stops at the leaf modules, which can be registers, glue logic, or predefined modules (IP cores). In the sensor node there were some large predefined modules which couldn't be further partitioned, so they appeared as unpartitionable blocks at the input of the partitioning algorithm. Thus, even though the number of LSMs was 5, the achieved attenuation wasn't equal to $20\log(5)$, because there were unpartitionable blocks with strong harmonics.

Let the n^{th} switching current harmonic for the entire synchronous system be $F_n\{i_{sys}\}$, and the switching current harmonic of the k^{th} block be $F_n\{i_{bk}\}$. In case that the targeted harmonics of block currents have the same or at least similar phase, the maximum partition number which makes sense for reducing the n^{th} harmonic is:

$$M_{max} = \text{ceil} \left\{ \frac{|F_n\{i_{sys}\}|}{\max\{|F_n\{i_{bk}\}|\}} \right\} \quad (4.67)$$

i.e. the ratio between the system harmonic and the maximum block harmonic, rounded to the next higher integer.

If there are blocks that have opposite current harmonic phase to the one of the block with the maximum harmonic, they can be combined with it within the same LSM and this way the large block harmonic can be further reduced, resulting in partition harmonic being smaller than the block

harmonic. Then, the above limit can be exceeded (i.e. in such cases the above limit is too conservative). However, most of the block harmonics, especially those at lower frequencies, usually have very similar phases, so in most cases the above limit can be used as an approximation to estimate the maximum number of partitions which would make sense. For the design example from Section 4.4.1, this would deliver $M_{max} = 3$, so the maximum theoretically achievable attenuation in this case would be 9.54 dB.

A phase difference can be introduced artificially, by applying a combined GALS clocking scheme, with two LSMs sharing the same plesiochronous frequency, but with opposite clock phases. This way, the limitation from (4.67) could be overcome. This idea can be a topic for future work and a further improvement of the methodology.

Note that a limitation similar to (4.67) also applies also to power balancing – if the non-partitionable block with the maximum power has power $\max\{P_{bk}\}$, and the synchronous system power is P_{sys} , the maximum number of partitions which makes sense would be:

$$M_{max} = \text{ceil} \left\{ \frac{P_{sys}}{\max\{P_{bk}\}} \right\} \quad (4.68)$$

The comparison of the harmonic-balanced plesiochronous GALS partitioning to the power-balanced plesiochronous GALS partitioning, as well as its comparison to the synchronous system-level methodologies for substrate noise reduction, will be given in the next section. Also, the main limitations of the methodology, one of them being the impact of granularity, will be summarized.

4.5. Comparison of harmonic-balanced plesiochronous GALS partitioning to the other system-level methodologies for substrate noise reduction

4.5.1. Main limitations of harmonic-balanced plesiochronous GALS partitioning

As discussed in Section 4.4.2, one of the drawbacks of the harmonic-balanced plesiochronous GALS partitioning methodology is the dependence of the applicability of the methodology on system granularity. Granularity which is too coarse limits the number of partitions M for which the attenuation $20\log(M)$ is still valid. However, by applying a combined GALS partitioning scheme, with introducing phase difference, this limitation could be effectively counteracted.

An additional drawback is that the methodology relies heavily on the correct estimation of the switching current profile, i.e. on the precision of the CAD tool providing the estimation. These estimations are currently less reliable than the much more straight-forward power estimations, on which the power-balanced plesiochronous GALS partitioning is based. However, CAD tools are steadily getting better in estimating switching current profiles. Nowadays there are even tools specialized for such estimations, such as Teklatech's FloorDirector™ [161]. With further development in this field, the estimations of the switching current profiles will become increasingly reliable, which will consequently contribute to reliability of harmonic-balanced plesiochronous GALS partitioning methodology.

It is also important to mention that there's a trade-off between the optimal noise reduction and the number of interfaces between the blocks. Partitioning which would only take noise reduction into account might result in an excessively large number of interfaces between the LSMs. This would make the galsification costly in terms of power and area. Thus, in practice, it would be important to take into account the number of interfaces between the LSMs of the resulting GALS system, when determining the optimal partitioning. This could be easily incorporated into the optimizing algorithm by accordingly modifying the cost function, as mentioned at the end of Section 4.3.1.

Finally, harmonic-balanced plesiochronous GALS partitioning methodology has been derived by applying the equal ground bounce application, which means that its applicability is limited to systems where package parasitics are dominant compared to on-chip parasitics. Extending the methodology to larger-scale systems is essentially possible, but it would require applying much more sophisticated models of on-chip power-delivery network. This could be an interesting topic for future work.

Beside these drawbacks, the methodology also has various advantages compared to other system-level methodologies for substrate noise reduction. In the following two subsections, the methodology will be compared to the existing system-level methodologies for substrate noise reduction – first to power-balanced plesiochronous GALS partitioning methodology, and then also to the synchronous methodologies – switching current shaping and clock modulation.

4.5.2. Comparison to power-balanced plesiochronous GALS partitioning

As stated in Section 4.2.2, the spectral peak attenuation achieved with harmonic-balanced plesiochronous GALS partitioning at a targeted harmonic is equal to the spectral peak attenuation at the fundamental and lower harmonics achieved with power-balanced plesiochronous GALS partitioning. In this section, the relation between these two GALS partitioning methodologies will be analyzed.

The power-balanced plesiochronous GALS partitioning methodology [16][160] was derived by taking an assumption that the switching current waveforms of the partitions can be modeled as periodic triangular pulses. The n^{th} harmonic of such a waveform is [160]:

$$F_n\{i_m\} = \frac{I_{pm}}{j2\pi n} \left(\text{sinc}(\pi n f_m t_{rm}) - \text{sinc}(\pi n f_m t_{fm}) e^{-j\pi n f_m (t_{rm} + t_{fm})} \right) e^{-j\pi n f_m t_{rm}} \quad (4.69)$$

where t_{rm} is the rise time of the switching current profile, t_{fm} the fall time of the switching current profile, f_m the clock frequency, I_{pm} the peak current and i_m the current profile (i.e. time-domain waveform) of the m^{th} LSM.

Accordingly, when an initial synchronous system is partitioned into M LSMs, it can be written:

$$F_n\{i\} = \sum_{m=1}^M F_n\{i_m\} \quad (4.70)$$

where $F_n\{i\}$ represents the n^{th} switching current harmonic of the initial synchronous system, and i the current profile of the initial synchronous system. Spectral peak attenuation of the n^{th} harmonic, per definition, can be expressed as:

$$SPA_n = 20 \log \frac{|F_n\{i\}|}{\max\{|F_n(i_m)|\}} = 20 \log \frac{|\sum_{m=1}^M F_n\{i_m\}|}{\max\{|F_n(i_m)|\}} \quad (4.71)$$

Following the harmonic-balanced plesiochronous GALS partitioning methodology, by choosing:

$$(\forall_m) \left(F_n\{i_m\} = \frac{F_n\{i\}}{M} = \frac{\sum_{m=1}^M F_n\{i_m\}}{M} \right) \quad (4.72)$$

the n^{th} harmonic will be optimally attenuated, with the attenuation of $20 \log(M)$.

On the other hand, as stated in [16][160], by applying the power-balanced plesiochronous GALS partitioning methodology, i.e. by selecting:

$$(\forall_m) \left(P_m = \frac{P_{sync}}{M} = \frac{\sum_{m=1}^M P_m}{M} \right) \quad (4.73)$$

for low harmonics, satisfying the condition:

$$nf_0 < \min \left(\frac{1}{\pi t_{rm}}, \frac{1}{\pi t_{fm}} \right) \quad (4.74)$$

the following attenuation is achieved:

$$SPA_n = 20 \log M + O(\Lambda) \quad (4.75)$$

where $O(\Lambda)$ is a factor dependent on current profile. For $M \gg 1$, it can be proven [16][160] that this factor is negligible, so the spectral peak attenuation achieved by power-balanced plesiochronous GALS partitioning can also be approximated as $20 \log(M)$.

In other words, power-balanced plesiochronous GALS partitioning can be viewed as an approximation of harmonic-balanced plesiochronous GALS partitioning, which is valid for lower harmonics and for a sufficiently large number of partitions.

There's also one more connection between the two methodologies. Let's assume that the initial synchronous system consists of N blocks. If the current profile of each block is modeled as a periodic triangular pulse, similar to (4.69) its n^{th} harmonic can approximately be expressed as:

$$F_n\{i_{bk}\} = \frac{I_{pbk}}{j2\pi n} \left(\text{sinc}(\pi n f_0 t_{rbk}) - \text{sinc}(\pi n f_0 t_{fbk}) e^{-j\pi n f_0 (t_{rbk} + t_{fbk})} \right) e^{-j\pi n f_0 t_{rbk}} \quad (4.76)$$

where t_{rbk} is the rise time of the switching current profile, t_{fbk} the fall time of the switching current profile, I_{pbk} the peak current and i_{bk} the switching current profile (i.e. time-domain waveform) of the k^{th} block within the system, and f_0 the clock frequency, very close to each of the later LSM frequencies f_m . Since the power of a periodic triangular waveform is:

$$P_{bk} = \frac{V_{DD} I_{pbk} (t_{rbk} + t_{fbk}) f_0}{2} \quad (4.77)$$

with V_{DD} being the supply voltage, the expression (4.76) can be written as:

$$F_n\{i_{bk}\} = -j \frac{P_{bk}}{V_{DD}} \frac{(\text{sinc}(\pi n f_0 t_{rbk}) - \text{sinc}(\pi n f_0 t_{fbk}) e^{-j\pi n f_0 (t_{rbk} + t_{fbk})}) e^{-j\pi n f_0 t_{rbk}}}{\pi n (t_{rbk} + t_{fbk}) f_0} \quad (4.78)$$

which can further be expressed as:

$$F_n\{i_{bk}\} = c P_{bk} \psi_n(\lambda_{bk}) \quad (4.79)$$

where c is a multiplicative constant, P_{bk} the power of the k^{th} block, and $\psi_n(\lambda_{bk})$ the factor which is dependent on the parameters of the current profile of the block (rise time t_{rbk} , fall time t_{fbk} and the switching frequency), and on the order of harmonic which is observed.

If for the observed harmonic all of the blocks have similar current profiles, so that the factor $\psi_n(\lambda_{bk})$ is approximately the same for each of the blocks, balancing the harmonics becomes equivalent with balancing the power P_{bk} . In this case, harmonic-balancing is also nearly equivalent to power-balancing, and these two methods would give very similar results.

To conclude, harmonic-balanced plesiochronous GALS partitioning can be approximated by power-balanced plesiochronous GALS partitioning in the following two cases:

- If only the harmonics of low order are targeted, and if the number of partitions is large
- For any harmonic or any number of partitions, if all of the blocks within the system have a similar switching current profile

Harmonic-balancing is more general and more flexible than the power-balancing. It's applicability doesn't depend on the current profiles of the individual blocks or on the number of partitions. Additionally, and contrary to power-balancing, it is also applicable for higher harmonics, and it can be optimized for a specific frequency band, or even for several frequency bands simultaneously.

Power balancing, on the other hand, has an important advantage of being much simpler to implement. No optimizing algorithm is required. Also, estimating power is much simpler and can be done much more precisely than estimating a switching current profile, which makes the power-based partitioning methodology more reliable.

Thus, in the two cases listed above, when harmonic-balancing can be approximated by power-balancing, a simpler power-balanced plesiochronous GALS partitioning methodology is recommended. For substrate noise attenuation, possible application would include a system where VCOs (which are sensitive only to low-frequency noise) are the only analog victims, or when there are victims on various frequencies, but all of the blocks within the system have similar switching current profiles.

If the blocks within the system have more diverse switching current profiles, and if there are also victims in the system which are sensitive to noise at higher frequency bands (such as LNAs), power-balanced plesiochronous GALS partitioning can't provide the desired attenuation at high frequencies. Thus, such cases can be recommended for application of harmonic-balanced plesiochronous GALS partitioning.

4.5.3. Comparison to other system level methodologies for substrate noise reduction

In Chapter 2, the existing synchronous system-level methodologies for substrate noise reduction have been presented in detail. These methodologies include switching current shaping (SCS) and clock modulation (CM). In this subsection, they will be compared to the harmonic-balanced plesiochronous GALS partitioning.

Spectral peak attenuation achieved by switching current shaping can be expressed as [159][160][P7]:

$$SPA_n = 40 \log \left(\frac{1}{\lambda_0} \right) \quad (4.80)$$

where λ_0 is the relative width of the current pulse for the initial system. The expression (4.80), however, is valid only for higher harmonics, which fulfill the condition:

$$nf_0 > \max \left(\frac{1}{\pi t_r}, \frac{1}{\pi t_f} \right) \quad (4.81)$$

with t_r and t_f being the rise time and fall time of the switching current profile of the initial system, and f_0 its clock frequency. For low harmonics, on the other hand, the dependence of the spectral peaks on the relative pulse width is marginal, and thus the spectral peak attenuation which can be achieved by this methodology is also marginal [160][P7].

Note that even for higher frequencies, the attenuation which can be achieved by this methodology, as per (4.80), strongly depends on the current profile of the initial system. In case that the initial system already has λ_0 which is close to 1, this methodology produces only marginal attenuation even for higher harmonics.

A special type of SCS is two-phase clocking [159]. By applying this methodology, for low switching current harmonics, satisfying the condition:

$$nf_0 < \min \left(\min \left(\frac{1}{\pi t_{r1}}, \frac{1}{\pi t_{f1}} \right), \min \left(\frac{1}{\pi t_{r2}}, \frac{1}{\pi t_{f2}} \right) \right) \quad (4.82)$$

with t_{r1} and t_{r2} being the rising switching current edges of the two clock domains and t_{f1} and t_{f2} being the falling switching current edges of the two clock domains, the spectral peak attenuation can be expressed as:

$$SPA_n = 20 \log \left| \frac{Q_1 + Q_2}{Q_1 - Q_2} \right| \quad (4.83)$$

if n is odd, while it becomes negligible if n is even. Q_1 and Q_2 in the equation (4.84) represent the areas below the single triangular pulse in the periodic triangular pulse model (i.e. the electric charge transported in one switching cycle) in the first and the second clock domain, respectively.

Similar to plesiochronous GALS, SCS also achieves spectrum power attenuation, with spectral peaks being replaced by lower spectral peaks of the same width. Current modulation, on the other hand, replaces each spectral peak with a lower but wider one, thus achieving spectral peak attenuation, but without attenuation of spectral power.

Spectral peak attenuation achieved by current modulation can approximately be calculated by using the Carson's law [176]:

$$SPA_n \approx 10 \log(2(n\beta + 1)) \quad (4.84)$$

with β being the modulation index of the frequency modulation of the clock, defined as:

$$\beta = \frac{\Delta f}{f_m} \quad (4.85)$$

with f_m being the clock modulation frequency, and Δf the peak deviation from the central clock frequency.

At low frequencies, i.e. for the harmonic orders under the cutoff limit $n_{cutoff} = 1/\beta$, the achieved attenuation is rather low. As the order of the harmonic to be attenuated gets higher, the achieved attenuation gets higher as well. The upper limit for this is the overlapping frequency, at which the two adjacent harmonic peaks become so wide that they start to overlap [160]:

$$n_{overlap} = \frac{1}{\beta} \left(\frac{f_0}{2f_m} - 1 \right) - \frac{1}{2} \quad (4.86)$$

Spectral peak attenuation achieved between the two limits can reach high values. Increasing β can improve the spectral peak attenuation, but it would also lower the overlapping limit from (4.86), so there's a trade-off.

In the following table, the main properties of the two synchronous substrate noise reduction methods are compared to both the power-balanced plesiochronous GALS partitioning methodology and the harmonic-balanced plesiochronous GALS partitioning methodology. Two-phase clocking, as a special case of switching current shaping, is separately included in the comparison.

Table 4-II - Harmonic peaks and harmonic power attenuations in dB (SCS - switching current shaping, 2Ph - two phased clocking, CM - clock modulation, PB-pG - power-balanced plesiochronous GALS partitioning, HB-pG - harmonic-balanced plesiochronous GALS partitioning)

	<i>Maximum spectral peak attenuation</i>		<i>Maximum in-band spectral power attenuation</i>		<i>Deliberate frequency-band selective optimization</i>
	<i>Low harmonics</i>	<i>High harmonics</i>	<i>Low harmonics</i>	<i>High harmonics</i>	
<i>SCS</i>	marginal	$40 \log(1/\lambda_0)$	marginal	$20 \log(1/\lambda_0)$	no
<i>2Ph</i>	only odd harmonics: $20 \log \left \frac{Q_1 + Q_2}{Q_1 - Q_2} \right$	not targeted	only odd harmonics: $10 \log \left \frac{Q_1 + Q_2}{Q_1 - Q_2} \right$	not targeted	no
<i>CM</i>	$10 \log(2(n\beta + 1))$ low due to low $n\beta$	$10 \log(2(n\beta + 1))$, for $n < n_{overlap}$	no	no	no
<i>PB-pG</i>	$20 \log(M)$	not targeted	$10 \log(M)$	not targeted	no
<i>HB-pG</i>	$20 \log(M)$	$20 \log(M)$	$10 \log(M)$	$10 \log(M)$	yes

It can be concluded that at low frequencies harmonic-balanced plesiochronous GALS partitioning can significantly outperform both switching current shaping and clock modulation. Note, however, that if attenuation is only required at low frequencies, harmonic-balanced plesiochronous

GALS partitioning can be replaced by power-balanced plesiochronous GALS partitioning. The only synchronous methodology also achieving high attenuation at low frequencies is two-phased clocking, however only at odd harmonics – at even harmonics, no attenuation is achieved.

Two-phased clocking and power-balanced plesiochronous GALS partitioning don't target attenuation at higher harmonics. On the other hand, the remaining three methodologies – switching current shaping, current modulation and harmonic-balanced plesiochronous GALS partitioning – can reach high spectral peak attenuation at high frequencies.

However, for switching current shaping, the attenuation which can be achieved is strongly dependent on the switching current profile of the initial system. If the system has a low critical slack and if the switching is not strongly concentrated at the beginning of the clock period, switching current triangular model will have λ_0 close to 1, and consequently no large attenuation can be achieved. Harmonic-balanced plesiochronous GALS partitioning doesn't have such a limitation and the attenuation it can achieve is independent on the slack and the switching current profile of the initial system.

Clock modulation, on the other hand, can reach high spectral peak attenuation only up to the overlapping limit. As stated in Section 4.2.3, harmonic-balanced plesiochronous GALS partitioning also has a conditional overlapping limit. This limit, however, can be avoided by a careful frequency planning. Another advantage of harmonic-balanced plesiochronous GALS partitioning compared to clock modulation is that, besides providing spectral peak attenuation, it also provides spectral power attenuation, which is not the case for clock modulation.

Finally, neither switching current shaping nor clock modulation can provide an attenuation which would be specifically optimized for a deliberate set of frequency-bands. Switching current shaping provides constant attenuation for all of the higher harmonics, without any optimization possibility, while at low frequencies it provides only marginal attenuation. If a single harmonic in the middle frequency range is targeted, there's a possibility to optimize noise reduction by positioning the notch of the spectrum envelope to match exactly this harmonic. However, this strongly depends on the switching current waveform of the initial system – having λ_0 close to 1 would significantly reduce the possibility of deliberate notch positioning. Clock modulation provides attenuation which gets better with higher frequency and is optimal for the highest harmonic under the overlapping limit. If only one harmonic is targeted, it is possible to optimize clock modulation so that this harmonic is exactly the one under the overlapping limit, i.e. the one with the highest attenuation without overlapping. Targeting a wider frequency band or separate frequency bands, however, is possible only with harmonic-balanced plesiochronous GALS partitioning.

4.6. Summary

In this chapter, a new system-level substrate noise reduction methodology called harmonic-balanced plesiochronous GALS partitioning has been presented. The methodology achieves substrate noise reduction by converting an initially synchronous digital system into a plesiochronous GALS system.

First, the requirements for a generally applicable substrate noise reduction methodology have been analyzed. It has been concluded that a generally applicable substrate noise reduction methodology would require spectral peak attenuation not only of the lower harmonic peaks, but also of the harmonic peaks of a higher order. Also, the possibility to optimize deliberate frequency bands is desired.

As a basis for the methodology, a plesiochronous GALS clocking scheme has been chosen. With this clocking scheme, the speed of the system and the processing capability of each block are maintained to the larger extent compared to the initial synchronous system. Also, for a plesiochronous clocking scheme, the spectral peak attenuation of the switching current harmonic peaks is equal to the spectral peak attenuation of the substrate noise harmonic peaks. This enables operating with switching current parameters instead of substrate voltage parameters, which is beneficial because switching current is much easier to estimate.

Further, it has been mathematically proven that the optimum spectral peak attenuation is achieved if the harmonic peaks of the LSM switching currents in the resulting GALS system are balanced to each other, both in phase and magnitude. Hence, the methodology has been named harmonic-balanced plesiochronous GALS partitioning. If the partitioning is performed this way, it theoretically results in spectral peak attenuation of $20 \log(M)$, where M is the number of LSMs in the resulting GALS system.

This result was derived by using a plesiochronous approximation, which states that the harmonic peaks of the switching current spectrum estimated at the synchronous frequency remain the same at LSM frequencies in the resulting GALS system. The plesiochronous approximation has been separately analyzed, and it has been shown that it doesn't impact the relevant part of the spectrum. Additionally, conditions for avoiding having an upper applicability limit due to harmonic overlapping have been defined.

In order to be able to practically apply the methodology, a suitable partitioning procedure, based on the simulated annealing algorithm, has been developed. The goal of the procedure is to distribute the hierarchical blocks of the initial synchronous system into the partitions (LSMs) of the resulting GALS system, so that the result is as close as possible to the theoretical "perfect partitioning", where all the switching current harmonics within the targeted frequency band would be perfectly balanced among the LSMs.

The procedure has further been evaluated numerically in MATLAB, with block current profiles represented by the periodic triangular model from Chapter 3. The applicability of the methodology for spectral peak attenuation at higher order harmonics has been demonstrated, as well as the possibility of dedicated frequency band targeting, and especially the possibility of simultaneously targeting more frequency bands. The evaluation results are close to the theoretical prediction in all of the test cases, especially at lower frequencies and when targeting narrower frequency bands.

The property of the total harmonic spectral power reduction within a targeted frequency band containing more harmonics has further been theoretically analyzed. Unlike the spectral peak

attenuation, which is the same for switching current harmonic peaks and for the corresponding substrate noise harmonic peaks, the in-band spectral power attenuation doesn't necessarily need to be the same for these two switching noise types. However, when the harmonic-balanced plesiochronous GALS partitioning is applied, it has been shown that the in-band spectral power attenuation converges to the same value for the switching current and for the substrate noise. The closer the partitioning is to the perfect harmonic balancing, the closer the two spectral power attenuations get. Only if the partitioning is far from the perfectly harmonic-balanced one, the difference in spectral power attenuation between the two noise types can become noticeable.

As the next step, the methodology has been incorporated into a low-noise GALS design flow. The partitioning algorithm has been embedded within the EMIAS CAD tool. The methodology has been evaluated on a real design example – a wireless sensor node. Two frequency bands have been simultaneously targeted for noise reduction – a frequency band containing only the fundamental, at 50 MHz, and the GSM-850 frequency band (800 MHz – 900 MHz). The harmonic peaks in the targeted frequency bands have indeed been reduced, thus proving the concept of noise reduction by harmonic-balanced plesiochronous GALS partitioning. The attenuation of up to 7.52 dB has been achieved, which, however, is lower than theoretically expected. The impact of granularity has been identified as the reason for achieving a lower attenuation than expected, and the limit set by granularity has subsequently been analyzed.

Finally, the methodology has been compared to the other system-level substrate noise reduction methodologies. The dependence on granularity, the relying on switching current profile estimation and the limitation of applicability to small systems with dominant package parasitics have been identified as the main drawbacks of the methodology.

The methodology, however, also has significant advantages compared to the synchronous system-level substrate noise reduction methodologies (SCS and CM):

- At lower frequencies, it outperforms both SCS and CM
- The attenuation is independent on the slack and the switching current profile of the initial system (an advantage compared to SCS)
- The upper harmonic order limit set by overlapping can be avoided (an advantage compared to CM)
- The in-band spectral power attenuation is provided (an advantage compared to CM)
- It has the possibility to specifically target one or more deliberate frequency bands (an advantage compared to both SCS and CM)

Additionally, the harmonic-balanced plesiochronous GALS partitioning (HB-pGp) methodology has been compared to the power-balanced plesiochronous GALS partitioning (PB-pGp). It has been shown that PB-pGp can be viewed as a special case of HB-pGp if the harmonics of lower order are targeted for reduction. The two methodologies are also equivalent if all the blocks within the system are similar to each other in terms of switching current profile. However, contrary to PB-pGp, HB-pGp can target harmonics of higher order, and it can also specifically target harmonics within one or more frequency bands; so, it is more general and it covers more use cases.

PB-pGp is much simpler to implement. Thus, in special cases when HB-pGp can be approximated by PB-pGp (i.e. for lower order harmonics, and if the blocks within the system have similar current profiles), PB-pGp should preferably be used. In more general cases when the blocks within the system are diverse and attenuation of higher-order harmonics is needed, HB-pGp should be used instead.

The harmonic-balanced plesiochronous GALS partitioning methodology, presented in this chapter, is applicable on both kinds of substrates, and it doesn't use the isolating property of lightly-doped substrates. In order to make use of this useful property of lightly-doped substrates, another GALS-based methodology has been developed. This methodology is presented in the following chapter.

5. Substrate noise reduction methodology based on a combination of GALS partitioning and physical placement

5.1. Introduction

In the previous chapter, harmonic-balanced plesiochronous GALS partitioning methodology for substrate noise reduction has been presented. This methodology, however, is a purely system-level methodology, and it doesn't make use of the main advantage of lightly doped substrates – its isolating property.

In the methodology which will be presented in this chapter, the isolating property of lightly doped substrates is used in combination with GALS design approach to reduce substrate noise in lightly doped substrates. Thus, the methodology presented here is a hybrid methodology, consisting of a physical aspect and a system-level aspect. Similar to the methodology from Chapter 4, the methodology is developed for application in a small system where package parasitics are dominant compared to on-chip parasitics. This enables using the equal ground bounce approximation during analysis.

The physical aspect of the methodology is based on a separation of power domains within a digital aggressor circuitry and their physical placement, and it is presented in Section 5.2. The separation of power domains represents an important difference compared to the method presented in the previous chapter, where there was just a single power domain within the digital circuitry, and consequently, due to the equal ground bounce approximation, just a single aggressor in the model. Here, there will be as many aggressors as there are power domains, since separate power domains can have different values of ground bounce. This creates a new degree of freedom in optimizing the noise, which will be exploited by the methodology presented in this chapter.

In Section 5.3, the basic methodology from Section 5.2 is extended by including the system level aspect – the galsification. In order to simplify the initial theoretical analysis, the impact of decoupling capacitance has been neglected in Sections 5.2 and 5.3. This impact is analyzed in Section 5.4.

The sections 5.2-5.4 deal with a theoretical case where granularity is infinite and a “perfect partitioning” is possible. In order to make the methodology applicable for a realistic system with finite granularity, a partitioning algorithm has been developed, and further evaluated by numerical simulations in MATLAB. This algorithm and its numerical evaluation are presented in Section 5.5.

Further, in Section 5.6, some special cases of the methodology are derived. The partitioning procedure is significantly simpler for these special cases, and the methodology is much easier to implement.

In Section 5.7, the methodology is compared to the existing methodologies. Also, the possibility of combining this methodology with other methodologies is discussed. Finally, in Section 5.8, the conclusions are drawn.

5.2. Substrate noise reduction by power domain separation in the digital aggressor

As mentioned in Chapter 1, lightly doped substrates are mostly homogenous, consisting of a high-resistivity bulk. This high resistivity enables isolation by separation – the larger the distance between two blocks, the larger the substrate impedance between them. If these two blocks are an aggressor and a victim, increasing the distance between them will reduce the noise coupling. This has been used in many physical noise reduction methodologies, as already shown in Section 2b. Note, however, that an important prerequisite for this approach to work is the absence of a conductive backplane (backside). At smaller distances, the conductive backplane doesn't make a large difference, but at larger distances, it has a similar effect as a high-conductivity bulk of an epi-type substrate, making noise coupling almost independent of physical placement [18].

The start point of the methodology, i.e. the initial system to be optimized for substrate noise at the victim's position, is a mixed signal integrated circuit (MSIC) with an analog victim and a digital synchronous circuit as an aggressor. As explained in section 2b, for MSICs it's common to have separated supply and ground lines for analog and digital modules. Common supply or ground lines would otherwise represent a direct path for noise coupling from a digital aggressor to an analog victim. For the initial system, it is assumed that the digital aggressor contains a single power domain. The on-chip supply and ground networks of the digital circuitry are biased to a stable off-chip supply and ground via N_d supply pads and N_d ground pads, respectively. Similar to the previous chapter, it is also assumed that the package parasitics are dominant to the on-chip PDN parasitics. Hence, the equal ground bounce approximation can be applied, and the digital circuitry in the initial system can be viewed as a single aggressor. If the parasitic package impedance of a single power pad (including the corresponding bonding wire) is Z_p' , the total impedance of package parasitics connected to the aggressor (either on the supply side or on the ground side) is $Z_p = Z_p'/N_d$. The underlying substrate of the MSIC is, as already stated, lightly doped, and the conductive backplane is not present. At the aggressor side, the substrate is biased by the digital ground network, while at the victim side, the substrate is biased by a separate analog PDN, with the total parasitic impedance Z_{pa} .

This initial system can be represented by the equivalent circuit shown in Fig. 5.1, with Z_{seff} being the extracted substrate impedance between the digital aggressor and the analog victim. For the time being, the capacitance of the digital circuitry and the possible intentional decoupling capacitance will be neglected – the impact of those will be taken into account in Section 5.4.

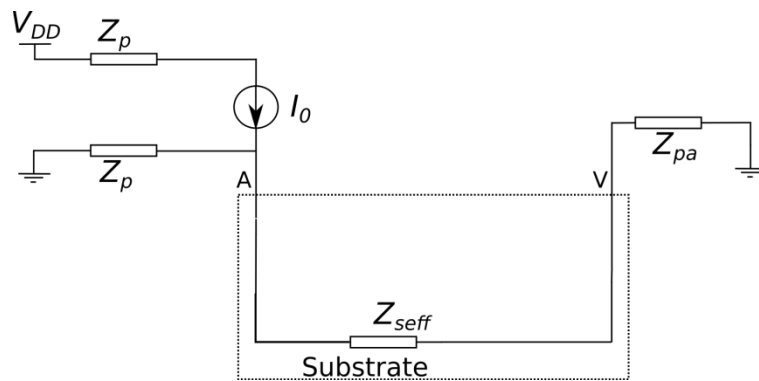


Figure 5.1 – Equivalent circuit for the initial system with one digital power domain

The current generator I_0 represents the switching current generated in the digital circuitry of the initial system. For the initial theoretical consideration, the switching current will be approximated as a single harmonic peak in frequency domain. This, of course, is a very coarse oversimplification, which doesn't correspond to a realistic digital circuitry. The goal of this simplification is to be able to analyze the effect this methodology has on a single frequency peak. Later, in Section 5.4, the methodology will be numerically evaluated by using a more realistic model – a periodic triangular pulse, as described in Chapter 3. A periodic triangular pulse in time domain corresponds to multiple harmonic peaks in frequency domain. By knowing the impact of the methodology to each of these harmonic peaks, it will be possible to deal with the spectrum of a realistic switching current, as will be shown in Section 5.4.

According to the model in Fig. 5.1, the ground bounce voltage can be calculated as:

$$V_{gb0} = Z_p I_0 = \frac{Z'_p}{N_d} I_0 \quad (5.1)$$

while the substrate voltage at the point of victim can be expressed as:

$$V_{sub0} = \frac{Z_{pa}}{Z_{pa} + Z_{seff}} V_{gb0} = \frac{Z_{pa}}{Z_{pa} + Z_{seff}} \frac{Z'_p}{N_d} I_0 \quad (5.2)$$

Note that all the voltage and current representations are in frequency domain, while the impedances are complex numbers.

If this initial system is now partitioned in power domains D1 and D2 which both retain the same supply level as the initial system, but have separated PDNs, the equivalent circuit will change. The supply and ground pads of the initial system will now be distributed among the two power domains, in such a way that N_{d1} supply pads and N_{d1} ground pads get assigned to the power domain D1, while N_{d2} supply pads and N_{d2} ground pads get assigned to the power domain D2. Since the equal ground bounce approximation was valid for the initial system, it will be valid for domains D1 and D2 as well, since they're both smaller than the initial system, and thus have smaller on-chip parasitics. Package parasitics for domains D1 and D2 can thus be expressed as $Z_{p1} = Z'_p/N_{d1}$ and $Z_{p2} = Z'_p/N_{d2}$, respectively.

In small systems, for which the equal ground bounce is applicable, the number of pads is usually limited and assigning an additional pad for noise reduction purposes is often not possible. Thus, it will be assumed that the total number of supply and ground pads remains the same:

$$N_{d1} + N_{d2} = N_d \quad (5.3)$$

As a consequence, the package impedance of the initial system Z_p can be expressed as the package impedances of the two domains in the resulting system connected in parallel, i.e:

$$Z_{p1} \parallel Z_{p2} = \frac{\frac{Z'_p}{N_{d1}} \frac{Z'_p}{N_{d2}}}{\frac{Z'_p}{N_{d1}} + \frac{Z'_p}{N_{d2}}} = \frac{Z'_p}{N_{d1} + N_{d2}} = \frac{Z'_p}{N_d} = Z_p \quad (5.4)$$

Additionally, if the overheads due to the interface between the two power domains can be neglected, it can be assumed that both the total area and the total switching current remain the same. In other words, the initial system area A is split into the area A_1 of the domain D1 and the area A_2 of the domain D2, so that $A_1 + A_2 = A$. Similarly:

$$I_0 = I_1 + I_2 \quad (5.5)$$

where I_1 and I_2 are the switching currents of the domains D1 and D2, respectively.

The equivalent circuit of the resulting system is presented in Fig. 5.2. Since the domains D1 and D2 have separated PDNs, each of them will have a different ground bounce. Thus, there are two aggressors in the system now, and consequently the substrate is modeled as a three impedance network.

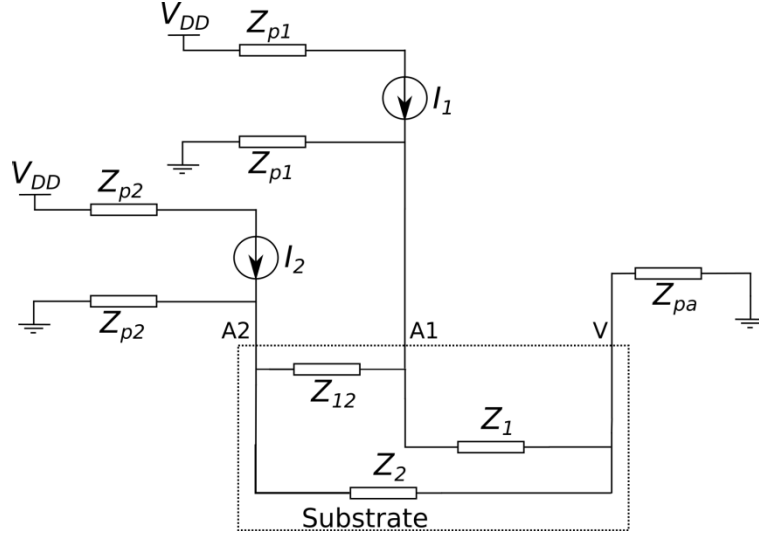


Figure 5.2 – Equivalent circuit after power domain separation of the aggressor

By comparing the two equivalent circuits it can be noticed that the circuit in Fig. 5.1 can be obtained from the circuit in Fig. 5.2 by shorting the two aggressor nodes A1 and A2 in Fig. 5.2 (and by shorting the supply nodes S1 and S2 as well). Indeed, this would mean connecting the current sources I_1 and I_2 in parallel, resulting in I_0 , as well as connecting Z_{p1} and Z_{p2} in parallel both in supply and in ground branch, resulting in Z_p in each of the branches. In other words, the nodes A1 and A2 (as well as the nodes S1 and S2) were shorted in the initial system, due to having the same supply and ground bounce.

Consequently, by shorting the nodes A1 and A2 in Fig. 5.2, the substrate network from Fig. 5.1 is obtained. This means that the equivalent substrate impedance Z_{seff} from Fig. 5.1 is actually equal to a parallel connection of impedances Z_1 and Z_2 from Fig. 5.2, i.e:

$$Z_{seff} = Z_1 \parallel Z_2 \quad (5.6)$$

By inserting (5.6) into (5.2), the substrate noise in the initial system can be expressed as:

$$V_{sub0} = \frac{Z_{pa}Z_2 + Z_{pa}Z_1}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z_p'}{N_d} I_0 \quad (5.7)$$

On the other hand, substrate voltage for the equivalent circuit from Fig. 5.2 is:

$$V_{subp} = \frac{Z_{pa}Z_2}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z_p'}{N_{d1}} I_1 + \frac{Z_{pa}Z_1}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z_p'}{N_{d2}} I_2 \quad (5.8)$$

The reduction of substrate noise, which has been achieved by applying the initial digital power domain separation into the domains D1 and D2, can be quantified as the absolute value of the ratio of substrate noise values from equations (5.7) and (5.8):

$$\frac{V_{subp}}{V_{sub0}} = \frac{1}{1 + \frac{Z_1}{Z_2} \frac{N_d I_1}{N_{d1} I_0}} + \frac{1}{1 + \frac{Z_2}{Z_1} \frac{N_d I_2}{N_{d2} I_0}} \quad (5.9)$$

The sum in (5.9) has two members. The first one originates from the domain D1, while the second one originates from the domain D2. If the impedances in substrate network can be considered as purely resistive (which is a valid approximation up to several GHz [18]), i.e. $Z_1 \approx R_1$ and $Z_2 \approx R_2$, the equation (5.9) can be rewritten as:

$$\frac{V_{subp}}{V_{sub0}} = \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d I_1}{N_{d1} I_0}} + \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d I_2}{N_{d2} I_0}} \quad (5.10)$$

Since switching current is assumed to be sinusoidal, i.e. its spectrum to consist of a single harmonic peak, the same will be valid for substrate noise as well. Spectral peak attenuation of substrate noise can thus be expressed as:

$$SPA = 20 \log \left(\frac{1}{\frac{V_{subp}}{V_{sub0}}} \right) = 20 \log \left(\frac{1}{\frac{1}{1 + \frac{R_1}{R_2} \frac{N_d I_1}{N_{d1} I_0}} + \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d I_2}{N_{d2} I_0}}} \right) \quad (5.11)$$

For a realistic switching current profile, and a switching current spectrum consisting of multiple harmonic peaks, the expression (5.11) can be used to calculate the spectral peak attenuation of each of the harmonic peaks of the substrate noise. In that case, for the purpose of calculating the spectral peak attenuation of the substrate noise at the frequency of the n^{th} harmonic, the switching currents in frequency domain I_0 , I_1 and I_2 have to be replaced by the n^{th} harmonic of the switching current I_{0n} , I_{1n} and I_{2n} , respectively.

As it can be seen from the equation (5.10), the substrate noise ratio (and thus also the substrate noise reduction) depends on the ratio of the equivalent substrate resistances R_1 and R_2 . The values of these resistances are floorplan-dependent – they are determined by the mutual position of the aggressors and by their areas. Since the total area of the aggressors is constant, instead of saying that the substrate resistances depend on the areas of the aggressors, one could also say that they depend on the area ratio of the aggressors, or on the ratio of the area of one of the aggressors to the total area.

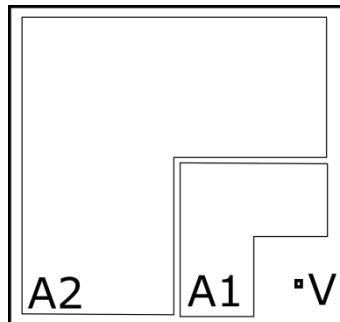


Figure 5.3 – Power domain separation of digital aggressor. D1, D2 – (digital) aggressor power domains, V – analog victim (adapted from [P6])

Now, let's assume a floorplan as shown in Fig. 5.3, where a closer domain D1 is „shielding“ the victim from the further away domain D2. For such a floorplan and an example 0.64 mm ×

0.64 mm chip, the values of R_1 and R_2 are extracted for various values of A_1/A ratio. The extraction has been performed by applying the approach described in Chapter 3, and by using the data for IHP 130 nm technology [25]. The extraction results are shown in Fig. 5.4.

As it can be seen from the graphs in Fig. 5.4, R_1 is monotonically decreasing with increasing the A_1/A ratio, while R_2 is monotonically increasing. Consequently, the ratio R_1/R_2 is monotonically decreasing with increasing the A_1/A ratio. In other words, the larger the domain D1, the lower the R_1/R_2 ratio. Thus, with increasing the size of the D1 domain, the factor $1/(1 + R_1/R_2)$ of the first member of the sum in (5.10) will monotonically increase, while the factor $1/(1 + R_2/R_1)$ of the second member of the sum in (5.10) will monotonically decrease.

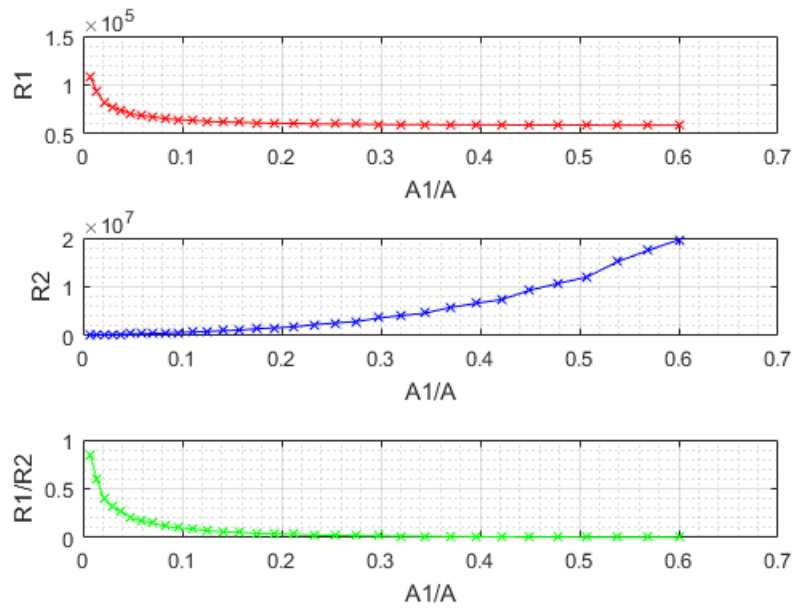


Figure 5.4 – Substrate resistances R_1 and R_2 , and their ratio R_1/R_2 (adapted from [P6])

Additionally, the switching current I_1 , i.e. its ratio to the total switching current, I_1/I_0 , will monotonically increase in its absolute value with increasing the domain D1. On the other hand, increasing the domain D1 means decreasing the domain D2, so the switching current I_2 , i.e. its ratio to the total switching current, I_2/I_0 , will monotonically decrease in its absolute value with increasing the domain D1.

Thus it can be concluded that the first member of the sum in (5.10) monotonically increases in its absolute value with increasing the area ratio A_1/A , i.e. with the amount of digital circuitry assigned to the domain D1, while the second member of the sum monotonically decreases. As a consequence, the absolute value of the sum in (5.10) has a minimum, which means that there is a size of the domain D1 for which the noise is optimally reduced.

Since the power domain separation impacts the ground bounce, it is important to make sure that ground bounce in the newly created power domains doesn't increase too much compared to the initial system. The same ground bounce constraints which were satisfied for the initial system should also be met for the new power domains D1 and D2. In order to achieve this, the power and ground pads have to be appropriately distributed.

As already explained, the ground bounce of the initial system can be represented as:

$$V_{gb0} = \frac{Z'_p}{N_d} I_0 \quad (5.12)$$

If the maximum allowed ground bounce amplitude is V_{gbm} , the following condition has to be satisfied:

$$|V_{gb0}| = \left| \frac{Z'_p}{N_d} I_0 \right| < V_{gbm} \quad (5.13)$$

According to this, a constraint for the number N_d of supply/ground pad pairs can be determined:

$$N_d - 1 < \frac{|Z'_p I_0|}{V_{gbm}} < N_d \quad (5.14)$$

The right portion of the inequality (5.14) originates directly from (5.13), and it's must be satisfied in order to meet the ground bounce constraint. On the other hand, the left portion of the inequality (5.14) is needed only to minimize the number of pads – it ensures that N_d is the minimum number of supply/ground pad pairs meeting the ground bounce constraint, i.e. that the ground bounce constraint can't be met for a lower number of supply/ground pads.

After separation, the domains D1 and D2 will have ground bounce values of $V_{gb1} = (Z'_p/N_{d1})I_1$ and $V_{gb2} = (Z'_p/N_{d2})I_2$. The ground bounce constraint, however, still remains the same:

$$|V_{gb1}| = \left| \frac{Z'_p}{N_{d1}} I_1 \right| < V_{gbm} \quad (5.15)$$

$$|V_{gb2}| = \left| \frac{Z'_p}{N_{d2}} I_2 \right| < V_{gbm} \quad (5.16)$$

Consequently, the values for N_{d1} and N_{d2} have to satisfy the conditions analogous to (5.14):

$$N_{d1} - 1 < \frac{|Z'_p I_1|}{V_{gbm}} < N_{d1} \quad (5.17)$$

$$N_{d2} - 1 < \frac{|Z'_p I_2|}{V_{gbm}} < N_{d2} \quad (5.18)$$

For simpler notation let's define the parameter K as:

$$K = \frac{|Z'_p I_0|}{V_{gbm}} \quad (5.19)$$

Also, let's define the parameter α as the ratio of the switching current of the domain D1 and the total switching current of the initial system:

$$\alpha = \frac{I_1}{I_0} \quad (5.20)$$

Note that, due to (5.5), it can be written that $I_2 = (1 - \alpha)I_0$.

Now the inequalities (5.14), (5.17) and (5.18) can be rewritten as:

$$N_d - 1 < K < N_d \quad (5.21)$$

$$N_{d1} - 1 < |\alpha|K < N_{d1} \quad (5.22)$$

$$N_{d2} - 1 < |1 - \alpha|K < N_{d2} \quad (5.23)$$

By using (5.3), the inequality (5.23) can further be rewritten as:

$$N_d - N_{d1} - 1 < |1 - \alpha|K < N_d - N_{d1} \quad (5.24)$$

From (5.21) and (5.22), the following inequality can be deduced:

$$|\alpha|(N_d - 1) < N_{d1} < |\alpha|N_d + 1 \quad (5.25)$$

on the other hand, from (5.21) and (5.24):

$$(1 - |1 - \alpha|)N_d - 1 < N_{d1} < (1 - |1 - \alpha|)N_d + |1 - \alpha| \quad (5.26)$$

From the conditions (5.25) and (5.26), for each partitioning of the initial system into domains D1 and D2, i.e. for each α , a corresponding N_{d1} can be chosen.

Note that, while (5.25) and (5.26) provide the best N_{d1} fit for a given α , this still doesn't mean that the ground bounce conditions (5.15) and (5.16) are satisfied for every possible α , since these conditions depend on K and thus indirectly also on the value of V_{gb0} before partitioning. This will be further clarified later in this section.

In Fig. 5.5, the absolute value of the substrate noise ratio from equation (5.10) is shown for an example $0.64 \text{ mm} \times 0.64 \text{ mm}$ chip in IHP 130nm technology [25]. The absolute values of the first and the second member of the sum from (5.10), i.e. the members originating from D1 and D2, respectively, are shown as well. The total number of pad pairs reserved for supply and ground of the digital circuitry for this example is $N_d = 4$. The values for N_{d1} and N_{d2} are set depending on the ratio $\alpha = I_1/I_0$, according to the inequalities (5.25) and (5.26).

As a conservative assumption, a uniform distribution of the switching current per aggressor surface has been assumed, which means that the switching current ratio of the domain D1 to the initial system corresponds to their area ratio: $\alpha = I_1/I_0 = A_1/A = a$. A non-uniform distribution of switching current would provide more freedom when assigning the digital blocks to one of the two domains based on their switching current, and thus also more possibilities for a further minimization of the noise, as will be shown later.

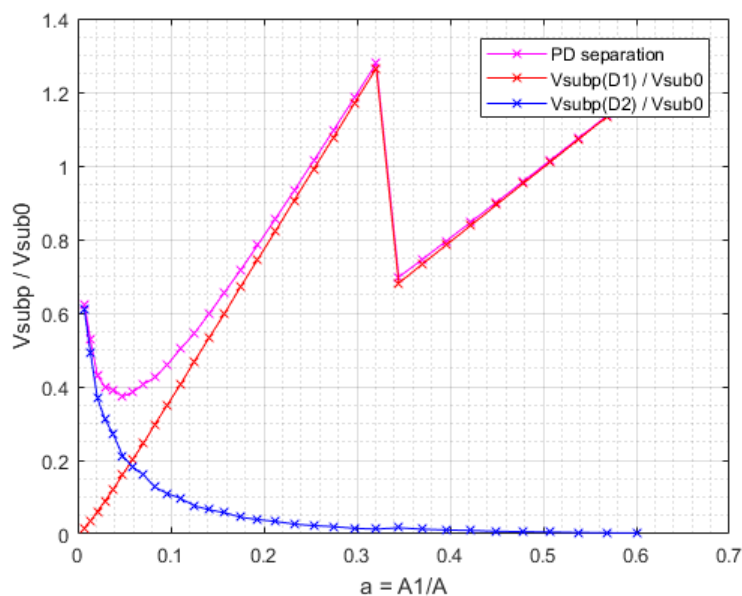


Figure 5.5 – Substrate noise ratio from equation (5.10) (adapted from [P6])

In Fig. 5.5 it can be seen that the domain D2 dominantly impacts the substrate noise only for very low values of $a = A_1/A$. As a increases, the impact of the domain D2 decreases rather sharply in an exponential-like manner. This happens as a consequence of both decreasing the size of the domain D2 (and thus also decreasing the amount of noise injected by the aggressor D2) and increasing the resistance R_2 between the domain D2 and the victim, as shown in Fig. 5.4.

On the other hand, the impact of the domain D1 increases almost linearly with the increase of a . This is mainly due to increasing the size of the domain, and thus the amount of noise injected into the substrate. The resistance R_1 decreases with increasing a , i.e. with increasing the size of the D1 domain, but this decrease reaches saturation already for lower values of a , as it can be seen in Fig. 5.4. Thus, the impact of the domain D1 is mainly determined by the amount of noise injected by the domain.

It can also be noticed that for higher values of a the ratio V_{subp}/V_{sub} becomes larger than 1, i.e. the impact of the domain D1 becomes even larger than the substrate noise in the initial system. In other words, for values of a which are too high, no substrate noise reduction is achieved. This continues until the value for a (and consequently, α) surpasses the limits of the constraints (5.25) and (5.26). As a consequence, for such values of a a new value for N_{d1} has to be applied, which causes a discontinuity which can be seen in the graph. In the part of the graph left of the discontinuity, $N_{d1} = 1$ is applied, while in the part of the graph right of the discontinuity $N_{d1} = 2$ is applied. For the values of a immediately right of the discontinuity, substrate noise reduction is achieved again, however to a lesser degree compared to the reduction achieved for small values of a , where $N_{d1} = 1$.

In Fig. 5.6, for the same example system the dependence of the spectral power attenuation on the ratio $a = A_1/A = I_1/I_0 = \alpha$ is shown. For this system, maximum attenuation is reached for $a = 0.048$, and its value is 8.56 dB.

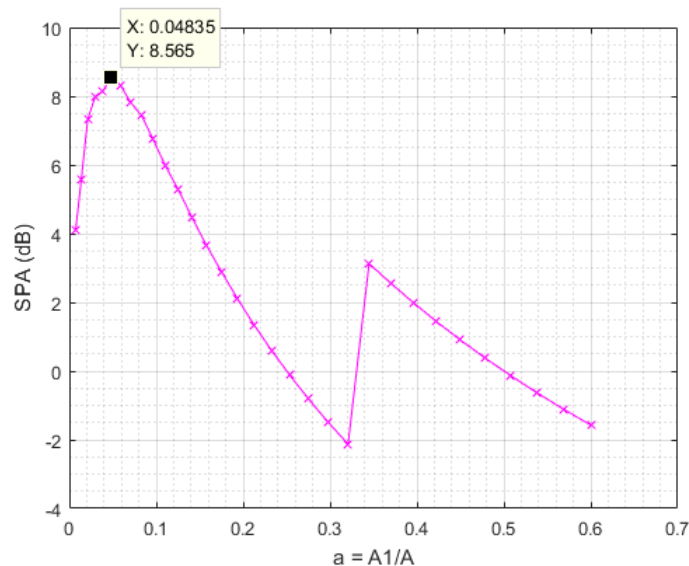


Figure 5.6 – Spectral peak attenuation (dB) achieved by separating power domains (adapted from [P6])

This optimum attenuation is achieved at the price of increasing ground bounce in D2, as it can be seen in Fig. 5.7. In this figure, the ground bounce ratios $K_{gb10} = |V_{gb1}/V_{gb0}|$ and $K_{gb20} = |V_{gb2}/V_{gb0}|$ are shown.

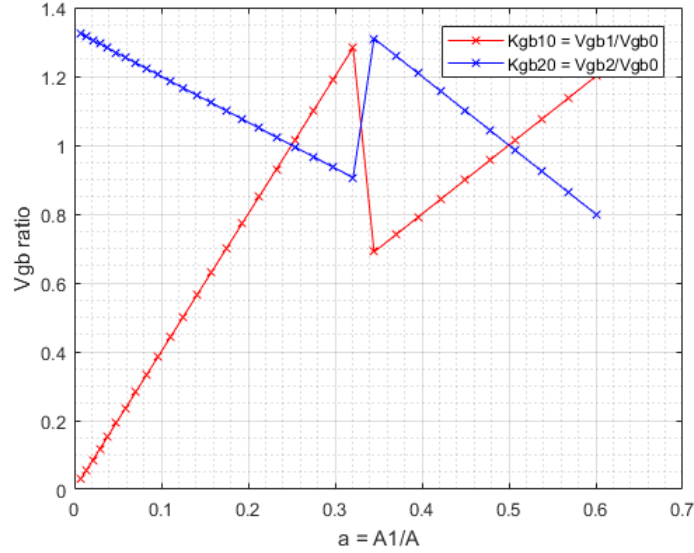


Figure 5.7 – The effect of power domain separation on ground bounce (adapted from [P6])

Now the statement that fulfilling the ground bounce constraint depends also on the initial V_{gb0} can be explained more in detail. The ground bounce constraint (5.13) is fulfilled for the initial system, and the ground bounce constraints (5.15) and (5.16) should be fulfilled for the newly created power domains D1 and D2. These constraints can also be formulated in the following way:

$$\frac{|V_{gb0}|}{V_{gbm}} = K_0 < 1 \quad (5.27)$$

$$\frac{|V_{gb1}|}{V_{gbm}} = K_1 < 1 \quad (5.28)$$

$$\frac{|V_{gb2}|}{V_{gbm}} = K_2 < 1 \quad (5.29)$$

By using (5.27) and (5.13), the parameter K defined in (5.19) can be represented as follows:

$$K = \frac{|Z'_p I_0|}{V_{gbm}} = K_0 \frac{|Z'_p I_0|}{V_{gb0}} = K_0 N_d \quad (5.30)$$

By combining (5.30) and the constraint (5.21), a complete constraint for the parameter K_0 can be derived:

$$\frac{N_d - 1}{N_d} < K_0 < 1 \quad (5.31)$$

Further, the ground bounce ratios from Fig. 5.7 can now be represented as follows:

$$K_{gb10} = \frac{|V_{gb1}|}{|V_{gb0}|} = \frac{K_1 V_{gbm}}{K_0 V_{gbm}} = \frac{K_1}{K_0} \quad (5.32)$$

$$K_{gb20} = \frac{|V_{gb2}|}{|V_{gb0}|} = \frac{K_2 V_{gbm}}{K_0 V_{gbm}} = \frac{K_2}{K_0} \quad (5.33)$$

By using (5.32) and (5.33), constraints (5.28) and (5.29) can be rewritten in the following way:

$$K_1 = K_{gb10} K_0 < 1 \quad (5.34)$$

$$K_2 = K_{gb20} K_0 < 1 \quad (5.35)$$

These two inequalities must be fulfilled, in order to meet the ground bounce constraint in both of the newly formed power domains D1 and D2. Whether or not these two inequalities are satisfied, depends on the values of K_{gb10} and K_{gb20} , which are represented in Fig. 5.7, but also on the value of K_0 , which is a property of the initial system.

In an extreme case where $K_0 = 1$, i.e. $|V_{gb0}| = V_{gbm}$, the ground bounce constraint in the initial system is just barely met, without any ground bounce margin left. For such system, the conditions (5.34) and (5.35) would turn into $K_{gb10} < 1$ and $K_{gb20} < 1$, respectively. As it can be seen from Fig. 5.7, both these condition would barely be met only at a single point. This case, however, would never occur in practice – for any realistic system, there’s always a margin between the actual ground bounce and a ground bounce constraint.

On the contrary, in an opposite extreme case where $K_0 = (N_d - 1)/N_d$, i.e. where $|V_{gb0}| = (N_d - 1)V_{gbm}/N_d$, the ground bounce constraint in the initial system is reached with a very high margin. It could have even been reached (although just barely, without any margin) if there was one supply/ground pad pair less. The conditions (5.34) and (5.35) would turn into $K_{gb10} < N_d/(N_d - 1)$ and $K_{gb20} < N_d/(N_d - 1)$, respectively. For the example which is shown in Fig. 5.7, with $N_d = 4$, this would turn into $K_{gb10} < 1.33$ and $K_{gb20} < 1.33$. As it can be seen from Fig. 5.7, these conditions are satisfied for every α . While this case can occur in practice, it is not very probable – the margin for most of the realistic systems would usually not be that high.

The margin for realistic systems, and thus the value of K_0 , is between these two extreme cases. Thus, there may exist a range of α for which the ground bounce constraints aren’t reached. If the value of K_0 is known, the values of K_{gb10} and K_{gb20} for which (5.34) and (5.35) are satisfied can be determined. From there, using the graph in Fig. 5.7, the range of α for which the ground bounce constraints are satisfied can be determined.

In this section, substrate noise reduction which can be achieved by power domain separation has been analyzed. In the following section, this methodology will be extended by applying galsification to the newly formed partitions.

5.3. Additional improvement by applying GALS design approach

In the previous section, substrate noise reduction which can be achieved by partitioning a digital aggressor in two separate power domains has theoretically been analyzed. The partitioned system, however, remains synchronous, i.e. both of the newly formed power domains is still clocked with the same clock as the original system. As a consequence, switching current spectra of both power domains will have peaks at the same frequencies. Thus, the substrate noise originating from each of the power domains will have peaks at the same frequency, and these peaks will add to each other in the frequency domain.

This can be clearly seen in the equation (5.10), where the substrate noise ratio between the new, partitioned system, and the original synchronous system is shown. This ratio represents a sum comprising of two members, the first of them originating from the domain D1, and the second originating from the domain D2.

However, if the two newly formed power domains have separate clocks at different frequencies, substrate noise in frequency domain will consist of two separate harmonic peaks, each

originating from a different power (and now also clock) domain. From the model shown in Fig. 5.2, the substrate noise originating from the domains D1 and D2 can be calculated as, respectively:

$$V_{subp1} = \frac{Z_{pa}Z_2}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z_p'}{N_{d1}} I_1 \quad (5.36)$$

$$V_{subp2} = \frac{Z_{pa}Z_1}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z_p'}{N_{d2}} I_2 \quad (5.37)$$

By clocking each of the power domains with a separate clock, the synchronous system is turned into a GALS system, with two LSMs corresponding to the power domains D1 and D2. The advantages of the plesiochronous GALS scheme, i.e. the scheme where the LSM clocks have different but very close frequencies, have been presented in detail in Chapter 4. This scheme maintains the processing capability of the digital circuitry compared to the original synchronous clocking scheme. Additionally, since the harmonic peak frequencies are very close to each other, substrate and package impedances are calculated at approximately the same frequency as for the initial synchronous system, which is advantageous for predicting the noise reduction.

Substrate noise ratio from (5.9) can now be redefined as the ratio between the highest of the substrate noise harmonic peaks of the new GALS system, and the substrate noise harmonic peak of the initial synchronous system:

$$\left| \frac{V_{subp}}{V_{sub0}} \right| = \max \left\{ \left| \frac{V_{subp1}}{V_{sub0}} \right|, \left| \frac{V_{subp2}}{V_{sub0}} \right| \right\} \quad (5.38)$$

which, by applying (5.7), (5.36) and (5.37), results in:

$$\left| \frac{V_{subp}}{V_{sub0}} \right| = \max \left\{ \left| \frac{1}{1 + \frac{Z_1}{Z_2} \frac{N_d}{N_{d1}} \frac{I_1}{I_0}} \right|, \left| \frac{1}{1 + \frac{Z_2}{Z_1} \frac{N_d}{N_{d2}} \frac{I_2}{I_0}} \right| \right\} \quad (5.39)$$

For frequencies up to a couple of GHz, where the resistive substrate approximation is applicable, (5.39) can be written as:

$$\left| \frac{V_{subp}}{V_{sub0}} \right| = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_{d1}} \frac{I_1}{I_0}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_{d2}}{N_{d2}} \frac{I_2}{I_0}} \right| \right\} \quad (5.40)$$

From (5.40) the spectral peak attenuation of the substrate noise can be calculated:

$$SPA = 20 \log \left(\min \left\{ \left(1 + \frac{R_1}{R_2} \right) \frac{N_{d1}}{N_{d1}} \frac{I_1}{I_0}, \left(1 + \frac{R_2}{R_1} \right) \frac{N_{d2}}{N_{d2}} \frac{I_2}{I_0} \right\} \right) \quad (5.41)$$

It can be noticed that the SPA for a GALS system, as expressed in (5.41), is larger than the SPA for a system where only power supply domain separation has been performed, as expressed in (5.41). According to the theory presented in Chapter 4, maximum additional attenuation by galsification is obtained in case when the two harmonic peaks, resulting from the two LSMs (i.e. the two power domains D1 and D2), are balanced. In other words, GALS introduces the largest additional attenuation when the two power domains have equal impact to the substrate noise at victim's position. This maximum additional attenuation, according to (4.21), is equal to $20 \log(2) = 6$ dB.

If the frequencies of the clock domains don't differ too much from the original synchronous frequency, the ground bounce constraints depend only on the power delivery networks and power

domains, not on the clock domains. Thus, the ground bounce constraints remain the same, regardless of whether the two domains are clocked synchronously or plesiochronously.

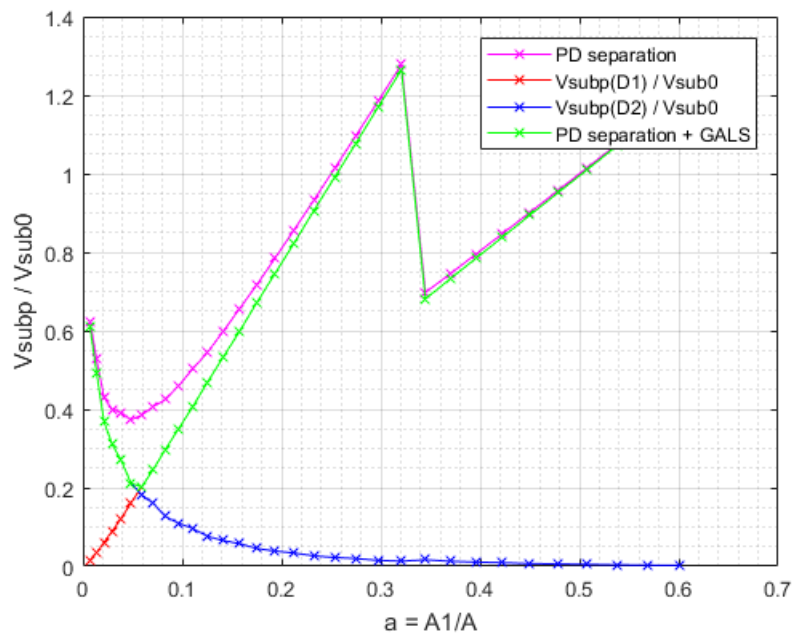


Figure 5.8 – Substrate noise ratio achieved only by separating power domains, and by separating power domains combined with galsification (adapted from [P6])

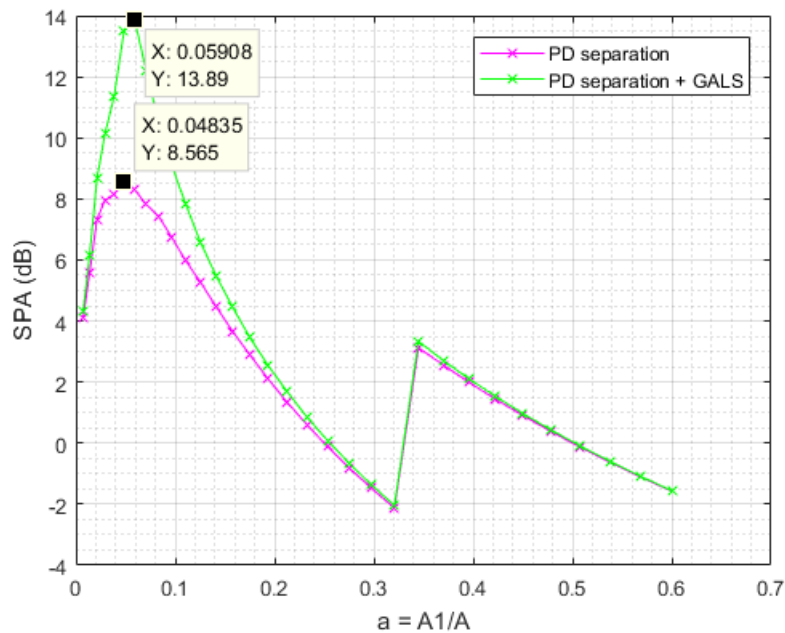


Figure 5.9 – Spectral peak attenuation (dB) achieved only by separating power domains, and by separating power domains combined with galsification (adapted from [P6])

In Fig. 5.8 and Fig. 5.9, the absolute value of the substrate noise ratio from Fig. 5.5 and the spectral peak attenuation from Fig. 5.6 are shown again for the same example system, but this time with an additional graph showing the results obtained by applying the additional galsification. As it can be seen, applying a plesiochronous GALS scheme instead of a standard synchronous clocking scheme contributes to further reducing the substrate noise. The benefit of galsification is mostly

pronounced for the values of a where the impacts of the two power domains have similar values. The optimum attenuation achieved by power domain separation and galsification is 5.325 dB larger than the optimum attenuation achieved only by power domain separation. Note that the optimum is achieved for almost the same value of a for both cases.

5.4. The impact of decoupling capacitance

In sections 5.2 and 5.3, the substrate noise reduction by power domain separation and galsification has been theoretically analyzed. In order to simplify the analysis, the existence of the on-chip decoupling capacitance has been neglected. In this section, the decoupling capacitance will be included into the model, and its impact will be taken into account.

As mentioned in Chapter 3, the decoupling capacitance can be introduced in the model by adding a capacitor in parallel to the current source modeling the switching current. Such a capacitor represents the total decoupling capacitance, i.e. the sum of the intentional on-chip decoupling capacitance and the inherent capacitance of the switching circuitry. According to this, the equivalent circuit of the initial system from Fig. 5.1 transforms into the equivalent circuit presented in Fig. 5.10, with C_d being the total decoupling capacitance of the original synchronous digital circuitry.

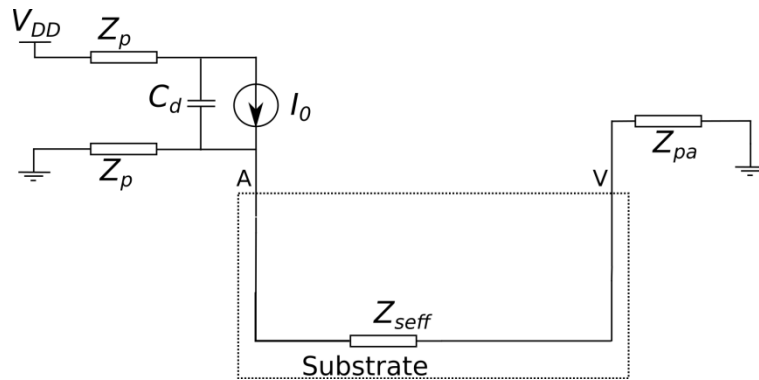


Figure 5.10 – Equivalent circuit for the original system with synchronous digital circuitry, with decoupling capacitance included into the model

By using this equivalent circuit, a new expression for substrate noise at victim's position can be calculated:

$$V_{sub} = \frac{Z_{pa}Z_2 + Z_{pa}Z_1}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z'_p}{N_d} \frac{1}{1 + 2j\omega C_d \frac{Z'_p}{N_d}} I_0 \quad (5.42)$$

where $\omega = 2\pi f$, and f is the frequency of the injected switching current harmonic peak, i.e. the frequency at which the substrate noise is calculated.

Similarly, the model from Fig. 5.2 is transformed into the model shown in Fig. 5.11, with C_{d1} and C_{d2} being the total decoupling capacitances of the power domains D1 and D2, respectively. Note that, in order to retain the same total area, the total amount of decoupling capacitance remains unchanged:

$$C_{d1} + C_{d2} = C_d \quad (5.43)$$

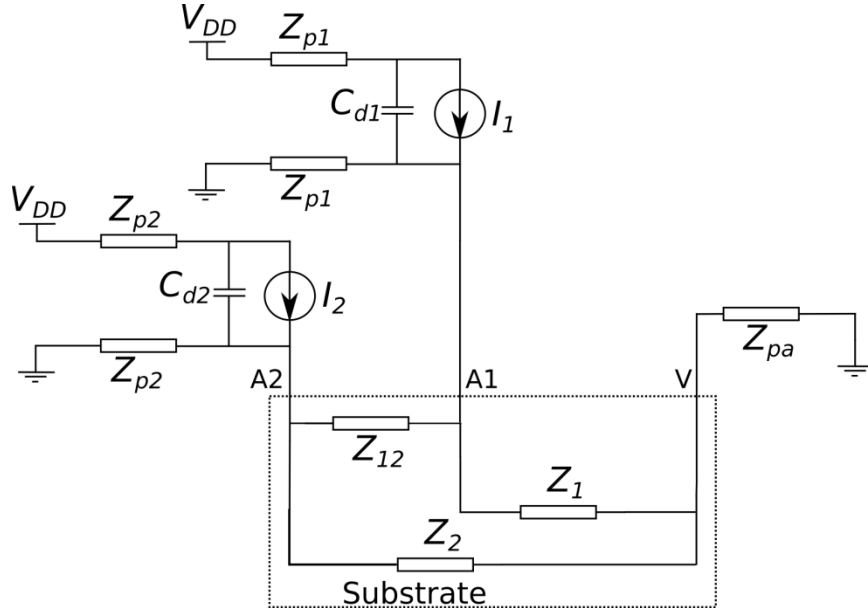


Figure 5.11 – Equivalent circuit for the system with power domain separation, with decoupling capacitance included into the model (adapted from [P6])

From this equivalent circuit, a new expression for the substrate voltage for the system with power domain separation can be calculated. In case that only power domain separation is performed:

$$\begin{aligned}
 V_{subp} = & \frac{Z_{pa}Z_2}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z'_p}{N_{d1}} \frac{1}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} I_1 \\
 & + \frac{Z_{pa}Z_1}{Z_1Z_2 + Z_{pa}Z_1 + Z_{pa}Z_2} \frac{Z'_p}{N_{d2}} \frac{1}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} I_2
 \end{aligned} \quad (5.44)$$

From here, the ratio between the reduced substrate noise and the initial substrate noise can be calculated:

$$\left| \frac{V_{subp}}{V_{sub}} \right| = \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_1}{I_0}} \frac{N_d}{N_{d1}} \frac{I_1}{I_0} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} + \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d}{N_{d2}} \frac{I_2}{I_0}} \frac{N_d}{N_{d2}} \frac{I_2}{I_0} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} \right| \quad (5.45)$$

as well as the spectral peak attenuation:

$$SPA = 20 \log \left(\frac{1}{\left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_1}{I_0}} \frac{N_d}{N_{d1}} \frac{I_1}{I_0} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} + \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d}{N_{d2}} \frac{I_2}{I_0}} \frac{N_d}{N_{d2}} \frac{I_2}{I_0} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} \right|} \right) \quad (5.46)$$

In case that both power domain separation and galsification are performed, equations (5.45) and (5.46) respectively transform into:

$$\left| \frac{V_{subp}}{V_{sub}} \right| = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_1}{I_0}} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d}{N_{d2}} \frac{I_2}{I_0}} \frac{1 + 2j\omega C_d \frac{Z'_p}{N_d}}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} \right| \right\} \quad (5.47)$$

and:

$$SPA = 20 \log \left(\min \left\{ \left| \left(1 + \frac{R_1}{R_2} \right) \frac{N_{d1}}{N_d} \frac{I_0}{I_1} \frac{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}}{1 + 2j\omega C_d \frac{Z'_p}{N_d}} \right|, \left| \left(1 + \frac{R_2}{R_1} \right) \frac{N_d}{N_{d2}} \frac{I_2}{I_0} \frac{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}}{1 + 2j\omega C_d \frac{Z'_p}{N_d}} \right| \right\} \right) \quad (5.48)$$

Substrate noise ratio and spectral peak attenuation for the model without decoupling capacitance, as it can be seen from equations (5.10), (5.11), (5.40) and (5.41), are dependent only on the ratio of substrate resistances, and on the parameters of the domains D1 and D2, such as the number of supply/ground pad pairs (N_{d1}/N_d) and the amount of switching circuitry (I_1/I_0) assigned to each of them. However, as it can be seen from equations (5.45)-(5.48), when decoupling capacitance is taken into account, substrate noise ratio and spectral peak attenuation depend also on the parasitic impedance of the package (Z'_p) and on the decoupling capacitance itself.

By introducing the decoupling capacitance C_d into the model, the ground bounce constraints also become modified. Ground bounce in the initial system without separation becomes:

$$V_{gb0} = \frac{Z'_p}{N_d} \frac{1}{1 + 2j\omega C_d \frac{Z'_p}{N_d}} I_0 \quad (5.49)$$

while for the D1 and D2 domain, respectively:

$$V_{gb1} = \frac{Z'_p}{N_{d1}} \frac{1}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} I_1 \quad (5.50)$$

$$V_{gb2} = \frac{Z'_p}{N_{d2}} \frac{1}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} I_2 \quad (5.51)$$

From here, the constraints for the maximum ground bounce are:

$$|V_{gb0}| = \left| \frac{Z'_p}{N_d} \frac{1}{1 + 2j\omega C_d \frac{Z'_p}{N_d}} I_0 \right| < V_{gbm} \quad (5.52)$$

$$|V_{gb1}| = \left| \frac{Z'_p}{N_{d1}} \frac{1}{1 + 2j\omega C_{d1} \frac{Z'_p}{N_{d1}}} I_1 \right| < V_{gbm} \quad (5.53)$$

$$|V_{gb2}| = \left| \frac{Z'_p}{N_{d2}} \frac{1}{1 + 2j\omega C_{d2} \frac{Z'_p}{N_{d2}}} I_2 \right| < V_{gbm} \quad (5.54)$$

In order to have the minimum possible number of supply/ground pads, the numbers N_d , N_{d1} and N_{d2} should be the least possible numbers satisfying these equations. Thus:

$$\left| \frac{Z'_p}{(N_d - 1)} \frac{1}{1 + 2j\omega C_d \frac{Z'_p}{(N_d - 1)}} I_0 \right| > V_{gbm} \quad (5.55)$$

$$\left| \frac{Z'_p}{(N_{d1} - 1) 1 + 2j\omega C_{d1} Z'_p / (N_{d1} - 1)} I_1 \right| > V_{gbm} \quad (5.56)$$

$$\left| \frac{Z'_p}{(N_{d2} - 1) 1 + 2j\omega C_{d2} Z'_p / (N_{d2} - 1)} I_2 \right| > V_{gbm} \quad (5.57)$$

From (5.52)-(5.57), knowing that $I_0 = I_1 + I_2$ (5.5), and by applying $\alpha = I_1/I_0$ (5.20) and $K = |Z'_p I_0|/V_{gbm}$ (5.19), the ground bounce constraints for a system with decoupling capacitance can be deduced as follows:

$$|N_d - 1 + 2j\omega C_d Z'_p| < K < |N_d + 2j\omega C_d Z'_p| \quad (5.59)$$

$$|N_{d1} - 1 + 2j\omega C_{d1} Z'_p| < |\alpha|K < |N_{d1} + 2j\omega C_{d1} Z'_p| \quad (5.60)$$

$$|N_{d2} - 1 + 2j\omega C_{d2} Z'_p| < |1 - \alpha|K < |N_{d2} + 2j\omega C_{d2} Z'_p| \quad (5.61)$$

From (5.59)-(5.61), knowing that $N_d = N_{d1} + N_{d2}$ (5.3) and $C_d = C_{d1} + C_{d2}$ (5.43), and by defining:

$$c = \frac{C_{d1}}{C_d} \quad (5.62)$$

$$K_p = 2j\omega C_d \frac{Z'_p}{N_d} \quad (5.63)$$

the following constraints for determining N_{d1} can be deduced:

$$|\alpha| |N_d - 1 + K_p N_d| < |N_{d1} + c K_p N_d| \quad (5.64)$$

$$|N_{d1} - 1 + c K_p N_d| < |\alpha| |N_d + K_p N_d| \quad (5.65)$$

$$|1 - \alpha| |N_d - 1 + K_p N_d| < |N_d - N_{d1} + (1 - c) K_p N_d| \quad (5.66)$$

$$|N_d - N_{d1} - 1 + (1 - c) K_p N_d| < |1 - \alpha| |N_d + K_p N_d| \quad (5.67)$$

Note that the constraints (5.64)-(5.67) correspond to the constraints (5.25) and (5.26) for a system without decoupling capacitance.

Substrate noise ratio and spectral peak attenuation have been evaluated again for the example system from Section 5.2, but this time with the decoupling capacitance included into the model. The initial system has $N_d = 4$ supply/ground pad pairs, while the number of supply/ground pad pairs assigned to D1 and D2, i.e. N_{d1} and N_{d2} , have been calculated from constraints (5.64)-(5.67). Package parasitics for a single pad in this example are $R'_p = 1 \Omega$ and $L'_p = 1 \text{ nH}$. Two values of the total decoupling capacitance have been used: $C_d = 1.5 \text{ nF}$, corresponding to a slightly underdamped system, and $C_d = 1.5 \text{ nF}$, corresponding to a slightly overdamped system. The result has been evaluated for the frequency of 100 MHz.

A uniform distribution of the decoupling capacitance per area has been assumed, which means that $c = C_{d1}/C_d = A_1/A = a$ has been assumed. Once more, this is a conservative assumption, since being able to control the decoupling capacitance distribution per area would introduce an additional level of freedom, which could be used to further improve noise reduction.

Substrate noise ratio is shown in Fig. 5.12 for the underdamped system ($C_d = 1.5 \text{ nF}$) and in Fig. 5.13 for the overdamped system ($C_d = 1.5 \text{ nF}$), while spectral peak attenuation is shown in Fig. 5.14 for the underdamped system and in Fig. 5.15 for the overdamped system. Both the result of power domain partitioning with retaining the synchronous clocking and the result of power domain partitioning with plesiochronous GALS clocking are presented.

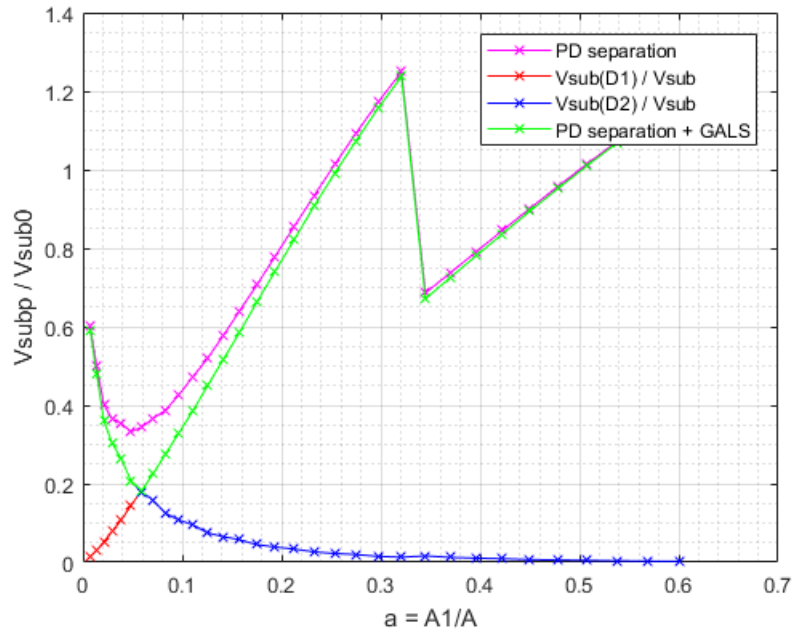


Figure 5.12 – Substrate noise ratio, model with C_d , underdamped system ($C_d = 1.5$ nF) (adapted from [P6])

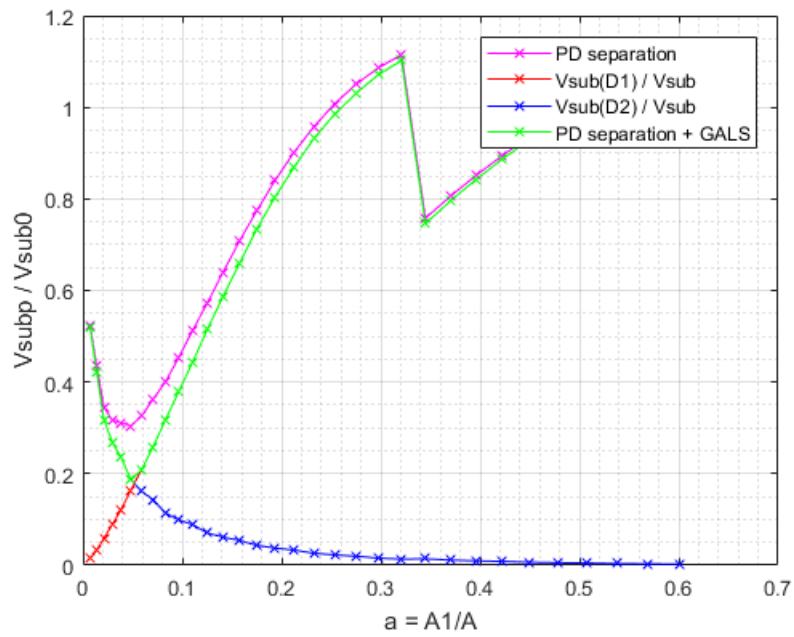


Figure 5.13 – Substrate noise ratio, model with C_d , overdamped system ($C_d = 2.5$ nF)

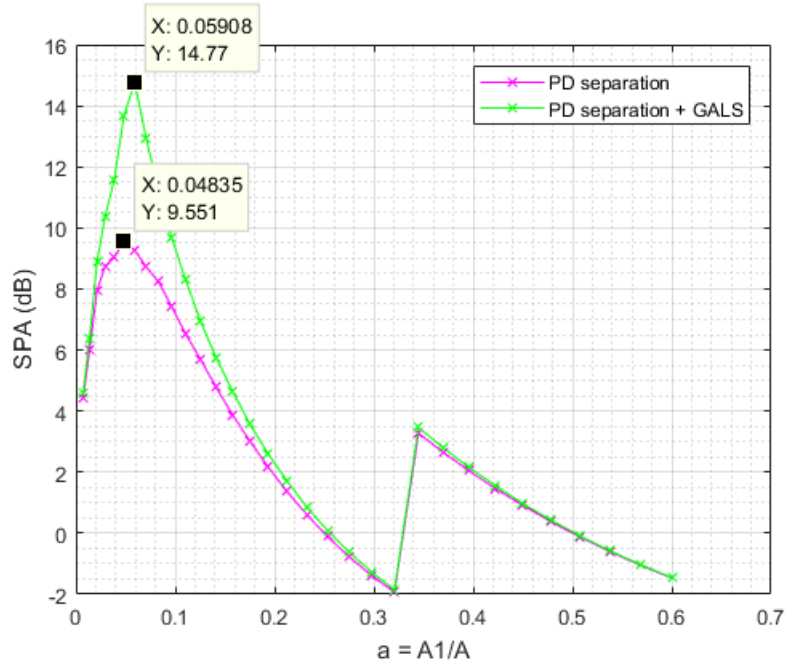


Figure 5.14 – Spectral peak attenuation, model with C_d , underdamped system ($C_d = 1.5$ nF) (adapted from [P6])

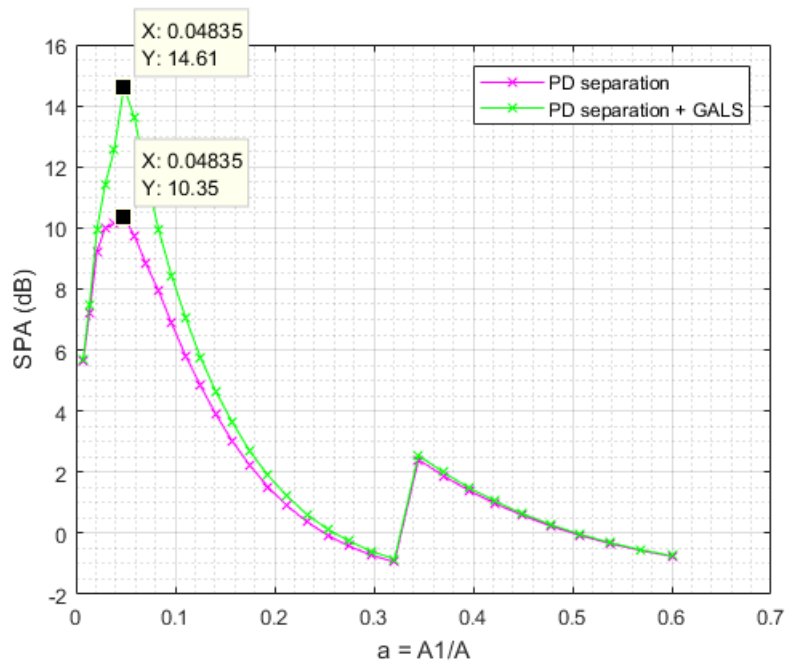


Figure 5.15 – Spectral peak attenuation, model with C_d , overdamped system ($C_d = 2.5$ nF)

Table 5-III – Spectral peak attenuation for the example system

Methodology	Result	Model		
		no C_d	$C_d = 1.5 \text{ nF}$ (underdamped system)	$C_d = 2.5 \text{ nF}$ (overdamped system)
Power domain separation	Maximum SPA:	8.565 dB	9.551 dB	10.35 dB
	reached at $a =$	0.04835	0.04835	0.04835
Power domain separation + GALS	Maximum SPA:	13.89 dB	14.77 dB	14.61 dB
	reached at $a =$	0.05908	0.05908	0.04835

The optimum attenuation results from Fig. 5.9, Fig. 5.14 and Fig. 5.15 are summarized in Table 5-I. As it can be seen in the table, the presence of the decoupling capacitance in the system further improves the spectral peak attenuation reached by power domain separation, with or without an additional galsification. The impact is more pronounced if no galsification is applied; in that case, the impact gets stronger with the increase of the decoupling capacitance. On the other hand, the value of a , at which the optimum attenuation is reached, shifts only slightly towards lower values when C_d is introduced.

The ground bounce penalty is also impacted by introducing the decoupling capacitance. The ground bounce ratios $K_{gb10} = V_{gb1}/V_{gb0}$ and $K_{gb20} = V_{gb2}/V_{gb0}$ are shown in Fig. 5.16 for an underdamped system (with $C_d = 1.5 \text{ nF}$) and in Fig. 5.17 for the overdamped system (with $C_d = 2.5 \text{ nF}$).

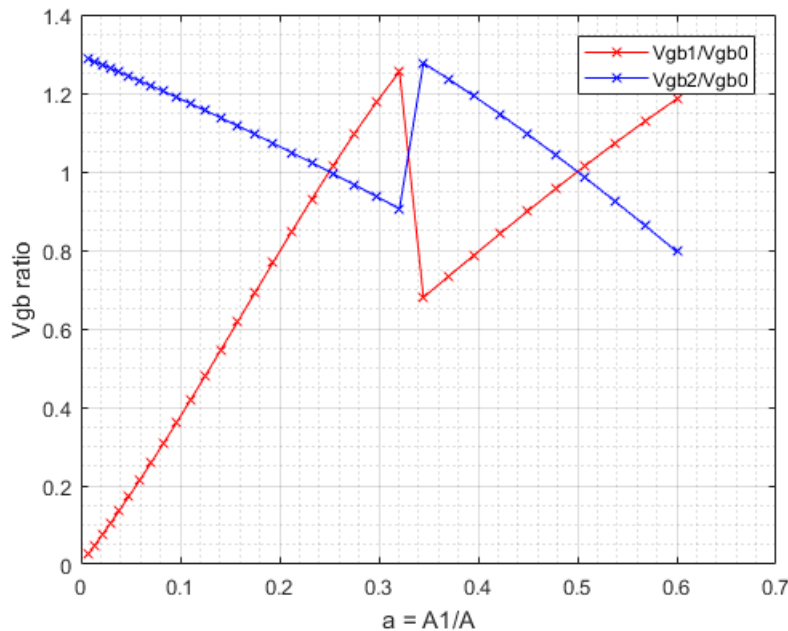


Figure 5.16 – The effect of power domain separation on ground bounce, model with C_d , underdamped system ($C_d = 1.5 \text{ nF}$)

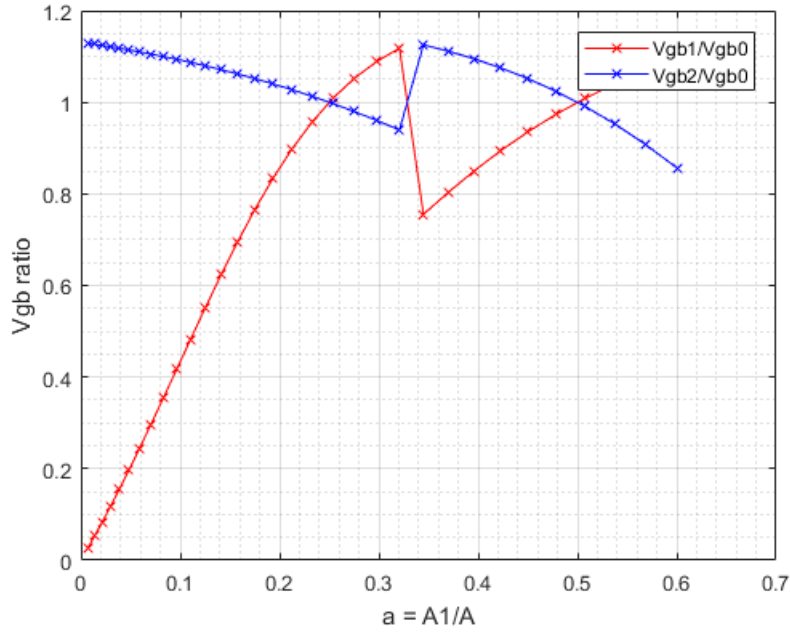


Figure 5.17 – The effect of power domain separation on ground bounce, model with C_d , overdamped system ($C_d = 2.5$ nF)

Note that the conditions (5.27)-(5.29), stating that the parameters K_0 , K_1 and K_2 have to be lower than 1, as well as the equations (5.32) and (5.33), showing the relation between K_{gb10} , K_1 and K_0 , and respectively K_{gb20} , K_2 and K_0 , remain the same also for the system with C_d . Consequently, (5.34) and (5.35), which are repeated here as (5.68) and (5.69) respectively, remain valid as well:

$$K_1 = K_{gb10}K_0 < 1 \quad (5.68)$$

$$K_2 = K_{gb20}K_0 < 1 \quad (5.69)$$

By comparing Fig. 5.16 and Fig. 5.17 with Fig. 5.7 it can be noticed that the ground bounce ratios K_{gb10} and K_{gb20} are lower for the system with decoupling capacitance included. While the difference compared to the system without the decoupling capacitance is only slight for the underdamped system, for the overdamped system the improvement is significant. As a consequence, for the same K_0 , the conditions (5.68) and (5.69) will be satisfied for a wider range of α , which increases the applicability of the methodology.

In the analysis so far, substrate noise reduction by power domain separation and GALS partitioning has been analyzed only theoretically, with a number of model simplifications applied. In the next section, an algorithm applicable on a realistic system will be presented and evaluated.

5.5. An algorithm for substrate noise reduction by GALS partitioning and power domain separation, and a numerical evaluation of the methodology in MATLAB

In Sections 5.2-5.4, the theoretical foundations of substrate noise reduction based on power domain separation and galsification were presented. These foundations have been analyzed by using a number of simplifications. First of all, noise was analyzed only at a single frequency, which corresponds to the case where switching current has a sinusoidal waveform. As already explained in Section 3, a more realistic model of the switching current in digital circuitry would be a periodic triangular pulse, which corresponds to multiple harmonic peaks at the multiples of clock frequency. Besides that, similar to the initial theoretical analysis from the previous chapter, the analysis in sections 5.2-5.4 doesn't take into account the finite granularity of the system, i.e. the fact that neither area nor harmonic magnitude of the initial system can be distributed into two domains with deliberate precision, due to the finite size of the blocks composing the system.

In practice, the optimization problem can be defined in the following way: A synchronous system is given, consisting of a set of digital modules, with known areas and switching current waveforms. These modules or blocks have to be distributed in two partitions, which will be turned into separate power domains and LSMs of a GALS system. The distribution should be done in such a way, that the optimum substrate noise reduction is achieved at the targeted frequency, i.e. that the targeted spectral peak is attenuated as much as possible.

The equation (5.48) expresses the spectral peak attenuation of substrate noise at a single frequency, which can be achieved by power domain separation and galsification. While the equation is derived for a sinusoidal switching current having a single frequency peak, it also remains valid for a single harmonic peak of a more complex periodic waveform, such as periodic triangular waveform. Achieving the maximum spectral peak attenuation at a targeted substrate noise harmonic is equivalent to achieving the minimum substrate noise ratio at the harmonic of interest, i.e. to minimizing the right side of the equation (5.49) at the frequency of the targeted harmonic. Thus, the right side of the equation (5.49), applied at the n^{th} harmonic of the switching current, can be used as a cost function which should be minimized by the optimizing algorithm:

$$c_1 = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_d} I_{0,n}} \frac{N_d I_{1,n}}{1 + 2jn\omega_1 C_{d1}} \frac{1 + 2jn\omega_0 C_d \frac{Z'_p(\omega_0)}{N_d}}{1 + 2jn\omega_0 C_d \frac{Z'_p(\omega_0)}{N_d}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_{d2}}{N_d} I_{0,n}} \frac{N_d I_{2,n}}{1 + 2jn\omega_2 C_{d2}} \frac{1 + 2jn\omega_0 C_d \frac{Z'_p(\omega_0)}{N_d}}{1 + 2jn\omega_2 C_{d2} \frac{Z'_p(\omega_2)}{N_{d2}}} \right| \right\} \quad (5.70)$$

where $I_{0,n}$, $I_{1,n}$ and $I_{2,n}$ are the n^{th} harmonics of the switching current of the initial system, the domain D1 (closer to the victim) and the domain D2 (further from the victim), respectively. Also, $\omega_0 = 2\pi f_0$, $\omega_1 = 2\pi f_1$, and $\omega_2 = 2\pi f_2$, where f_0 , f_1 and f_2 are the clock frequencies of the initial system, the domain D1 and the domain D2, respectively. Note that, for a plesiochronous system, $f_0 \approx f_1 \approx f_2$, i.e. $\omega_0 \approx \omega_1 \approx \omega_2$.

The cost function c_1 from equation (5.70) contains the package parasitic impedance Z'_p . In other words, it requires a package parasitic model in order to be calculated, which is often not available. Additionally, in a realistic system, the decoupling capacitances C_d , C_{d1} and C_{d2} are

composed not only of the intentional decoupling capacitance, but also from the inherent capacitance of the switching circuitry, which is difficult to estimate. However, if this data is not available, the right side of the equation (5.40), which has been derived for a system without decoupling capacitance, can be used instead:

$$c_2 = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d I_{1,n}}{I_{0,n}}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d I_{2,n}}{I_{0,n}}} \right| \right\} \quad (5.71)$$

As shown in Section 5.4, while the decoupling capacitance does impact the optimum substrate noise ratio and the spectral peak attenuation (making the SPA larger), it doesn't significantly impact the position of the optimum, which justifies using (5.71) instead of (5.70).

Similar to the algorithm presented in Chapter 4, the optimizing algorithm for this methodology should also go through the possible partitionings, i.e. the possible assignments of blocks into the domains D1 and D2, and for each partitioning it should evaluate the cost function. The partitioning with the minimum cost function should be chosen as the optimum one. If the number of blocks to be distributed is not large, an exhaustive search through all the possible combinations can be performed. The number of the possible assignments is $2^N - 2$, where N is the number of blocks. For a large number of blocks this can become an excessively large number – in such cases, a simulated annealing algorithm, similar to the one from Chapter 4, can be used [200].

Regardless of whether an exhaustive search or simulated annealing is applied, the cost function has to be evaluated in every iteration of the optimizing algorithm. Both cost functions (5.70) and (5.71) contain the ratio of substrate resistances R_1/R_2 . Evaluation of this ratio, i.e. estimation of the substrate resistances, would normally require substrate extraction. Performing substrate extraction in every optimization step would be too timing costly. Thus, it is necessary to approximate the ratio of substrate resistances R_1/R_2 by an analytical function.

From Fig. 5.4 it can be seen that R_1/R_2 can be described as a function of the area ratio $a = A_1/A$ which is monotonically decreasing in an exponential-like manner. Thus, as an appropriate fitting function for estimating the resistance ratio R_1/R_2 , an exponential function

$$\frac{R_1}{R_2} = k_1 e^{-k_2 a} \quad (5.72)$$

or a rational function

$$\frac{R_1}{R_2} = k_1 / a^{k_2} \quad (5.73)$$

can be selected, with k_1 and k_2 being the fitting parameters. Both the exponential and the rational approximation require just two extracted values of R_1/R_2 in order for the fitting parameters k_1 and k_2 to be determined. The extraction can be performed prior to starting the optimization process. From the analysis in the sections 5.2-5.4 it can be concluded that the optimum point is expected at low values of a . Thus, in order to achieve the best accuracy of the analytical approximation, the two fitting points should be chosen in the range where $a < 1/N_d$.

In Fig. 5.18, substrate resistance ratio R_1/R_2 from Fig. 5.4 is shown again, together with the two approximate analytical functions – the exponential one and the rational one. Fitting parameters have been calculated from the values of the extracted function at $a = 0.04835$ and $a = 0.1104$. As

it can be seen, the analytical estimation fits the extracted function well, except for the very small values of a . Note, however, that for a realistic system, extremely small values of a are not possible due to finite granularity – the lower limit of a is determined by the area of the smallest block which can be assigned to some of the partitions. Thus, it can be said that in the range of interest of the area ratio a , both approximations are applicable.

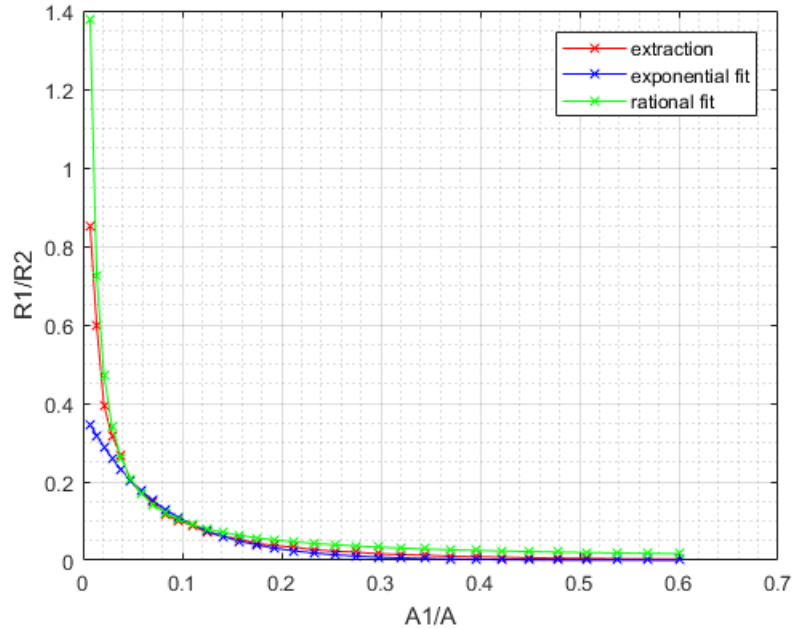


Figure 5.18 – Substrate resistivity ratio R_1/R_2 , extraction result and the two approximate analytical functions

Besides optimizing substrate noise, the algorithm should also take care about the negative side-effect of the methodology - the increase of the ground bounce in the newly formed domains. As explained in Chapter 2, the impact of the ground bounce on digital circuitry is determined by the time-domain parameters, primarily by the maximum ground bounce amplitude in time domain. A simplified sinusoidal switching current model, as used in sections 5.2-5.4, contains only one spectral peak in the frequency domain, so limiting the amplitude of that single peak in the frequency domain is equivalent to limiting the amplitude in time-domain. A more complex and more realistic switching current waveform, such as periodic triangular pulse, would contain multiple harmonic peaks at integer multiples of clock frequency, which all contribute to the maximum amplitude in time domain. Thus, all of those harmonic peaks have to be taken into account for limiting the increase of ground bounce when applying this substrate noise reduction methodology.

If the package parasitics and the total decoupling capacitance are known, the resonant frequency of the system can be determined, and the frequency range of the dominant ground bounce harmonics can be determined. In that case, ground bounce constraints (5.64)-(5.67) should be applied, and it would be enough to apply them only to the dominant ground bounce harmonics:

$$|\alpha_n| |N_d - 1 + K_{pn} N_d| < |N_{d1} + c K_{pn} N_d| \quad (5.74)$$

$$|N_{d1} - 1 + c K_{pn} N_d| < |\alpha_n| |N_d + K_{pn} N_d| \quad (5.75)$$

$$|1 - \alpha_n| |N_d - 1 + K_{pn} N_d| < |N_d - N_{d1} + (1 - c) K_{pn} N_d| \quad (5.76)$$

$$|N_d - N_{d1} - 1 + (1 - c) K_{pn} N_d| < |1 - \alpha_n| |N_d + K_{pn} N_d| \quad (5.77)$$

with $\alpha_n = I_{n1}/I_n$, $c = C_{d1}/C_d$ and $K_{pn} = 2jn\omega_0 C_d Z'_p(\omega_n)/N_d$. Note that an approximation $\omega_0 \approx \omega_1 \approx \omega_2$ is used here.

On the other hand, if the data about the package parasitics and the total decoupling capacitance are not known, the ground bounce constraints (5.25)-(5.26) can be applied:

$$|\alpha_n|(N_d - 1) < N_{d1} < |\alpha_n|N_d + 1 \quad (5.78)$$

$$(1 - |1 - \alpha_n|)N_d - 1 < N_{d1} < (1 - |1 - \alpha_n|)N_d + |1 - \alpha_n| \quad (5.79)$$

If the frequency range of the dominant ground bounce harmonics can't be estimated, the constraints have to be applied to all of the available harmonics. The larger the number of harmonics which satisfy the constraints, the lower the resulting ground bounce in time domain.

According to this, the ground bounce constraints are embedded in the optimizing algorithm in the following way: if for a particular iteration, i.e. for a particular partitioning solution, the number of harmonics satisfying the ground bounce constraints is lower than a chosen minimum number, the partitioning solution is rejected regardless the substrate noise attenuation achieved.

The proposed optimizing algorithm (an exhaustive search version) is presented in Fig. 5.19.

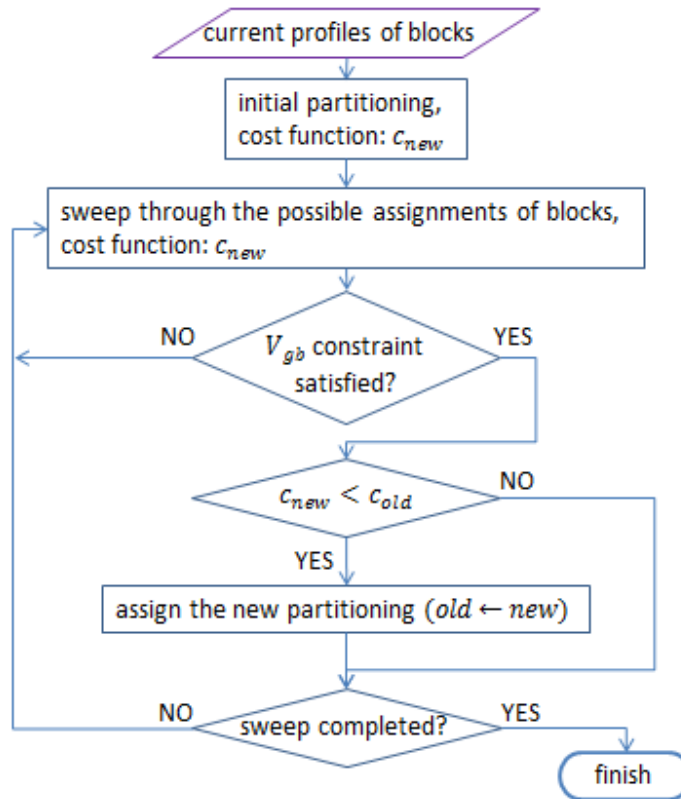


Figure 5.19 – An exhaustive search algorithm for optimizing the substrate noise by power domain separation and galsification

The proposed algorithm has been numerically evaluated in MATLAB on an example system consisting of 10 modules. The switching current waveforms of the modules have been modeled as periodic triangular pulses, with randomly assigned amplitudes, rise times and fall times. Total area of the chip is $0.64 \times 0.64 \text{ mm}^2$, like in the example from Fig. 5.4, while the areas of the blocks have been randomly assigned. Package has been modeled by a lumped RLC network, with values for the package parasitics being $R_p = 1 \Omega$, $L_p = 1 \text{ nH}$ and $C_p = 100 \text{ fF}$, while the decoupling capacitance has been selected to be $C_d = 10 \text{ pF}$, which corresponds to approximately 0.01 mm^2 in IHP 130nm

SiGe:C technology [25]. It has been assumed that the decoupling capacitance of the initial system remains uniformly distributed per area after partitioning. Substrate resistances have been extracted using the parameters of the for the IHP 130nm SiGe:C technology [25], while the rational function k_1/a^{k_2} has been used as an analytical approximation of the substrate resistance ratio R_1/R_2 . Clock frequency of the initial system is $f_0 = 100$ MHz, while the clock frequency offsets of the plesiochronous GALS clocking scheme after partitioning are selected as 1% and -1% for the domains D1 (closer to the victim) and D2 (further away from the victim), respectively.

The clock frequency of the initial system (i.e. the first spectral harmonic) has been selected as the target frequency of the algorithm. In other words, the fundamental spectral peak is targeted for minimization. This would correspond to the situation where the analog victim is a VCO, which is especially sensitive to substrate noise at lower frequencies. Ground bounce constraints from equations (5.78) and (5.79) have been used. They have been evaluated for all of the 32 available switching current harmonics, and the partitionings for which the constraints are satisfied for less than 31 harmonics have been rejected. As a cost function, the function c_2 from equation (5.71) has been used, which is the version without the decoupling capacitance C_d .

The algorithm uses a simpler model, without the decoupling capacitance and with the analytical approximation of the substrate resistance ratio. In order to have a better evaluation of the optimal partitioning delivered by the algorithm, the resulting spectra have been evaluated and plotted by using the more precise model, including the decoupling capacitance C_d , and using the extracted values of substrate resistances.

In Fig. 5.20, the substrate noise spectra of the initial synchronous system and the resulting GALS system with power domain separation have been presented. It can be seen that at each of the harmonics, a single peak of the synchronous spectrum has been replaced by two smaller peaks, corresponding to the two LSMs of the GALS system, i.e. the two power domains D1 and D2.

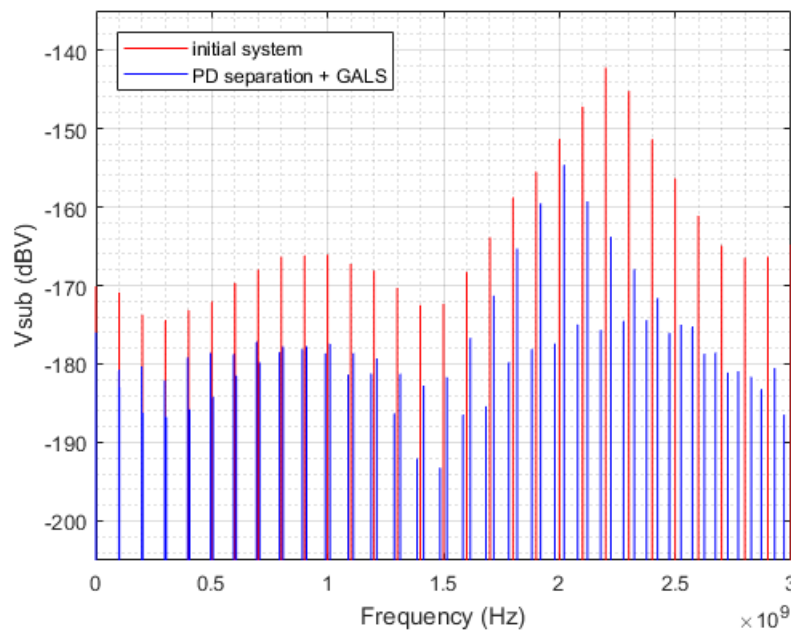


Figure 5.20 – Substrate noise spectrum of the initial synchronous system and the GALS system with power domain separation

In Fig. 5.21, the substrate noise spectrum is shown once more, but this time only around the targeted harmonic peak – the fundamental, at 100 MHz. The achieved spectral peak attenuation is $SPA_1 = 9.98$ dB. This is somewhat higher than the SPA value estimated by the algorithm, which is 9.69 dB. The difference is due to using a simpler model in the algorithm (neglecting the decoupling capacitance C_d in the cost function) and the more precise model for the evaluation. As expected based on the analysis from the Section 5.5, the presence of the decoupling capacitance contributes to a slight increase in the attenuation achieved, compared to the results obtained with the simpler model without the decoupling capacitance C_d .

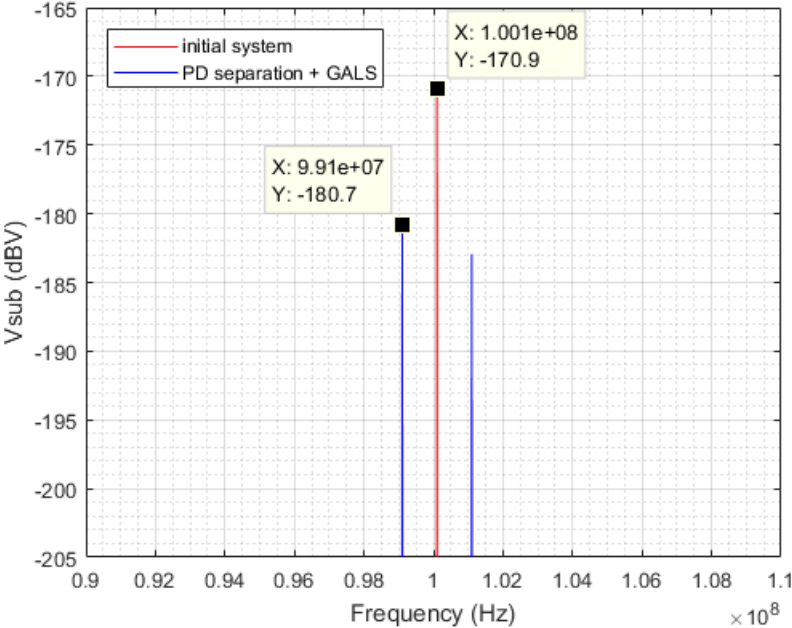


Figure 5.21 – Substrate noise spectrum of the initial synchronous system and the GALS system with power domain separation – the part of the spectrum around the targeted harmonic peak

The ground bounce has also been evaluated and it is presented in Fig. 5.22. It can be seen that the power domain separation resulted in an increased ground bounce in the domain D2. The highest ground bounce harmonic peak of the domain D2 is higher than the highest ground bounce harmonic peak of the initial system for 2.09 dB. The main parameter of the ground bounce in time-domain, the maximum ground bounce amplitude, can be calculated as a sum of magnitudes of all the spectral peaks. This way, from the data presented in this graph, it has been determined that the maximum ground bounce amplitude in the domain D2 has increased 1.17 times compared to the maximum ground bounce amplitude in the initial system.

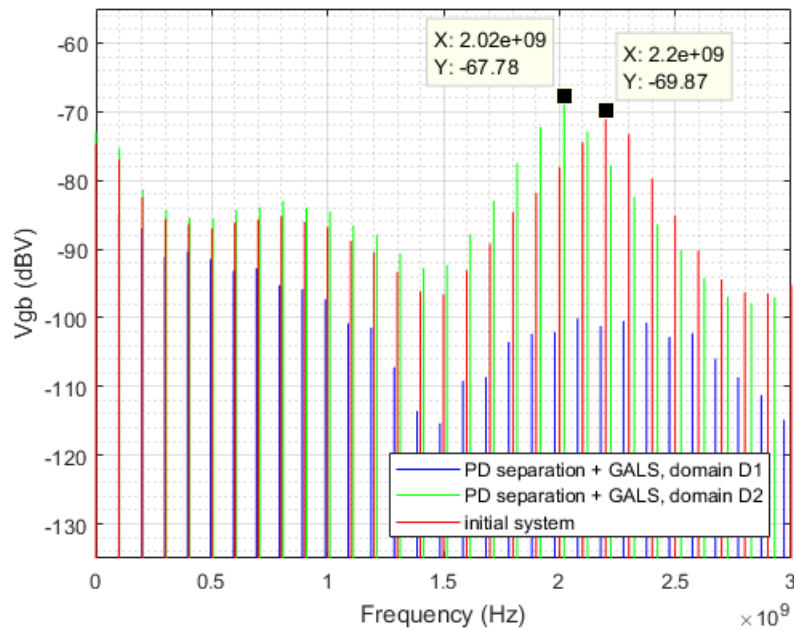


Figure 5.22 – Ground bounce spectrum of the initial synchronous system and the GALS system with power domain separation

In this section, the algorithm for implementing the methodology for substrate noise reduction by power domain separation and galsification has been presented, and the methodology has been numerically evaluated in MATLAB. The application of the methodology has resulted in a substantial reduction of the substrate noise – the spectral peak attenuation of 9.98 dB has been achieved at the targeted harmonic peak. The price to be paid is the 17% increase in the maximum ground bounce amplitude in the larger of the two resulting power domains.

In the next section, the simplifications of the methodology in some special cases will be discussed.

5.6. Some special cases of the methodology

5.6.1. Targeting low order substrate noise harmonics

In Chapter 4, the relation between power-balanced GALS partitioning and harmonic-balanced GALS partitioning has been analyzed. It has been concluded that for lower harmonics power-balanced GALS partitioning can be viewed as a special case of harmonic-balanced GALS partitioning. Since power is much easier to estimate than switching current harmonics, using power-based partitioning instead of harmonic-based partitioning is desirable whenever possible. Thus, it is useful to analyze whether in some special cases power can be used instead of switching current harmonics also in the substrate noise reduction methodology by power domain separation and galsification.

According to the equation (5.40), which has also been used for the cost function c_2 in Section 5.5, substrate noise ratio between the newly partitioned system and the initial system at the n^{th} harmonic can be represented as:

$$\left| \frac{V_{subp,n}}{V_{sub0,n}} \right| = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_{1,n}}{I_{0,n}}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d}{N_{d2}} \frac{I_{2,n}}{I_{0,n}}} \right| \right\} \quad (5.80)$$

Let's assume that for the n^{th} harmonic which is being observed the impact of the domain D1 is stronger, i.e. that:

$$\max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_{1,n}}{I_{0,n}}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d}{N_{d2}} \frac{I_{2,n}}{I_{0,n}}} \right| \right\} = \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d}{N_{d1}} \frac{I_{1,n}}{I_{0,n}}} \right| \quad (5.81)$$

This assumption is made without losing generality – the analysis would have been completely analogous if it had been assumed that the impact of the domain D2 was stronger.

Under the assumption (5.81), spectral peak attenuation at the n^{th} harmonic can be represented as:

$$SPA_n = 20 \log \left(\left(1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_d} \left| \frac{I_{0,n}}{I_{1,n}} \right| \right) \right) \quad (5.82)$$

which can further be written as:

$$SPA_n = SPA_{RN} + 20 \log \left(\left| \frac{I_{0,n}}{I_{1,n}} \right| \right) \quad (5.83)$$

where:

$$SPA_{RN} = 20 \log \left(\left(1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_d} \right) \right) \quad (5.84)$$

Note that if expression (5.47), for which the decoupling capacitance C_d is taken into account, was used as the starting point instead of the expression (5.40), spectral peak attenuation could also be represented as (5.81), only with a more complex expression for SPA_{RN} :

$$SPA_{RN} = 20 \log \left(\left(1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_d} \left| \frac{1 + 2jn\omega_1 C_{d1} \frac{Z'_p(\omega_1)}{N_{d1}}}{1 + 2jn\omega_0 C_d \frac{Z'_p(\omega_0)}{N_d}} \right| \right) \right) \quad (5.85)$$

Now, let's again assume periodic triangular pulse waveforms for the switching currents of both of the resulting LSMs i.e. power domains D1 and D2. As already explained in Chapter 2b, the n^{th} harmonic of a periodic triangular pulse switching current waveform i is:

$$F_n\{i\} = \frac{I_p}{j2\pi n} \left(\text{sinc}(\pi n f_{clk} t_r) - \text{sinc}(\pi n f_{clk} t_f) e^{-jn\pi f_{clk}(t_r+t_f)} \right) e^{-jn\pi f_{clk} t_r} \quad (5.86)$$

where t_r is the rise time of the switching current profile, t_f the fall time of the switching current profile, f_{clk} the clock frequency and I_p the peak current. In [16][160] it has been shown that, for low order harmonics, satisfying the condition:

$$n f_{clk} < \min \left(\frac{1}{\pi t_r}, \frac{1}{\pi t_f} \right) \quad (5.87)$$

the expression (5.86) can be approximated as:

$$F_n\{i\} \approx -\frac{j}{V_{DD}} P e^{-j\pi n \lambda} \quad (5.88)$$

where P is the average power, V_{DD} is the supply voltage, and λ is a parameter dependent on the shape of the switching current waveform:

$$\lambda = \left(t_r + \frac{t_r + t_f}{2} \right) f_{clk} \quad (5.89)$$

As discussed in Chapter 4, in case of plesiochronous GALS partitioning, the n^{th} harmonic of the switching current waveform i_0 of the initial synchronous system can be represented as the sum of the n^{th} harmonics of the switching current waveforms of the resulting LSMs. In case of partitioning into two LSMs, i.e. the domains D1 and D2, with switching current waveforms i_1 and i_2 :

$$F_n\{i_0\} = F_n\{i_1\} + F_n\{i_2\} \quad (5.90)$$

By using (5.90) and (5.88), it can now be written:

$$\frac{I_{0,n}}{I_{1,n}} = \frac{F_n\{i_0\}}{F_n\{i_1\}} = 1 + \frac{F_n\{i_2\}}{F_n\{i_1\}} = 1 + \frac{P_2 e^{-j\pi n \lambda_2}}{P_1 e^{-j\pi n \lambda_1}} = 1 + \frac{P_2}{P_1} e^{-j\pi n (\lambda_2 - \lambda_1)} \quad (5.91)$$

Spectral power attenuation of the substrate noise at the n^{th} harmonic can now be expressed as:

$$SPA_n = SPA_{RN} + 20 \log(|1 + k_P e^{-j\pi n (\lambda_2 - \lambda_1)}|) \quad (5.92)$$

where $k_P = P_2/P_1$.

Now, let's assume that power was used instead of switching current harmonics when estimating the substrate noise ratio and the spectral power attenuation, and let's analyze the error introduced by this estimation. The approximation introduced this way can be formulated as:

$$\alpha_n = \frac{I_1}{I_0} \approx \frac{P_1}{P_0} = \alpha_P \quad (5.93)$$

while the substrate noise ratio estimated this way would be:

$$\left| \frac{V_{subp,n}}{V_{sub0,n}} \right|_{P_est} = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_{d1}}{N_d} \frac{P_1}{P_0}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_{d2}}{N_d} \frac{P_2}{P_0}} \right| \right\} \quad (5.94)$$

which, under the assumption that the impact of the domain D1 is dominant for the observed harmonic, would result with the following estimation for the spectral peak attenuation:

$$SPA_{n,P_est} = SPA_{RN} + 20 \log\left(\left|\frac{P_0}{P_1}\right|\right) = SPA_{RN} + 20 \log(|1 + k_P|) \quad (5.95)$$

The error introduced by using power instead of the n^{th} switching current harmonic can further be calculated as:

$$\Delta SPA_n = SPA_{n,P_est} - SPA_n = 20 \log(|1 + k_P|) - 20 \log(|1 + k_P e^{-j\pi n (\lambda_2 - \lambda_1)}|) \quad (5.96)$$

which can also be expressed as:

$$\Delta SPA_n = SPA_{n,P_est} - SPA_n = 20 \log\left(\left|\frac{1 + k_P}{1 + k_P e^{-j\pi n (\lambda_2 - \lambda_1)}}\right|\right) \quad (5.97)$$

As it has been shown in the previous sections of this chapter, for optimal attenuation, the domain D2 is much larger in area and switching current than the domain D1. In case that the average

power consumption of the domain D2 is much bigger than the average power consumption of the domain D1, i.e. that:

$$k_p = \frac{P_2}{P_1} \gg 1 \quad (5.98)$$

the equation (5.97) reduces to:

$$\Delta SPA_n \approx 20 \log \left(\left| \frac{k_p}{k_p e^{-j\pi n(\lambda_2 - \lambda_1)}} \right| \right) = 20 \log \left(\left| \frac{1}{e^{-j\pi n(\lambda_2 - \lambda_1)}} \right| \right) = 0 \quad (5.99)$$

This means that for low order harmonics, satisfying the equation (5.87), and in case that the power consumption in the domain D2 is much larger than the power consumption in the domain D1, the error introduced by using power consumption instead of a switching current harmonic becomes negligible, which justifies the simplification of the methodology by using power consumption.

It can also be noticed that for most of the possible values of λ , the error is much smaller than the worst-case error, which makes the simplification of the methodology applicable for moderate values of k_p as well. The range of λ is from 0 (corresponding to an infinitesimally narrow triangle right after the clock edge) to 1.5 (corresponding to the case where $t_r = T_{clk}$ and $t_f \rightarrow 0$). However, most of the theoretically possible values of λ are little probable, and wouldn't occur when representing a realistic system. The shape of the switching current waveform of a block depends on the distribution of the signal paths within the block. Register switching happens at the beginning of the period, right after the clock edge, and in the combinatorial logic, most of the paths are usually significantly shorter than the critical path. Thus, for most of the systems, the peak of the triangular waveform would be in the first half of the period. Under this assumption, the range for λ narrows from the upper side from 1.5 down to 1. In Fig. 5.23, the error from (5.97) has been evaluated at the fundamental harmonic for $k_p = 5$, in the range of λ_1 and λ_2 between 0 and 1. As it can be seen, even for this rather small value of k_p , the difference caused by the simplification is less than 2 dB for most of the values of λ_1 and λ_2 . Only for the extreme different values of λ_1 and λ_2 (one of them being close to 0 while another of them is close to 1), the error is somewhat larger, with the maximum value being 3.39 dB.

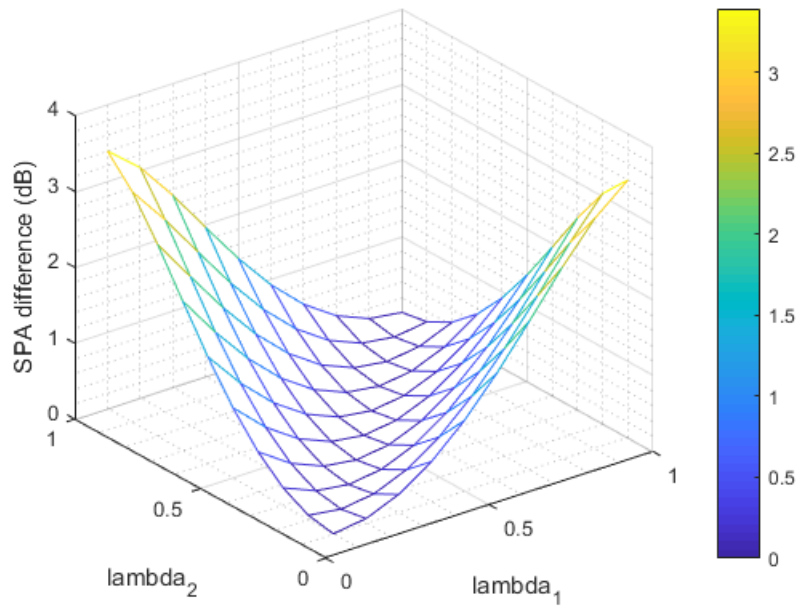


Figure 5.23 – The error in SPA estimation introduced by using power consumption instead of a power harmonic, fundamental harmonic, $k_p = P_2/P_1 = 5$

The same graph is shown again in Fig. 5.24, this time for $k_p = 10$. This value of k_p is already so large, that the error is below 2 dB for all of the values of λ_1 and λ_2 .

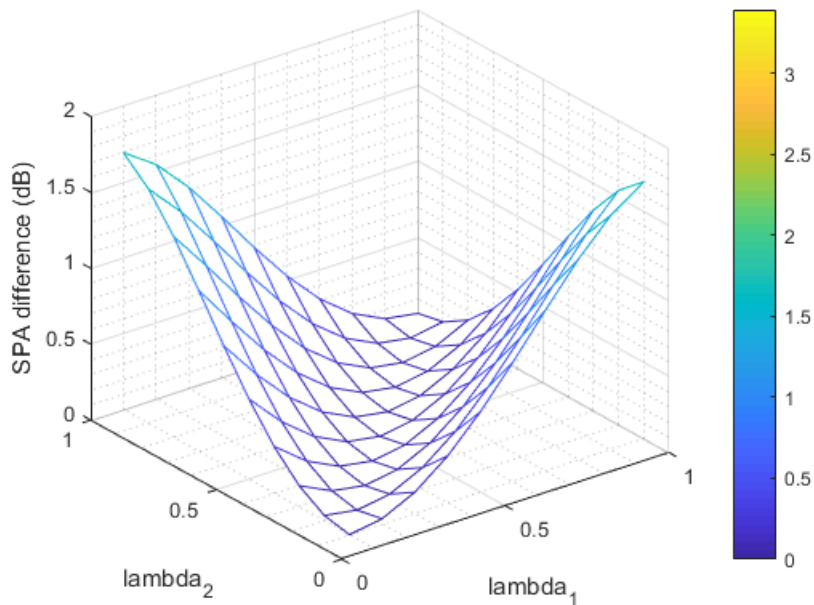


Figure 5.24 – The error in SPA estimation introduced by using power consumption instead of a switching current harmonic, fundamental harmonic, $k_p = P_2/P_1 = 10$

The simplification of using power consumption instead of a switching current harmonic can easily be incorporated into the algorithm described in Section 5.5. The cost function c_2 from the equation (5.71) then turns into:

$$c_3 = \max \left\{ \left| \frac{1}{1 + \frac{R_1}{R_2} \frac{N_d P_1}{N_{d1} P_0}} \right|, \left| \frac{1}{1 + \frac{R_2}{R_1} \frac{N_d P_2}{N_{d2} P_0}} \right| \right\} \quad (5.100)$$

The ground bounce constraints can also be modified by replacing $\alpha_n = I_{1,n}/I_{0,n}$ with $\alpha_p = P_1/P_0$, which, knowing that α_p is a real number, results in:

$$\alpha_p(N_d - 1) < N_{d1} < \alpha_p N_d + 1 \quad (5.101)$$

$$\alpha_p N_d - 1 < N_{d1} < \alpha_p(N_d - 1) + 1 \quad (5.102)$$

Note, however, that this approximation is valid only for lower switching current harmonics. Thus, while the ground bounce at low frequencies will still remain properly constrained, this doesn't have to apply to the higher harmonics, which may have large impact on the total ground bounce. As a consequence, the ground bounce constraints (5.101) and (5.102) are less tight than the ground bounce constraints (5.78) and (5.79), and applying them in the optimizing algorithm would result in a larger increase in the ground bounce. This limits the application to the cases where the power delivery network is well dumped, so that the dominant part of the ground bounce spectrum is at lower frequencies, or to the cases where the ground bounce margin in the initial system is large, so that the increase in ground bounce is not critical.

5.6.2. Power approximately uniformly distributed per area

A further simplification can be achieved if the modules out of which the system is composed don't differ much from each other in terms of power per area, i.e. if the switching activity is approximately uniformly distributed per area. Let's observe a system which consists of N_m modules, with average power consumptions $P_{m1}, P_{m2}, \dots, P_{mN_m}$ and areas $A_{m1}, A_{m2}, \dots, A_{mN_m}$, such that for each of those modules:

$$\frac{P_{m1}}{A_{m1}} \approx \frac{P_{m2}}{A_{m2}} \approx \dots \approx \frac{P_{mN_m}}{A_{mN_m}} \approx \frac{P_0}{A} \quad (5.103)$$

where P_0 and A are the total average power consumption and area of the initial system, respectively.

If the approximation (5.103) is valid, it can be written that:

$$\alpha_{pmi} = \frac{P_{mi}}{P_0} \approx \frac{A_{mi}}{A} = a_{mi} \quad , \quad i = 1, 2, \dots, N_m \quad (5.104)$$

This, combined with the previously introduced approximation (5.93), roughly corresponds to the case where switching activity is uniformly distributed per area, which has been theoretically discussed in sections 5.2-5.4. Note that, for every feasible partitioning, the following condition is fulfilled:

$$\alpha \approx \alpha_p = \frac{P_1}{P_0} \geq \min\{\alpha_{pmi}\} = \min\left\{\frac{P_{mi}}{P_0}\right\} \quad (5.105)$$

since at least one module has to be assigned to the domain D1.

Let α_{opt} be the value of $\alpha \approx \alpha_p = P_1/P_0$ for which the optimum substrate noise spectral peak attenuation could theoretically be reached, if the system granularity was infinite. As already explained in sections 5.2-5.4, in case that the switching activity is uniformly distributed per area, the

optimum attenuation is theoretically expected for small values of α , in the range where the appropriate choice for N_d is $N_{d1} = 1$. When $N_{d1} = 1$ is applied to the constraints (5.101) and (5.102), a following constraint for α can be derived:

$$0 < \alpha < \frac{1}{N_d - 1} \quad (5.106)$$

The exact value of α_{opt} depends on the substrate resistance ratio R_1/R_2 , and it has approximately the same value both for the case when only power domain separation is applied and for the case when galsification is applied as well. As seen in Section 5.5, the substrate resistance ratio can be approximated by a rational fitting function (5.73). Under the assumption (5.104), for any pair of parameters k_1 and k_2 , the theoretical optimum α_{opt} can be approximately calculated from:

$$\frac{d}{d\alpha} \left(\frac{1}{1 + \frac{k_1}{\alpha^{k_2}}} N_d \alpha + \frac{1}{1 + \frac{\alpha^{k_2}}{k_1}} \frac{N_d}{N_d - 1} (1 - \alpha) \right) = 0 \quad (5.107)$$

For example, in the example chip from previous sections, in IHP 130nm SiGe:C technology [25] and with dimensions $0.64 \times 0.64 \text{ mm}^2$, the fitting parameters are, as calculated in Section 5.5, $k_1 = 0.01$ and $k_2 = 1$. In that case, equation (5.107) results in:

$$\alpha_{opt} = \frac{1}{100} \left(\sqrt{\frac{N_d + 100}{N_d - 1}} - 1 \right) \quad (5.108)$$

which, for example, would result in $\alpha_{opt} = 0.049$ for $N_d = 4$ and $\alpha_{opt} = 0.062$ for $N_d = 3$.

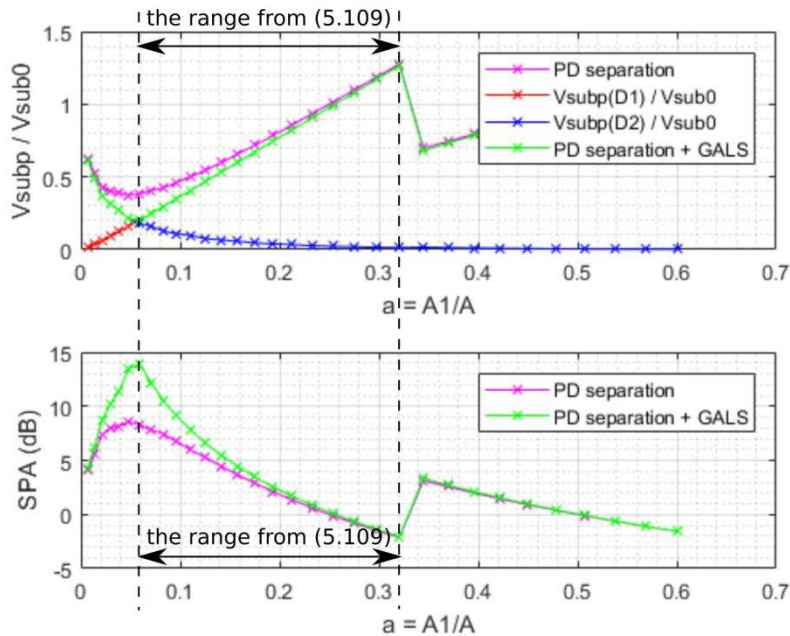


Figure 5.25 – Substrate noise ratio and spectral peak attenuation, for an example system with uniform distribution of power per area, with the range from (5.105) marked

Now, let's observe the range of α between the theoretical optimum point $\alpha = \alpha_{opt}$ and the point $\alpha = 1/(N_d - 1)$, where N_{d1} has to be changed from 1 to 2 due to the ground bounce constraint (5.106):

$$\alpha_{opt} < \alpha < \frac{1}{N_d - 1} \quad (5.109)$$

In this range, the impact of the close domain D1 to the substrate noise is dominant, and the substrate noise ratio V_{subp}/V_{sub0} is a monotonically rising function. In Fig 5.25, substrate noise ratio V_{subp}/V_{sub0} and spectral peak attenuation for the example from Section 5.2 are shown again, with this range marked.

If, additionally, each of the modules of the system is large enough in terms of power and area, so that:

$$(\forall_i)(\alpha_{pmi} > \alpha_{opt}) \quad (5.110)$$

no feasible partitioning resulting in α below the range (5.109) is possible, due to (5.105). Consequently, taking into account the monotonicity of V_{subp}/V_{sub0} in the range (5.109), the optimum feasible partitioning for this system is reached for the smallest possible α , which is:

$$\alpha = \min \{ \alpha_{pmi} \} \quad (5.111)$$

In other words, assigning only one module to the domain D1, and precisely the model with the lowest average power, would result in the optimum feasible partitioning for such a system.

5.7. A comparison to other physical and system level substrate noise reduction methodologies

The methodology presented in this chapter is partially a physical-level methodology (due to using the isolating property of lightly-doped substrates and applying a power-domain separation) and partially a system-level methodology (due to applying galsification to an initially synchronous design). Thus, in this section, it will be compared to both physical- and system- level methodologies. Prior to that, main limitation of the methodology will be discussed.

5.7.1. Main limitations of the substrate noise reduction methodology based on power domain separation and galsification

Similar to PB-pGp and HB-pGp, the methodology presented in this chapter also has a dependence on system granularity. The minimum achievable size of the power domain closer to victim is limited by the minimum block – if this block is too large, the optimum point as shown in Sections 5.2 to 5.4 cannot be reached. On the other hand, the PD closer to the victim shouldn't be too small. Its minimum allowable size is limited by the size of the victim – it has to be large enough, so that a floorplan as shown in Fig. 5.3 is feasible, i.e. so that it can „shield“ the victim from the larger partition which is further away from it.

If higher order harmonics are targeted, i.e. if the power-based version of the methodology can't be used, the dependence on the correct estimation of the switching current profile becomes a drawback, similar to HB-pGp. Here, it's even more expressed, since the EDA tool would have to provide the estimation for two different PDN configurations. The reliability of the methodology can be expected to improve with the development in EDA tools for estimating current profiles. However, if lower order harmonics are targeted, i.e. if the power-based version of the methodology can be used, the power estimations, which are already quite reliable with standard tools, are used instead.

Finally, similar to HB-pGp, the equal ground bounce approximation has been used to derive this methodology as well. Thus, the methodology is only applicable to systems where package parasitics are dominant compared to on-chip parasitics.

5.7.2. Comparison to other physical-level methodologies for substrate noise reduction

As it has been presented in Chapter 2, a variety of physical level methodologies can be used for substrate noise reduction, ranging from standard ones like using p+ guard rings, deep trench isolation or triple-wells, to more sophisticated and complex ones.

More sophisticated physical-level methodologies, such as porous silicon, silicon post-processing or using of Faraday-cage-like isolation can provide huge substrate noise attenuations, up to several tens of dB. However, all of these methodologies are expensive both in terms of area usage and in terms of additional technology options and steps needed for them. Thus, despite outperforming the methodology presented in this chapter, they would not be suitable for a low-cost MSIC solution.

Attenuation reported for SOI is comparable to that achievable by the methodology described in this chapter. However, at frequencies in the order of GHz, SOI loses its isolating property. This can be counteracted by using some technology enhancements, such as GPSOI, but this again means introducing additional non-standard technology steps, and is thus not applicable for a low-cost MSIC solution. Triple wells can provide even better attenuation at lower frequencies, but they are also not readily available in every technology. Similar to SOI, they lose their isolating property at higher frequencies, where they can even worsen the substrate noise coupling. Additionally, their performance is strongly dependent on biasing parasitics.

The p+ guard rings, on the other hand, are readily available in every technology, they have relatively low area cost, and may provide comparable or even better attenuation even at higher frequencies. Their performance, however, is highly dependent on biasing. Bias parasitics, if too big, can completely remove the attenuation. On the other hand, for proper operation, a separate, quiet biasing is needed. Biasing a p+ guard ring with digital ground would only introduce an additional noise injection point closer to the victim and worsen the noise coupling. Also, biasing a p+ guard ring with analog ground would introduce an additional noise reception point, by which the substrate noise would couple directly to the victim's ground. Thus, for an optimal performance of a p+ guard ring, an extra pad for biasing to the quiet ground would be needed.

Power domain separation and galsification, on the other hand, doesn't require an extra pad – it is based on redistributing the existing pads among the two power domains. Thus, it is especially fitting for a low-cost solution in a small ASIC, where the number of package pins is limited, so that no extra pin can be reserved for the quiet ground.

The frequency range which can be targeted by power domain separation and galsification depends on whether the galsification is power or harmonic based. Power based galsification targets lower harmonics for reduction. This frequency limitation, however, is not physical (like for SOI or DNW); it is defined by the clock frequency instead. For example, let's assume a system with a 1 GHz clock. Some physical methodologies, such as SOI and DNW, lose their noise reduction performance in the GHz range. So, if strong substrate noise harmonics are present in the 1 GHz - 10 GHz range, these methodologies most probably can't provide a satisfactory substrate noise attenuation. On the other hand, power domain separation and power based galsification would provide a large attenuation in this range, because the harmonics in the range are of lower order. Additionally, if harmonic and area based attenuation is applied, targeting even higher harmonics at even higher frequencies would also be possible.

5.7.3. Comparison to other system-level methodologies for substrate noise reduction

As already explained in Section 2.3.2 and 4.5.3, synchronous methodologies for substrate noise reduction perform generally well at spectral peak attenuation of higher harmonics, but they offer marginal attenuation (if any) at lower harmonics. Power domain separation and galsification, on the other hand, can provide quite a high attenuation at low harmonics as well – in the numerical evaluation in Section 5.5, a spectral peak attenuation of almost 10 dB was reached.

Harmonic-balanced plesiochronous GALS partitioning, presented in Chapter 4, as well as power-balanced plesiochronous GALS partitioning as its special case, provide an attenuation of $20\log(M)$, where M is the number of LSMs. Thus, they can also deliver a superior result at lower order harmonics. The power and area overheads of a GALS partitioning depend on the type of clock domain crossing wrapper which is applied, as well as on the number of partitions and the number of signals crossing between the LSMs. Thus, HB-pGp and PB-pGp could be expected to have similar overheads to the power domain separation and galsification methodology, if the resulting GALS system has the same number of LSMs.

Note that in power domain separation and galsification, the initially synchronous system is partitioned into two LSMs. If a single power domain was kept and PB-pGp or HB-pGp into two power domains was applied, the achieved attenuation would be around 6 dB attenuation, which is a lower result for the same power and area overhead. Thus, power domain separation and galsification can be preferable to PB-pGp and HB-pGp, if allowable power and area overheads are small, which might be especially the case in small ASICs. If there's enough room for more power and area overheads, so that partitioning with a larger number of LSMs is feasible, a larger attenuation can be reached by PB-pGp and HB-pGp.

The methodology presented in this chapter can also easily be combined with other system level noise methodologies. For example, it is possible to first separate the initially synchronous

design into two power domains which are then galsified, and further partition the larger power domain into several LSMs by using PB-pGp or HB-pGp. That way, an even larger attenuation could be achieved. The investigation of such combined methodologies could be an interesting topic for a future work.

5.8. Summary

In this chapter, a substrate noise reduction methodology based on power domain separation and galsification has been presented. Unlike the methodology presented in Chapter 4, this methodology makes use of the isolating property of lightly-doped substrate, and is thus applicable only in this kind of substrates.

The methodology achieves substrate noise reduction in two steps. The first step consists of splitting the initial synchronous design into two power domains – the first one smaller and closer to the victim, and the second one larger and further away from the victim. The supply and ground pads are distributed among the power domains so that the ground bounce in the smaller domain reduces compared to the original system, while the ground bounce in the larger domain increases. The floorplanning of the two power domains should be such that the smaller one effectively „shields“ the victim from the larger one. The achievable substrate noise reduction by power domain separation has been theoretically examined by using a simplified PDN model without decoupling capacitance, and the optimal point for the separation has been determined. The penalty of increased ground bounce in the larger power domain has been analyzed as well.

In the second step, the design is galsified, so that the two power domains become the LSMs of the resulting GALS design. A plesiochronous GALS clocking scheme has been used. The further improvement of substrate noise reduction by galsification has been examined on the same model, and it has been shown that several additional dB of attenuation can be achieved that way. Further, the model has been extended by introducing the decoupling capacitance, and the impact of this capacitance was analyzed. It has been shown that in this, more realistical model, the attenuation which can be reaches becomes even larger, while the penalty of the increased ground bounce in the larger power domain decreases.

Based on the theoretical analysis, an algorithm for reaching the optimal substrate noise reduction by power domain separation and GALS partitioning has been developed. In the most general version of the methodology, switching current harmonics generated by the blocks within the design, as well as their area, are used in order to determine this optimal partitioning. The attenuation is dependent, among else, on the ratio of the equivalent substrate resistances between each of the aggressor to the victim. An analytical function to estimate this ratio based on the area ratio between the LSMs has been proposed, thus enabling faster execution of the algorithm. This function needs two substrate resistivity extractions as fitting parameters, which can be calculated prior to starting the optimization procedure. The methodology has been numerically evaluated in MATLAB with the fundamental harmonic being targeted, resulting in an attenuation of 9.98 dB.

Further it has been shown that in some special cases the methodology can be significantly simplified. If low order substrate noise harmonics are targeted for reduction, average power consumption of the blocks can be used in the algorithm instead of switching current harmonics. This makes the methodology easier to implement, because power consumption is easier to estimate than switching current harmonics. If the power is approximately uniformly distributed per area, and if the granularity of the design is coarse, the methodology can even be simplified so that the optimal partitioning can be determined manually, without applying the optimization algorithm.

Finally, the methodology has been compared to the other substrate noise reduction methodologies, both physical-level and system-level. The methodology has similar drawbacks as the methodology presented in Chapter 4 - the dependence on granularity, the relying on switching current profile estimation and the limitation of applicability to small systems where package parasitics are dominant to on-chip parasitics.

Compared to the more sophisticated and complex physical-level methodologies, such as porous silicon and Faraday-Cage structures, this methodology provides much less attenuation, but it also introduces less area overheads, and it doesn't require any special technology enhancements. If switching current harmonics are used in the optimization algorithm, the methodology can be used for higher frequencies, unlike SOI or triple wells. Even if average power consumption is used instead of switching current harmonics, and the methodology becomes applicable only to low order harmonics, the frequency limitation is not physically hard, but only relative to the initial synchronous clock frequency. At low frequencies, the methodology can provide attenuation comparable to that provided by SOI. P+ type guard rings, on the other hand, can provide superior results both at high and low frequencies. However, the performance of guard rings is extremely dependent on the quality of biasing, and for optimal operation they would need to be biased to a separate quiet ground, which would require a separate package pin. The methodology based on power domain separation and galsification, on the other hand, only redistributes the existing supply and ground pins among the two power domains, and is thus suitable for MSICs in small packages where the number of available pins is limited.

Compared to the synchronous system-level substrate noise reduction methodologies (SCS and CM), this methodology provides a significantly better performance at lower frequencies. Compared to harmonic-balanced and power-balanced plesiochronous GALS partitioning, it provides better attenuation for the same number of LSMs (and thus the same overheads). However, if HB-pGp and PB-pGp have a larger number of LSMs, they would outperform the methodology presented in this chapter. It is important to notice that GALS methodologies presented in Chapter 5 and Chapter 4 are complementary, i.e. they can be combined in order to further improve the substrate noise reduction.

In the next chapter, the application of a special case of the methodology described in this chapter on a test chip is presented as a case study, and the results of measurements on silicon are demonstrated.

6. Case study: *SGE* chip

In the previous chapter, harmonic-and-area-based plesiochronous GALS partitioning with power domain separation has been presented. The methodology has been theoretically analyzed and evaluated by numerical simulations in MATLAB. At low frequencies, harmonic-and-area-based plesiochronous GALS partitioning can be simplified by using the average power instead of switching current harmonics. This special case has been named power-and-area-based plesiochronous GALS partitioning with power domain separation. For the purpose of silicon validation of this methodology, a test chip named *SGE* (*power domain Separation and Galsification Experiment*) has been developed.

The *SGE* chip contains two completely separate cores with the identical functionality: a synchronous core, which is the reference design, and a GALS core resulting from galsification of the reference synchronous design. In this chapter, details of implementation of this test ASIC are presented, as well as the results of noise measurements on silicon.

In Section 6.1, the architecture of the original synchronous reference design, the GALS partitioning strategy and the architecture of the resulting GALS design are presented, together with the floorplan, layout and the packaging of the test chip. In Section 6.2, the results of noise measurements are presented. The noise reduction achieved by applying the methodology is demonstrated, and the power and area overheads of the GALS design are discussed. Finally, in Section 6.3, the conclusions are drawn.

6.1. Architecture of the *SGE* chip and GALS partitioning strategy applied

6.1.1 Architecture of the reference design

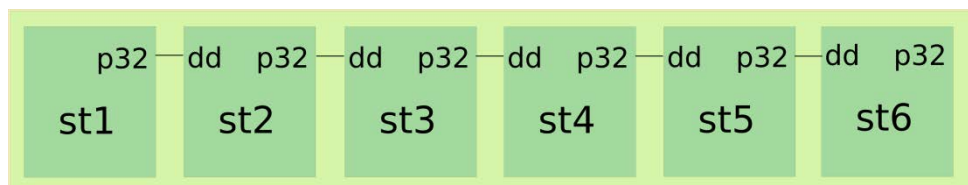


Figure 6.1 – Cascaded stages of the reference design – a simplified representation

The reference design consists of 6 cascaded stages, as shown in Fig. 6.1. Each stage has one 8-stage 32/16 bit divider, and one 8-stage 32*16 bin multiplier. Each of the dividers has a 32-bit dividend dd_{32} and a 16-bit divisor ds_{16} as inputs, and a 32-bit quotient q_{32} and a 16-bit remainder r_{16} as outputs. Each of the multipliers has a 32-bit factor f_{32} and a 16-bit factor f_{16} as an input, and a 32-bit product p_{32} as an output. A shift register is used to set all the configurable registers, and to capture outputs of each divider and multiplier. BIST checks outputs for each stage. A

simplified block diagram of the first stage, *st1*, is shown in the Figure 6.2. Stages *st2-st6* are similar, except that they don't have a linear-feedback shift register (LFSR), but receive their *dd32* inputs from the previous stage instead.

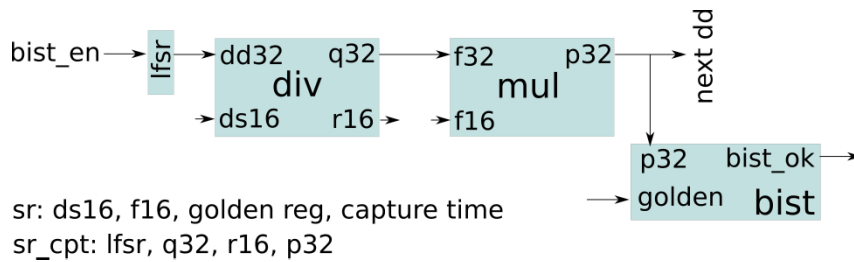


Figure 6.2 – A simplified block diagram of the first stage in the pipeline of the SGE chip

The reference design is synchronous. It contains only one clock domain using the CLK clock, which is provided to the chip as an input signal. The nominal clock frequency is 50 MHz. The operation of the core, including the BIST operation, is enabled by a BIST_EN signal. By an output multiplexer, it can be selected which one of the six *bist_ok* signals will be forwarded to the BIST_OK output. This way, in case of an error, the stage in which the error occurs can be localized. In the GALS version, the multiplexer can additionally be used to forward the two LSM clocks to the BIST_OK output, thus enabling the trimming of the GALS clock frequencies. The multiplexer is shown in Figure 6.3.

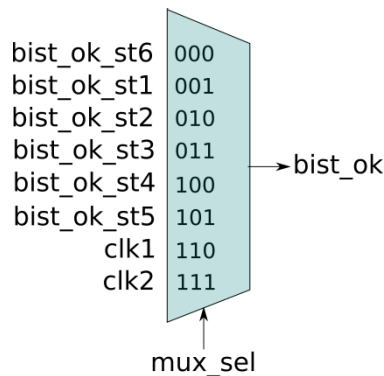


Figure 6.3 – BIST_OK multiplexer

For configuring the core, and for capturing the results, a shift register is used. The shift register has a serial input SREG_IN and a serial output SREG_OUT, its read/write operation is enabled via the SREG_EN signal. In both core versions, the shift register is clocked by the CLK clock, and it is reset by the RSTN_S reset signal.

Inputs for *ds16*, *f16*, and golden registers for BIST for each of the stages, as well as the *mux_sel* signal for configuring the BIST_OK multiplexer, are set via the shift register. Shift register can capture the output of LFSR, the outputs of each divider (*q32*, *r16*) and the output of each multiplier (*p32*). Additionally, in the GALS version, shift register can capture the output of the asynchronous FIFO, and it sets the frequency of the LSM clocks.

Capturing the data via the shift register is shown in Figure 6.4. The capture counter register *cpt_cnt* is loaded from the shift register, and it starts counting when the operation is enabled via BIST_EN. The purpose of this counter is to determine the capture time, i.e. the number of clock cycle in which the data is captured. When it counts down to zero, capture enable pulse signal

`cpt_en` is risen, which triggers the capturing of data to the capture register `cpt_reg`. When the shift register clock starts again, the content of the capture register `cpt_reg` is written into the shift register, and when the `SREG_EN` is raised, the output of the shift register is forwarded to the serial output. This way of capturing the output data was chosen because it can also be used for GALS, where the signals which need to be captured belong to a different clock domain than the shift register.

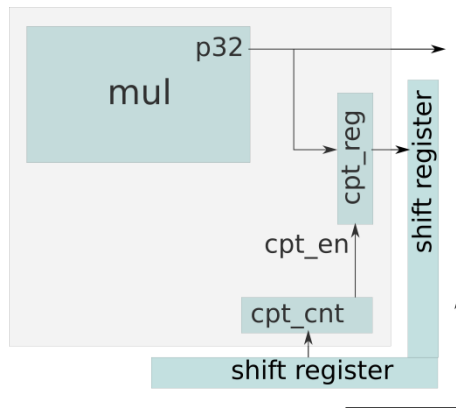


Figure 6.4 – Capturing the data via the capture register

A detailed architecture of the original synchronous design is presented in Fig. 6.6, at the end of the Subsection 6.1.2. The total number of supply/ground pad pairs assigned to the original synchronous design is $N_d = 3$.

6.1.2 GALS partitioning goal, the applied partitioning methodology and the architecture of the resulting GALS design

The goal of the partitioning is to optimally reduce substrate noise at lower frequencies, i.e. at lower clock harmonics. Thus, the power and area based plesiochronous GALS partition methodology can be used, provided that the power in the domain which is further from the victim is sufficiently larger than the power of the domain which is closer to the victim.

Prior to determining the optimum partitioning, the supply/ground pad pairs have to be properly distributed among the two newly created power domains. According to the discussion in Chapter 5, the optimum attenuation is achieved when only one supply/ground pad pair is assigned to the power domain closer to the victim. Since the original synchronous system has $N_d = 3$ supply/ground pad pairs, this means that $N_{d1} = 1$ pad pair gets assigned to the power domain closer to the victim, while $N_{d2} = 2$ pad pairs get assigned to the power domain further away from the victim.

As the next step in applying the methodology, the modules of the initial system which shouldn't be split during partitioning have to be identified. It is important to keep the power and area overheads as small as possible. Thus, the interfaces between the modules should have as few signals as possible. Also, in order to enable the testability of each of the partitions, regardless of the modules assigned to it, each of the modules should be testable. Pipeline stages `st1-st6` satisfy both of these conditions; thus, they have been identified as the modules of the system to be used by the partitioning methodology.

Table 6-I – Power/Area ratios for each of the modules (pipeline stages) in the SGE chip. P_{sw} – switching power; P_T – total power; A – area

Module	P_{sw} (mW)	P_T (mW)	α_{Psw}	α_{PT}	A (μm^2)	P_{sw}/A (mW/ μm^2)	P_T/A (mW/ μm^2)
ST1	1.45×10^{-3}	4.14×10^{-3}	0.179	0.179	152171	9529	27206
ST2	1.48×10^{-3}	4.05×10^{-3}	0.183	0.175	139391	10618	29055
ST3	1.25×10^{-3}	3.61×10^{-3}	0.154	0.156	137787	9072	26200
ST4	1.3×10^{-3}	3.7×10^{-3}	0.160	0.160	139319	9331	26558
ST5	1.38×10^{-3}	3.94×10^{-3}	0.170	0.171	139196	9914	28305
ST6	1.24×10^{-3}	3.66×10^{-3}	0.153	0.158	138837	8931	26362

The power (switching and total), area and power to area ratio are shown in Table 6-I for each of the stages st1-st6, i.e. for each of the modules of the initial system. For power estimation, dynamic vector based PrimeTime power analysis on the first available netlist of the synchronous design was used, while for area estimation, the cell area reported in synthesis was used.

As it can be seen from the table, all of the modules have relatively similar power to area ratios, deviating only up to 11% from the medium value. Thus, it can be considered that power is approximately uniformly distributed per area on module level. Additionally, $\alpha_{pm} = P_{module_m}/P_{total}$ parameters are in the range 0.153 – 0.183, which should be substantially above the α_{opt} value. Thanks to this, the simplification from Section 5.6.2 can be applied: The optimum partitioning is reached if only the module with the lowest power is assigned to the closer partition, while the remaining modules are assigned to the partition which is further away from the victim. As it can be seen from the table, the lowest values of power are obtained for the modules st3 and st6. However, for st3 two synchronization FIFOs would be needed (towards st2 and towards st4), while for st6 only one is enough (towards st5). Since the difference in power between st3 and st6 is very small, and the overheads for st3 would be significantly larger, st6 was selected to be assigned to the closer partition.

A simplified representation of the resulting GALS design is shown in Fig. 6.5. The GALS design has two locally synchronous modules (LSMs), which are placed in two power domains, P1 and P2. Each of the modules has its own clock, which is provided by a corresponding ring oscillator. Synchronization is done by asynchronous FIFO with Johnson code. The FIFO is placed in the larger power domain (P2). Besides these two clock domains, there's also a third one, containing only the shift register and using the clock from the CLK pad. This third clock domain, however, is not active during normal core operation.

When the asynchronous FIFO is full, a `fulln` signal is set to "0", blocking the data flow in the `clk2` clock domain. This has been achieved by adding a `ready` signal to each flip-flop (FF). When this `ready` signal is "1", the FF operates normally – the value from the input is written to it at a rising clock edge. When the `ready` signal is "0", instead of taking a new value from the input, the current state gets rewritten in the FF. In the `clk2` clock domain, the `fulln` signal is used as the `ready` signal. This way, when the `fulln` signal is set to "0", all of the registers of the clock domain `clk2` keep the same state. This lasts until a data set is read from the FIFO, leaving the space for writing a new data set. Then, the `fulln` signal is set back to "1", and the registers in the `clk2` domain resume their normal operation.

In an analogous way, when the asynchronous FIFO is empty, an `emptyn` signal is set to “0”, blocking the data flow in the `clk1` domain, i.e. causing the registers of this clock domain to keep their state. This lasts until a data set is written in the FIFO. Then, the `emptyn` signal is set back to “1”, and the registers in the `clk1` domain resume their normal operation.

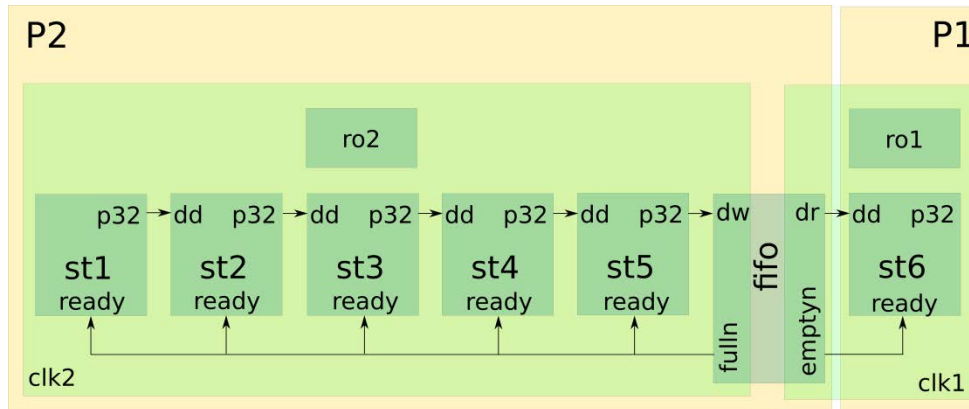


Figure 6.5 – A simplified representation of SGE chip partitioning
 A detailed architecture of the resulting GALS design is shown in Fig. 6.7.

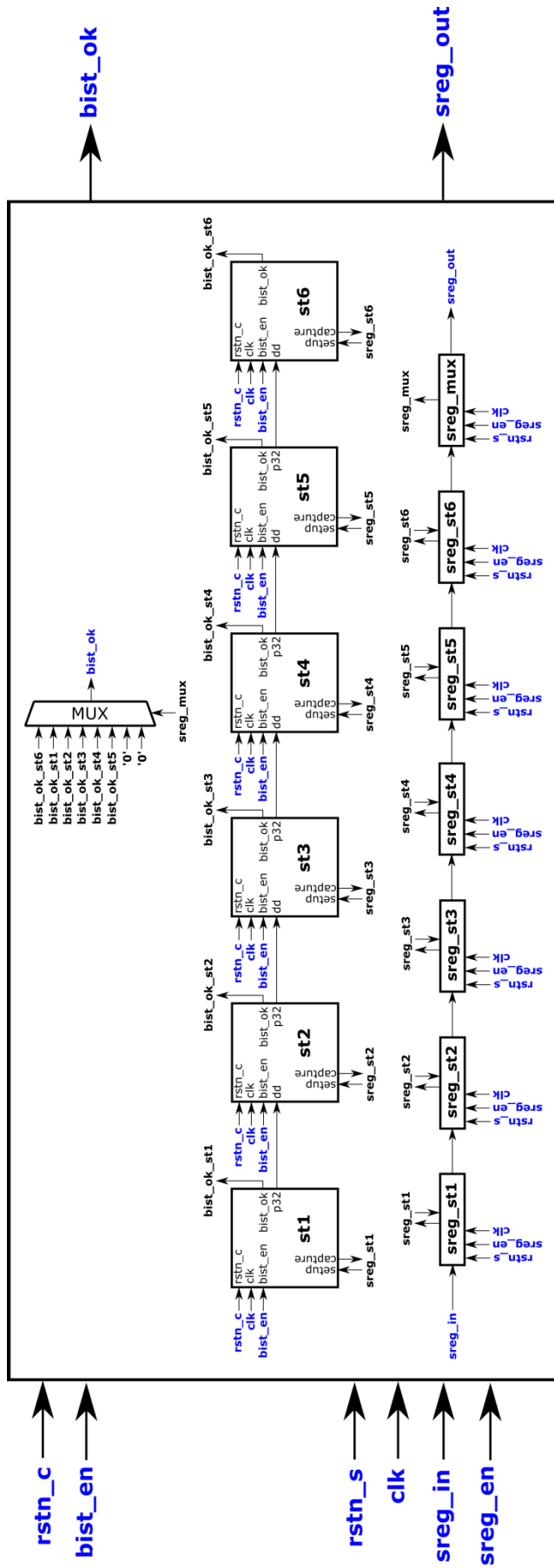


Figure 6.6 – The architecture of the original synchronous design

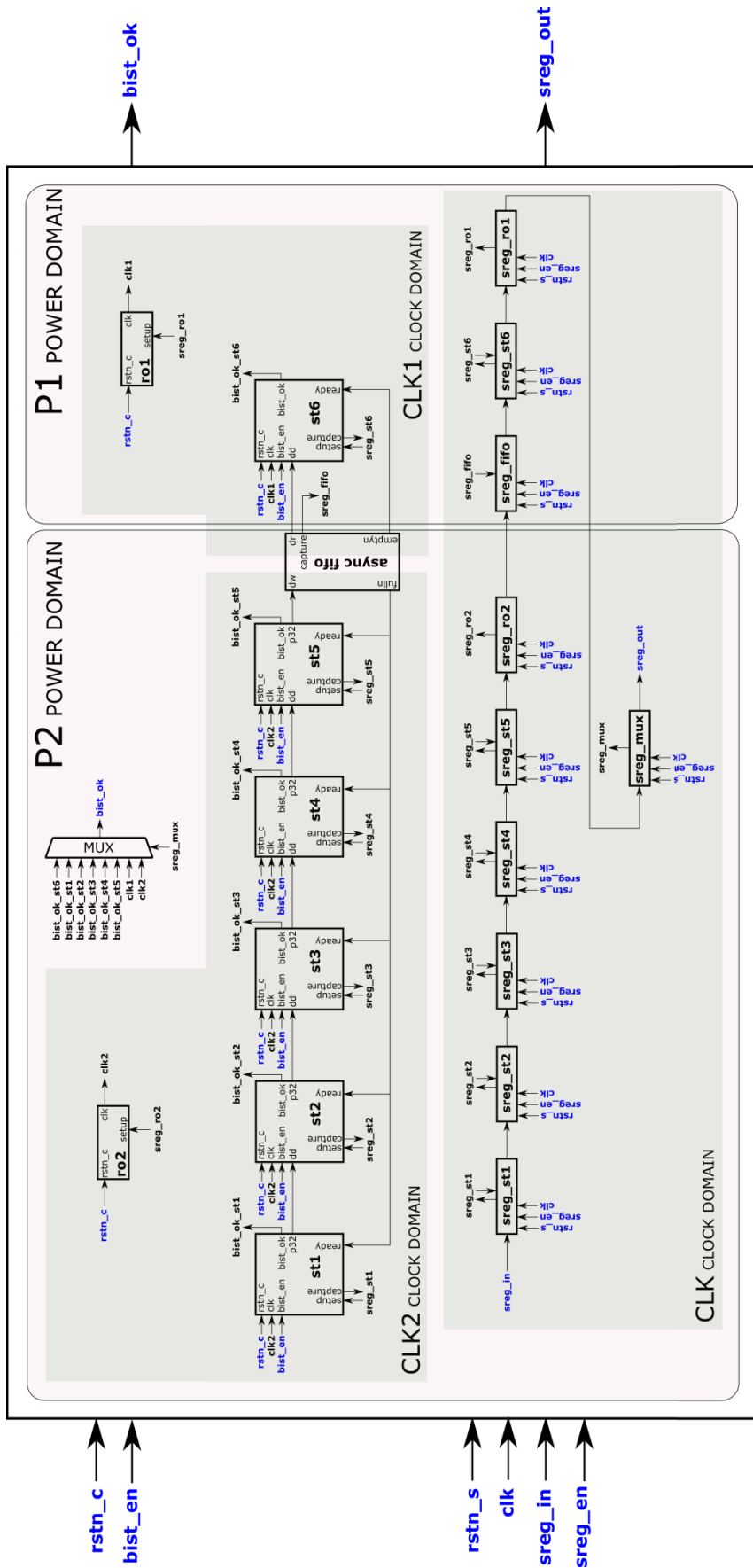


Figure 6.7 – The architecture of the resulting GALs design

6.1.3 Floorplan, layout and packaging of the test chip

The original synchronous core described in Section 6.1.1 and the GALS core with power domain separation described in Section 6.1.2 have been integrated on the same chip. In order to minimize the interaction of the cores, a wide NWell guard stripe is placed between them. The NWell guard stripe is biased via the pads NW_BIAS and NW_BIAS_G. The higher the biasing voltage applied to these pads, the better the isolation.

The layout of the chip is shown in the Figure 6.6, with the synchronous core on the left, the GALS core on the right, and NWell guard stripe in the middle. The total area of the chip is approximately 8 mm² (4 x 2 mm²).

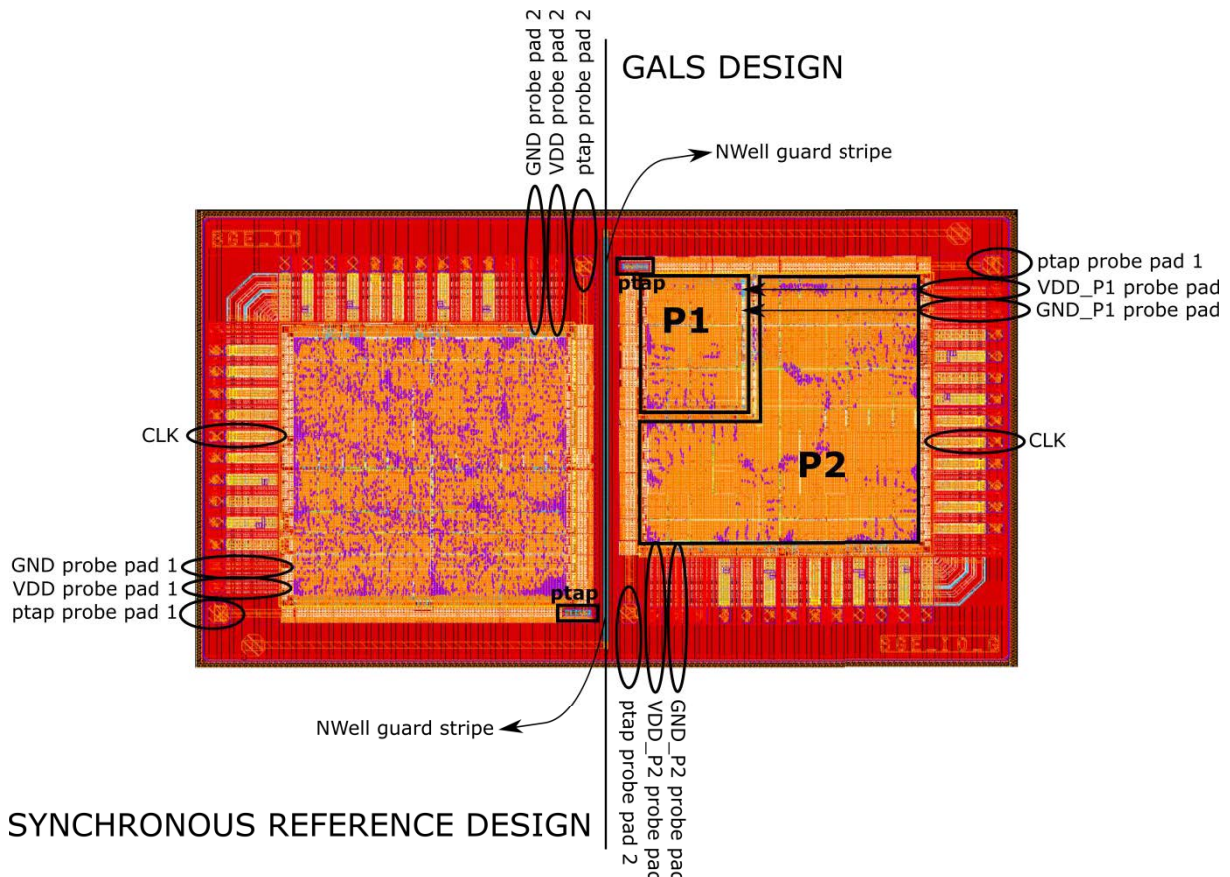


Figure 6.8 – SGE chip layout

Each half of the chip has a ptap substrate contact for probing the substrate voltage, which is connected to two sensing (probe) pads. Also, two pads for probing core supply voltage, and two pads for probing core ground voltage are provided per each core. In the GALS core, one of the supply voltage probe pads is connected to the supply ring of P1 power domain, while the other is connected to the supply ring of P2 power domain; the same applies for ground probe pads. The ptap contact is acting as the victim of substrate noise coupling. Thus, the floorplanning of the GALS core is done so that the domain P1 is closer to the ptap and “shields” it from the larger P2 domain.

In the synchronous half of the chip, there’s one pad with high switching activity – it’s the CLK pad, through which the core is clocked. Since ring oscillators are used for clocking the GALS core, there aren’t any pads with high switching activity in the GALS half of the chip. If the noise from the

pads would significantly propagate to the victim, it might lead to an unfair comparison in favour of GALS design version. Thus, it's important to keep the pad ring as far as possible from the sensing ptap victim. In order to achieve that, the pads have been put to the core sides opposite of the corner in which the ptap is placed. This way, the noise which might propagate from the pads is "shielded" by the substrate biasing in the core. Note, however, that this floorplanning decision also has a drawback: the P1 power domain has to be biased with long power and ground lines. Consequently, the on-chip parasitic PDN impedance in power domain P1 is increased, which might negatively impact the effectiveness of the noise reduction methodology.

The chip has been packaged in a QFN64 package, with dimensions 9 x 9 mm², as shown in Fig. 6.9. In order for the methodology to be applicable, there must be no conductive backplate connected to the die. Thus, the die has been isolated on the backside during packaging.



Figure 6.9 – Samples of the SGE chip

6.2. Testing and measurement results

Since noise measurements make sense only on package level, only package level tests have been performed. In order to functionally test the SGE chip and to be able to measure the noise, an adapter board has been developed, as shown in Fig. 6.10. The board is mounted on the baseboard of the Advantest V93000 testing system, which provides signal stimuli and performs the functional testing. For noise measurements via the spectrum analyzer, SMA connectors are provided on the board. During the testing and measurements, the chip is set in a socket.

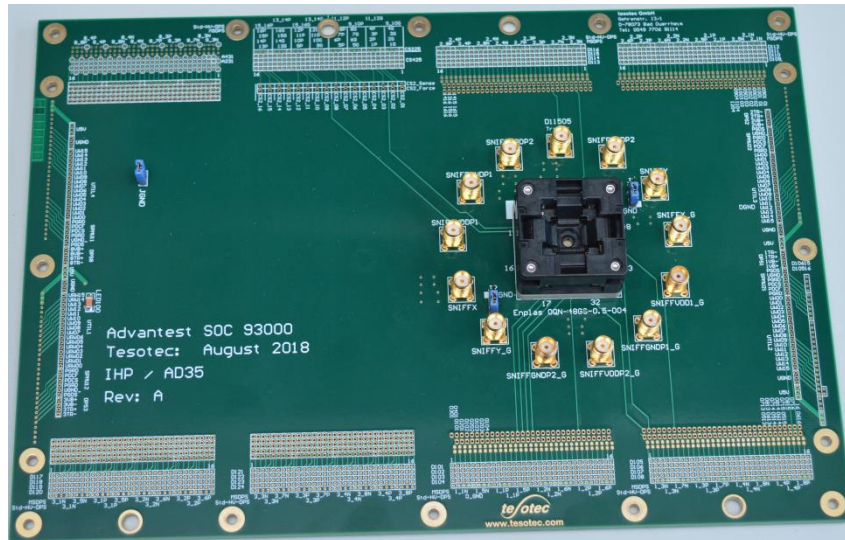


Figure 6.10 – The adapter board used for package level functional testing and noise measurements

During the measurements on one half of the chip, another half has been kept inactive, with supply and all the input signals connected to ground. The NWell isolation stripe has been biased to 3.3 V in order to further reduce the interaction between the two halves of the chip.

All of the noise measurements have been performed using the same resolution bandwidth, $RBW = 10$ kHz.

In Fig. 6.11, the measured substrate noise spectra in synchronous and GALS case are shown. As it can be seen, each of the harmonic peaks from the substrate noise spectra of the synchronous core is replaced by a pair of lower spectral peaks in the GALS core. That way, spectral peak attenuation has been achieved.

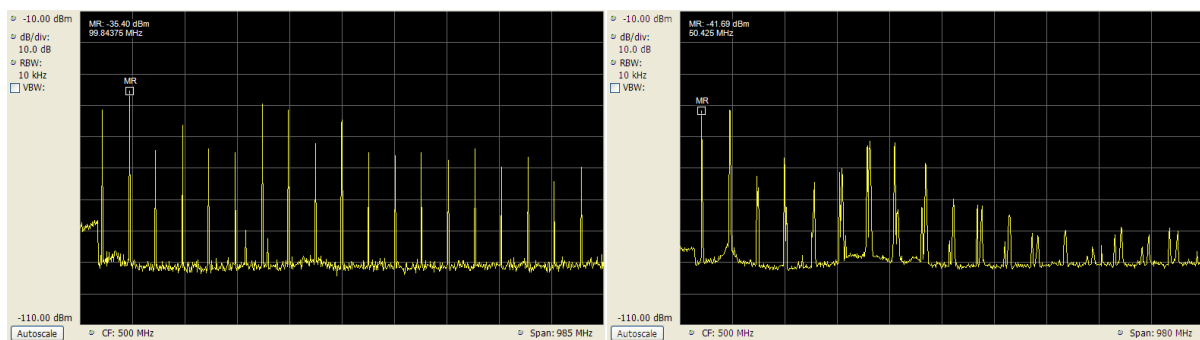


Figure 6.11 – Substrate noise spectra (left – synchronous core; right – GALS core)

The first two harmonics, as the highest spectral peaks, are shown more precisely in Fig. 6.12. and Fig. 6.13, for the synchronous core and the GALS core, respectively.

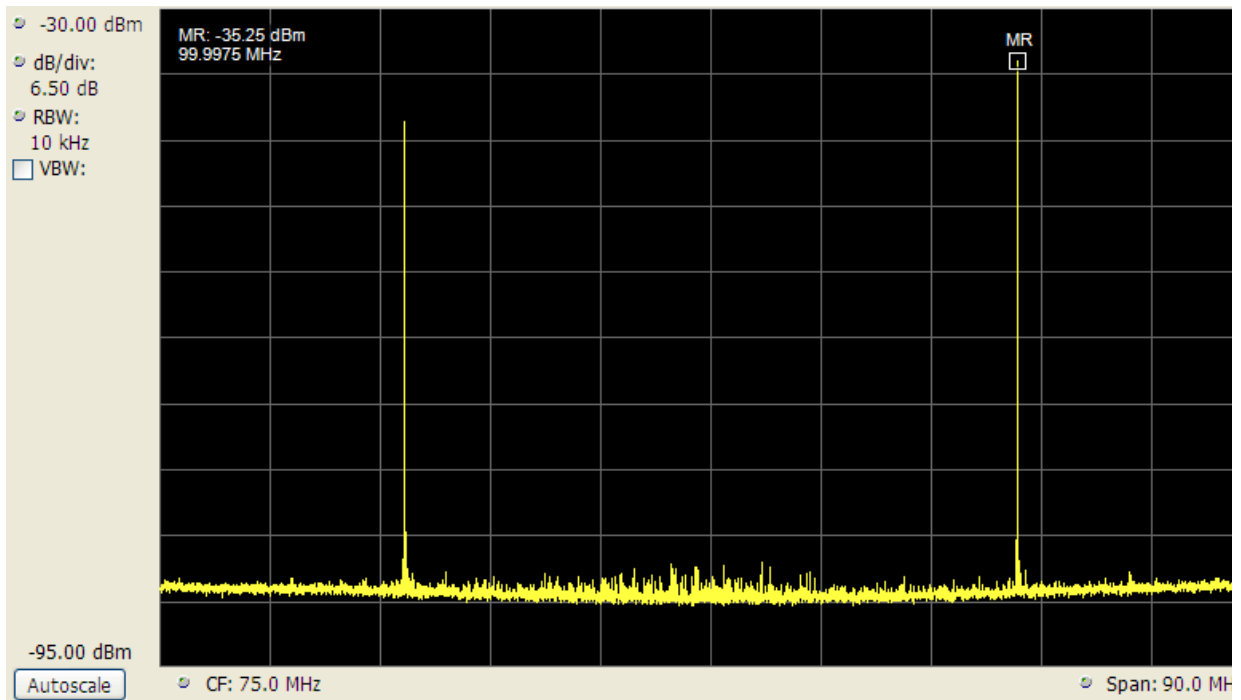


Figure 6.12 – Substrate noise spectra for the synchronous core – the first two harmonics

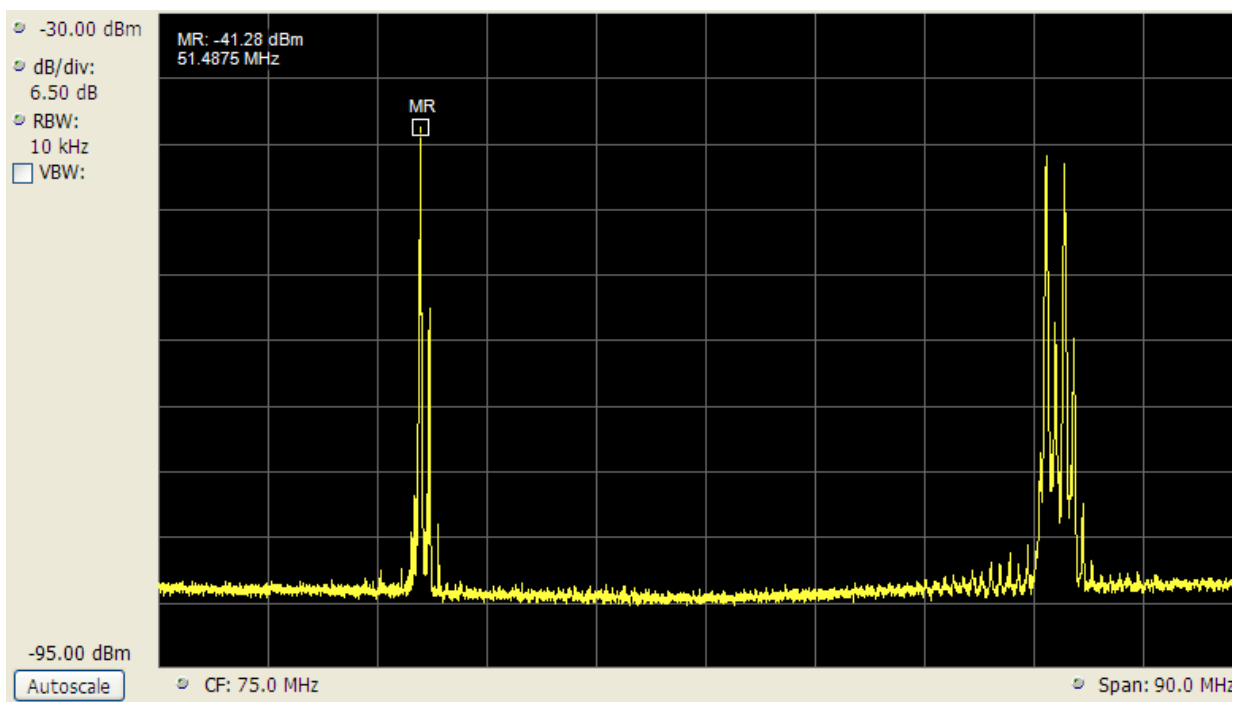


Figure 6.13 – Substrate noise spectra for the GALS core – the first two harmonics

The highest harmonic peak in the synchronous system is the second harmonic peak, at about 100 MHz and with an amplitude of -35.25 dBm. In the GALS system, this peak has been transformed into two approximately equal harmonic peaks, corresponding to the impacts of the two domains P1 and P2. In other words, the impacts of the two domains to the substrate noise are well-balanced, and consequently a substantial attenuation, somewhat larger than 6.5 dB, has been achieved at this peak.

On the other hand, the first harmonic peak in the synchronous system, at about 50 MHz, also turns into two harmonic peaks in the GALS system, but one of these two is much higher than the other one. In other words, the impacts of the two domains are not balanced in the GALS system. Thus, the first harmonic peak of the substrate noise remains approximately the same in the synchronous and in the GALS system, becoming the highest peak in the GALS system with an amplitude of -41.28 dBm.

Consequently, the difference between the largest substrate noise harmonic peak in the synchronous core and the largest substrate noise harmonic peak in the GALS core is 6.03 dB. Thus, it has been confirmed that by applying the methodology, substrate noise reduction can be achieved.

The ground bounce spectra in the synchronous and GALS cores have been measured as well. The comparison between the ground bounce spectrum in the synchronous system and the ground bounce spectrum in the partition P1 is shown in Fig. 6.14, while the comparison between the ground bounce spectrum in the synchronous system and the ground bounce spectrum in the partition P2 is shown in Fig. 6.15.

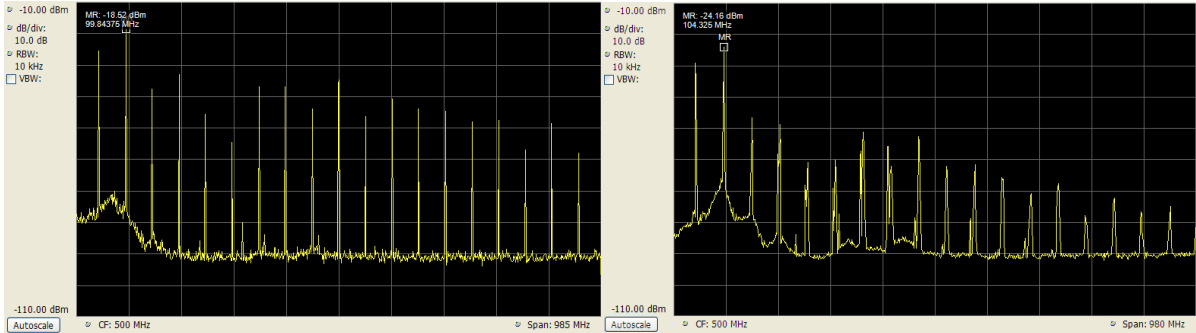


Figure 6.14 – The comparison of ground bounce spectrum in the synchronous system (left) and the P1 partition of the GALS system (right)

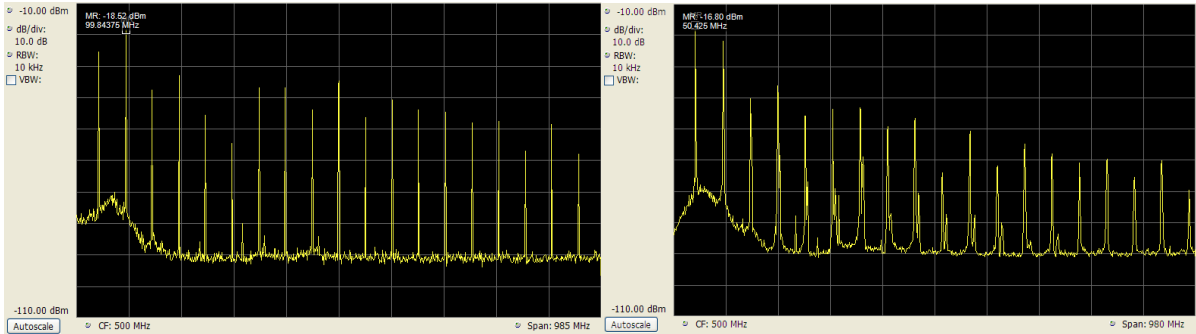


Figure 6.15 – The comparison of ground bounce spectrum in the synchronous system (left) and the P2 partition of the GALS system (right)

As it can be seen in Fig. 6.14, the ground bounce in the partition P1 is significantly lower than in the synchronous system. The highest ground bounce peak in both cases is the second one, with an amplitude of -18.52 dBm in the synchronous system, and -24.16 dBm in the GALS system.

On the other hand, as shown in Fig. 6.15, the ground bounce in the partition P2 increases compared to the initial system. Here, the first harmonic peak becomes the highest one, with an

amplitude of -16.80 dBm , which is 1.72 dBm higher than the highest (i.e. the second) ground bounce peak in the synchronous systems. However, this increase is not large enough to impact the operation of the logic gates in the partition P2, and thus it is acceptable.

The first harmonic peak in the domain P2 increased for more than 8 dB compared to the same peak in the synchronous domain. This large increase is an explanation for the absence of the spectral peak attenuation of the first harmonic of the substrate noise. Due to this huge increase, the impact of the domain P2 at this harmonic remains almost as strong as the impact of the original synchronous system, despite the higher substrate resistivity on the path to the victim.

Besides the cost in increasing the ground bounce of the P2 partition, the power and area overheads should be taken into account as well. These are presented in Table 6-II.

Table 6-II – Power and area overheads of the GALS design compared to the original synchronous design

Parameter		Synchronous	GALS	Overhead
Core area (post synthesis estimation)		0.997 mm^2	1.175 mm^2	17.9%
Estimated average power	From dynamic vector-based estimation on the post layout netlist in Synopsys PrimeTime	27.1 mW	29.7 mW	9.6%
	From dynamic vector-based estimation on the post layout netlist in Cadence Voltus	23.67 mW	25.28 mW	6.8%
Measured operating current	$f_{RO1} = 52.8 \text{ MHz} = 1.056 f_{sync}$ $f_{RO2} = 51.65 \text{ MHz} = 1.033 f_{sync}$	22.048 mA	25.524 mA	15.8%
	$f_{RO1} = 49.4 \text{ MHz} = 0.988 f_{sync}$ $f_{RO2} = 48.72 \text{ MHz} = 0.974 f_{sync}$	21.484 mA	23.132 mA	7.7%

As it can be seen in Table 6-II, the galsification has created a quite large area overhead of 17.9% . The reason is the GALS interface with the asynchronous FIFO and ready signals for blocking the data flow in case of a full or empty FIFO. Adding the ready signals to the FFs means that some additional gates are added in the feedback loop for each of the flip-flops (FFs) of the original design. Since there are many FFs in the system, this makes the most of the area overhead.

A more detailed look into the area data from the post synthesis estimation confirms this observation. The area of the asynchronous FIFO itself is $0,0143 \text{ mm}^2$, the area of each of the two ring oscillators is 0.0026 mm^2 , while the shift registers for setting the ring oscillators and reading the FIFO output, as well as the capture registers at the FIFO output, together make additional 0.0136 mm^2 of area. Thus, the asynchronous FIFO, the ring oscillators and the related circuitry for configuration and testing together comprise only about 3.3% area overhead. Most of the remaining 14.6% area overhead is due to the additional gates needed for introducing the ready signals to the FFs.

If pausable clocking [160] was used instead of the asynchronous FIFO interface with ready signals, the additional gates around the FFs wouldn't be needed. In that case, the area overhead would mainly consist of the ring oscillators, the asynchronous FIFO and the related setup and test circuitry, and it would be much smaller.

Estimated average power overhead, on the other hand, is moderate. Depending on the software tool used for the dynamic vector-based estimation, the power overhead estimation was 6.8 % - 9.6 %.

Similarly to the area overheads, a more detailed look in the dynamic vector-based estimation on the post-layout netlist in Synopsys PrimeTime reveals that a large part of the power overheads could have been avoided if a more efficient GALS interface was used instead. The total power consumed by the asynchronous FIFO is 0.515 mW, the total power of the two ring oscillators is 0.532 mW, while the shift registers for setting the ring oscillators and reading the FIFO output, as well as the capture registers at the FIFO output, together consume very little power in normal operating mode – only 0.067 mW. In total, the asynchronous FIFO, the ring oscillators and the related circuitry for configuration and testing comprise together only about 4.2 % power overhead. The remaining 5.4 % power overhead is mainly due to the additional gates needed for introducing the ready signals to the FFs.

Once more, if a more efficient GALS interface such as pausable clocking [160] was used instead, a significantly smaller power overheads would be achieved. Note that the power overheads also contribute to the ground bounce, and consequently also to the substrate noise. Thus, using a more efficient GALS interface could further improve the noise reduction results as well.

Finally, the data for the measured operating current has been provided. It is known [202] that the switching power is linearly proportional to the clock frequency. On the other hand, the average power is linearly proportional to the average operating current. Thus, for the GALS system, the average operating current depends on the average frequency of the ring oscillators, used for clocking the two LSMs of the GALS system. The frequency of the ring oscillators is process dependent. In Table 6-II, two examples have been shown – a GALS system with ring oscillator frequencies 5.6 % and 3.3 % higher than the synchronous clock frequency, and a GALS system with ring oscillator frequencies 1.2 % and 2.6 % lower than the synchronous clock frequency. For a faster GALS system, operating current overhead is 15.8 %, while for the slower GALS system, the operating current overhead is 7.7 %.

For a system with constant supply voltage as this one, the average operating current is proportional to the average power, which makes the operating current overhead basically the same as power overhead. As already mentioned in the overview of the estimated power overhead above, using a more efficient GALS interface such as pausable clocking [160] would reduce this overhead.

6.3. Summary

In this chapter, a test chip *SGE*, demonstrating the power-and-area-based plesiochronous GALS partitioning methodology for substrate noise reduction, has been presented. The chip contains two cores – a reference synchronous design and a GALS design derived from the reference design.

The architecture of the original synchronous design has been presented, and the modules to be taken into account for partitioning have been identified. Since the modules have similar power to area ratios, a simplified version of the methodology could be applied – the module with the lowest

power has been assigned to the partition closer to the victim, i.e. to the point where the substrate noise is measured. After determining the GALS partitioning, a GALS interface has been added. An asynchronous FIFO with Johnson counter has been used. Also, two ring oscillators have been added as the clock generators for the LSMs of the new GALS system. The chip has been fabricated, packaged in a QFN64 package, and tested functionally on an Advantest V93000 testing system.

Finally, substrate noise spectra and ground bounce spectra have been measured by a spectrum analyzer. Substrate noise attenuation has been demonstrated. Each of the substrate noise harmonic peaks in the synchronous design has been transformed into two lower harmonics peaks in the GALS design. The highest substrate noise harmonic peak of the GALS design is 6.03 dB lower than the highest harmonic peak of the synchronous design.

The price to be paid is a slight increase in the ground bounce in the larger GALS domain P2. The highest ground bounce harmonic peak of the P2 domain of the GALS design is 1.72 dB higher than the highest ground bounce harmonic peak of the original synchronous design. This increase, however, is acceptable, since it doesn't impact the operation of the core.

Finally, the power and area overheads have been analyzed. A significant area overhead of 17.9 % has been estimated. The estimated power overhead is 6.8 % – 9.6 %, depending on the tool used for dynamic power analysis. The measured operating current overhead (proportional to the average power overhead) is 7.7 % – 15.8 %, and it depends strongly on the frequency of the ring oscillators in the GALS system. The GALS interface used in this design is not very efficient in terms of power and area, since it requires additional logic for stopping the data flow at each of the FFs. Using a more efficient GALS interface, such as pausable clocking [160], would have reduced these overheads significantly, and would have further improved the noise reduction achieved.

7. Conclusion

In this thesis, a new approach for substrate noise reduction in MSICs is explored. An initially synchronous digital design is converted into a GALS design, thus naturally spreading the switching activity and reducing the switching noise. GALS-based noise reduction methodologies had already been used for reducing other forms of simultaneous switching noise. However, to the best of the author's knowledge, no GALS-based methodology targeting specifically substrate noise reduction had been developed prior to this work.

As an auxiliary task of this work, a coarse, high abstraction level model for substrate noise propagation in lightly doped substrates has been developed. The focus has been on lightly-doped substrates, because this substrate type is mostly used for MSICs. The purpose of this model is to enable the theoretical analysis of the impact of system level configurations to the substrate noise. By using this model, two new GALS-based substrate noise reduction methodologies have been developed: harmonic-balanced plesiochronous GALS partitioning (HB-pGp) and harmonic-and-area-based plesiochronous GALS partitioning with power domain separation (HAB-PDS&G). While HB-pGp is a purely system-level methodology, and is applicable independent on the substrate type, HAB-PDS is partially also a physical-level methodology, because it makes use of the isolating property of lightly doped substrates (and is thus specific only for such kind of substrates).

A theoretical analysis has shown that HB-pGp can reach substrate noise attenuation of up to $20\log(M)$, where M is the number of LSMs of the resulting GALS system. On the other hand, the attenuation achievable by HAB-PDS depends on the distribution of switching current harmonics and area among the partitions, as well as from the substrate itself. For each of the two methodologies, a suitable partitioning procedure for a practical application has been developed; these partitioning procedures have been numerically evaluated in MATLAB. HB has further been embedded within the EMIAS CAD tool, where it has been evaluated on a real design example – a wireless sensor node. A special case of HAB-PDS&G for low frequencies has been applied for developing a test chip called SGE. The measurements on silicon have proved the applicability of the methodology.

Both methodologies presented in this thesis have similar limitations. The effectiveness of both methodologies is dependent on granularity of the system, which can limit the achievable attenuation. Also, they both rely on switching current profile estimation, and are thus dependent on the EDA software delivering such estimation. This, however, can be avoided in special case when attenuation lower order harmonics is targeted. Finally, both methodologies assume the applicability of equal ground bounce approximation, which means that their usage is limited to small ASICs where package parasitics are dominant to on-chip parasitics.

On the other hand, both HB-pGp and HAB-PDS&G have some significant advantages to the existing methodologies.

At lower frequencies, HB-pGp outperforms synchronous system-level methodologies (SCS and CM). The attenuation it provides is independent on the slack and the switching current profile of the initial system, which is an advantage compared to SCS. Unlike for CM, the upper harmonic order limit set by overlapping can be avoided, and the in-band spectral power attenuation is provided. An additional advantage compared to both SCS and CM is that HB-pGp has the possibility to specifically

target one or more deliberate frequency bands. Compared to PB-pGp, which had already been used for reducing switching current harmonic peaks at lower frequencies, HB-pGp is more difficult to implement, but it can target harmonics of higher order, and it can also specifically target harmonics within one or more frequency bands; so, it is more general and it covers more use cases. PB-pGp can actually be viewed as a special case of HB-pGp if the harmonics of lower order are targeted for reduction.

HAB-PDS&G doesn't require any sophisticated technology enhancements. Unlike SOI and triple-wells, it can also be used for higher frequencies. Moreover, even for its lower-frequency version, PAB-PDS&G, the frequency limitation is not physically hard, but only relative to the initial synchronous clock frequency. Also, unlike p+ guard rings, it doesn't require any additional biasing pads for optimal performance, which makes it suitable for application for MSICs in small packages, where the number of pins is limited. It can deliver a significantly larger attenuation at lower frequencies compared to synchronous system-level methodologies (SCS and CM). It provides a larger attenuation compared to HB-pGp and PB-pGp with the same number of LSMs ($M = 2$), but HB-pGp and PB-pGp outperform it when the partitioning into a larger number of LSMs is done. It is also important to notice that the two methodologies presented in this thesis are complementary, which means that they can be combined in order to further improve the substrate noise reduction.

The author hopes that this work will lead to further research, improvement and finally wider acceptance of GALS-based methodologies for substrate noise reduction.

List of References

- [1] Behzad Razavi, "Analog CMOS Integrated Circuits," 2nd ed., McGraw Hill Education, 2017
- [2] Domine Leenaerts and Peter de Vreede, "Influence of substrate noise on RF performance," *Proceedings of the 26th European Solid-State Circuits Conference*, Stockholm, Sweden, 2000, pp. 328-331.
- [3] Min Xu, David K. Su, Derek K. Shaeffer, Thomas H. Lee and Bruce A. Wooley, "Measuring and modeling the effects of substrate noise on the LNA for a CMOS GPS receiver," *IEEE Journal of Solid-State Circuits*, vol. 36, no. 3, pp. 473-485, March 2001.
- [4] Tallis Blalack and Bruce A. Wooley, "The effects of switching noise on an oversampling A/D converter," *Proceedings ISSCC '95 - International Solid-State Circuits Conference*, pp. 200-201, San Francisco, CA, USA, 1995.
- [5] David England, "Substrate noise in mixed signal circuits: two case studies [CMOS]," *Proceedings Eleventh Annual IEEE International ASIC Conference*, pp. 37-41, Rochester, NY, USA, 1998.
- [6] Athanasios Stefanou and Georges Gielen, "Analyzing the performance degradation of flash A/D converters due to substrate noise coupling," *18th European Conference on Circuit Theory and Design*, pp. 360-363, Seville, Spain, 2007.
- [7] Athanasios Stefanou and Georges Gielen, "Effect of Mismatch on Substrate Noise Coupling on Flash A/D Converters," *14th IEEE International Conference on Electronics, Circuits and Systems*, pp. 1228-1231, Marrakech, Morocco, 2007.
- [8] Athanasios Stefanou and Georges Gielen, "Analyzing and modeling the performance degradation of flash A/D converters due to substrate noise coupling," *Analog Integrated Circuits and Signal Processing*, vol. 65, pp. 185-195, 2010.
- [9] Athanasios Stefanou and Georges Gielen, "A Volterra Series Nonlinear Model of the Sampling Distortion in Flash ADCs Due to Substrate Noise Coupling," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 12, pp. 877-881, December 2011.
- [10] Charlotte Soens, Geert Van der Plas, Piet Wambacq and Stéphane Donnay, "Performance degradation of an LC-tank VCO by impact of digital switching noise," *Proceedings of the 30th European Solid-State Circuits Conference*, pp. 119-122, Leuven, Belgium, 2004.
- [11] Charlotte Soens, Geert Van der Plas, Piet Wambacq, Stéphane Donnay and Maarten Kuijk, "Performance degradation of LC-tank VCOs by impact of digital switching noise in lightly doped substrates," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 7, pp. 1472-1481, July 2005.
- [12] Huailin Liao, Subhash C. Rustagi, Jinglin Shi and Yong Zhong Xiong, "Characterization and modeling of the substrate noise and its impact on the phase noise of VCO," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, 2003*, pp. 247-250, Philadelphia, PA, USA, 2003.
- [13] Sebastian Magierowski, Kris Iniewski and Chris Siu, "Substrate noise coupling effect characterization for RF CMOS LC VCOs," *The 3rd International IEEE-NEWCAS Conference, 2005.*, pp. 199-202, Quebec, QC, Canada, 2005.
- [14] Cristian Andrei, Olivier Valorge, Francis Calmon, Jacques Verdier and Christian Gontrand, "Impact of substrate perturbation on a 5 GHz VCO spectrum," *Proceedings of the 16th International Conference on Microelectronics (ICM)*, pp. 684-687, Tunis, Tunisia, 2004.
- [15] Nisha Checka, David D. Wentzloff, Anantha Chandrakasan and Rafael Reif, "The effect of substrate noise on VCO performance," *2005 IEEE Radio Frequency integrated Circuits (RFIC) Symposium - Digest of Papers*, pp. 523-526, Long Beach, CA, USA, 2005.
- [16] Xin Fan, Oliver Schrape, Miroslav Marinkovic, Peter Dähnert, Miloš Krstić, and Eckhard Grass, "GALS design for spectral peak attenuation of switching current," *Proceedings of the 19th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 83-90, Santa Monica, CA, USA, 2013.
- [17] Xin Fan, Miloš Krstić, Christoph Wolf, and Eckhard Grass, "GALS design for on-chip ground bounce suppression," *Proceedings of the 17th IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 43-52, Ithaca, NY, USA, 2011.
- [18] Xavier Aragonès, José Luis González, and Antonio Rubio, "Analysis and Solutions for Switching Noise Coupling in Mixed-Signal ICs," *Kluwer Academic Publishers*, 1999.
- [19] Frank Herzel and Detmar Kissinger, "Design and layout strategies for integrated frequency synthesizers with high spectral purity," *International Journal of Microwave and Wireless Technologies*, vol. 9(9), pp. 1791-1797, November 2017.
- [20] S. R. Vemuru, "Effects of simultaneous switching noise on the tapered buffer design," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 5, no. 3, pp. 290-300, September 1997.
- [21] Martin Saint-Laurent and Madhavan Swaminathan, "Impact of power-supply noise on timing in high-frequency microprocessors," *IEEE Transactions on Advanced Packaging*, vol. 27, no. 1, pp. 135-144, February 2004.
- [22] Emre Salman, Renatas Jakushokas, Eby G. Friedman, Radu M. Secareanu, and Olin L. Hartin, "Methodology for efficient substrate noise analysis in large-scale mixed-signal circuits," *IEEE Transactions on Very-Large Scale Integration (VLSI) Systems*, vol. 17, no. 10, pp. 1405-1418, October 2009.
- [23] Stéphane Donnay and Georges Gielen, "Substrate Noise Coupling in Mixed-Signal ASICs," *Kluwer Academic Publishers*, 2004.

- [24] Rajendran Panda, Savithri Sundareswaran, David Blaauw, "On the interaction of power distribution network with substrate," *Proceedings of the 2001 International Symposium on Low Power Electronics and Design (ISLPED'01)*, pp. 388-393, Huntington Beach, CA, USA, 2001.
- [25] <https://www.ihp-ffo.de/en/services/mpw-prototyping/sigec-bicmos-technologies.html>
- [26] Anthony Stamper, Renata Camillo-Castillo, Hanyi Ding, James Dunn, Mark Jaffe, Vibhor Jain, Alvin Joseph, Ian McCallum-Cook, Kim Newton, Shyam Parthasarathy, Robert Rassel, Nicholas Schmidt, Srikanth Srihari, Randy Wolf, and Michael Zierak, "High-resistivity SiGe BiCMOS technology development," *2014 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, pp. 25-28, Coronado, CA, USA, 2014.
- [27] Le Zhang, Xiao-Peng Yu and Er-Ping Li, "The investigation of the substrates noise suppression using guard rings in CMOS technology," *2016 IEEE International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, pp. 505-507, Beijing, China, 2016.
- [28] Le Zhang, Er-Ping Li, Xiao-Peng Yu and Ran Hao, "Modeling and Optimization of Substrate Electromagnetic Coupling and Isolation in Modern Lightly Doped CMOS Substrate," in *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 2, pp. 662-669, April 2017.
- [29] Mustafa Badaroglu, Piet Wambacq, Geert Van der Plas, Stéphane Donnay, Georges G.E. Gielen, and Hugo J. De Man, "Digital ground bounce reduction by supply current shaping and clock frequency modulation," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 1, pp. 65-76, January 2005.
- [30] Miguel A. Méndez, Diego Mateo, Antonio Rubio, and José Luis González, "Analytical and experimental verification of substrate noise spectrum for mixed-signal designs," *IEEE Transactions on Circuits and Systems*, vol. 53, no. 8, pp. 1803-1815, August 2006.
- [31] Giorgio Boselli, Gabriella Trucco, and Valentino Liberali, "Properties of digital switching currents in fully CMOS combinatorial logic," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 12, pp. 1625-1638, December 2010.
- [32] Howard H. Chen and J. Scott Neely, "Interconnect and circuit modeling techniques for full-chip power supply noise analysis," *IEEE Transactions on Components, Packaging, and Manufacturing Technology - Part B*, vol. 21, no. 3, pp. 209-215, August 1998.
- [33] Jun Chen and Lei He, "Efficient in-package decoupling capacitor optimization for I/O power integrity," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 4, pp. 734-738, April, 2007.
- [34] Matthias Eireiner, Doris Schmitt-Landsiedel, Paul Wallner, and Andreas Schöne, "Adaptive circuit block model for power-supply noise analysis of low-power system-on-chip," *Proceedings of 2009 International Symposium on System-on-Chip (SOC)*, pp. 013-018, Tampere, Finland, 2009.
- [35] Li-Rong Zheng and Hannu Tenhunen, "Fast modeling of core switching noise on distributed LRC power grid in ULSI circuits," *IEEE Transactions on Advanced Packaging*, vol. 24, no. 3, pp. 245-254, August 2001.
- [36] Divya Akella Kamakshi, Matthew Fojtik, Brucec Khailany, Sudhir Kudva, Yaping Zhou, Benton H. Calhoun, "Modeling and analysis of power supply noise tolerance with fine-grained GALS adaptive circuits," *Proceedings of the 22nd IEEE International Symposium on Asynchronous Circuits and Systems (ASYNC)*, pp. 75-82, Porto Alegre, Brazil, 2016.
- [37] Balsha R. Staniscic, Nishath K. Verghese, Rob A. Rutenbar, Richard Carley, and David J. Allstot, "Addressing substrate coupling in mixed-mode IC's: simulation and power distribution synthesis," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 3, pp. 226-238, March 1994.
- [38] Renatas Jakushokas, Emre Salman, Eby G. Friedman, Radu M. Secareanu, Olin L. Hartin, and Cynthia L. Recker, "Compact substrate models for efficient noise coupling and signal isolation analysis," *Proceedings of 2010. IEEE Symposium on Circuits and Systems (ISCAS)*, pp. 2346-2349, Paris, France, 2010.
- [39] Renatas Jakushokas, "Physical Resource Allocation for On-Chip Power Delivery Systems," *PhD Thesis*, Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York, USA, 2011.
- [40] Ming Shen, Jan Hvolgaard Mikkelsen, Ke Zhang, Ole Kiel Jensen, Tong Tian and Torben Larsen, "Modeling and Design Guidelines for P+ Guard Rings in Lightly Doped CMOS Substrates," *IEEE Transactions on Electron Devices*, vol. 60, no. 9, pp. 2854-2861, September 2013.
- [41] Ranjit Ghapurey and Robert G. Meyer, "Modeling and analysis of substrate coupling in integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 344-353, March 1996.
- [42] Nishath K. Verghese, David J. Allstot, and Mark A. Wolfe, "Verification techniques for substrate coupling and their application to mixed-signal IC design," *IEEE Journal of Solid-State Circuits*, vol. 31, no. 3, pp. 354-365, March 1996.
- [43] Thomas A. Johnson, Ronald W. Knepper, Victor Marcello, and Wen Wang, "Chip substrate resistance modeling technique for integrated circuit design," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 3, issue 2, pp. 126-134, April 1984.
- [44] Sujoy Mitra, Rob A. Rutenbar, L. Richard Carley, and David J. Allstot, "Substrate-aware mixed-signal macrocell placement in WRIGHT," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 3, pp. 269-278, March 1995.
- [45] Mustafa Badaroglu, Marc van Heijningen, Vincent Gravot, John Compier, Stéphane Donnay, Georges G.E. Gielen, and Hugo J. De Man, "Methodology and experimental verification for substrate noise reduction in CMOS mixed-signal ICs with synchronous digital circuits," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1383-1395, November 2002.
- [46] Mustafa Badaroglu, Stéphane Donnay, Hugo J. De Man, Yann A. Zinzus, Georges G. E. Gielen, Willy Sansen, Tony Fonden, and Svante Signell, "Modeling and experimental verification of substrate noise

- generation in a 220-k-gates WLAN system-on-chip with multiple supplies," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 7, pp. 1250-1260, July 2003.
- [47] Marc van Heijningen, John Compier, Piet Wambacq, Stéphane Donnay, and Ivo Bolsens, "Analysis and experimental verification of digital substrate noise generation for epi-type substrates," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 7, pp. 1002-1008, July 2000.
- [48] Mustafa Badaroglu, Geert Van der Plas, Piet Wambacq, Stéphane Donnay, Georges G. E. Gielen, and Hugo J. De Man, "SWAN: high-level simulation methodology for digital substrate noise generation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 14, no. 1, pp. 23-33, January 2006.
- [49] Sujoy Mitra, R.A. Rutenbar, L.R. Carley and D.J. Allstot, "A methodology for rapid estimation of substrate-coupled switching noise," *Proceedings of the IEEE 1995 Custom Integrated Circuits Conference*, pp. 129-132, Santa Clara, CA, USA, 1995.
- [50] Paolo Miliozzi, Luca Carloni, Edoardo Charbon, and Alberto Sangiovanni-Vincentelli, "SubWave: a methodology for modeling digital substrate noise injection in mixed-signal ICs," *Proceedings of the IEEE 1996 Custom Integrated Circuits Conference*, pp. 385-388, San Diego, CA, USA, 1996.
- [51] Luis Elvira, Ferran Martorell, Xavier Aragonés, and José Luis González, "A macromodelling methodology for efficient high-level simulation of substrate noise generation," *Proceedings of the Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pp. 1362-1363, vol. 2, Paris, France, 2004.
- [52] Nisha Checka, Anantha Chandrakasan, and Rafael Reif, "Substrate noise analysis and verification for the efficient noise prediction of a digital PLL," *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference*, pp. 473-476, San Jose, CA, USA, 2005.
- [53] Ali Afzali-Kusha, Makoto Nagata, Nishath K. Verghese and David J. Allstot, "Substrate Noise Coupling in SoC Design: Modeling, Avoidance, and Validation," *Proceedings of the IEEE*, vol. 94, no. 12, pp. 2109-2138, December 2006.
- [54] Keith A. Jenkins, "Substrate coupling noise issues in silicon technology," *Digest of Papers, 2004 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 91-94, Atlanta, GA, USA, 2004.
- [55] Xavier Aragonés and Antonio Rubio, "Experimental comparison of substrate noise coupling using different wafer types," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 10, pp. 1405-1409, October 1999.
- [56] Marc Molina, Xaier Aragonés and Jose Luis Gonzalez, "Experimental analysis of substrate isolation techniques for RF-SOC integration," *2009 IEEE International SOC Conference (SOCC)*, pp. 199-202, Belfast, UK, 2009.
- [57] William H. Kao and Xiaopeng Dong, "Digital Block Modeling and Substrate Noise Aware Floorplanning for Mixed Signal SOCs," *2007 IEEE International Symposium on Circuits and Systems*, pp. 1935-1938, New Orleans, LA, USA, 2007.
- [58] Hai Lan and Robert Dutton, "Synthesized compact models (SCM) of substrate noise coupling analysis and synthesis in mixed-signal ICs," *Proceedings Design, Automation and Test in Europe Conference and Exhibition*, Vol.2, pp. 836-841, Paris, France, 2004.
- [59] Marcin Jeske, Grzegorz Blakiewicz, Malgorzata Chrzanowska-Jeske and Benyi Wang, "Substrate noise-aware floorplanning for mixed-signal SOCs," *2004 IEEE International Symposium on Circuits and Systems*, pp. II-445-448, Vancouver, BC, Canada, 2004.
- [60] Grzegorz Blakiewicz, Marcin Jeske and Malgorzata Chrzanowska-Jeske, "Substrate noise optimization in early floorplanning for mixed signal SOCs," *IEEE International SOC Conference, 2004. Proceedings*, pp. 301-304, Santa Clara, CA, USA, 2004.
- [61] Grzegorz Blakiewicz, Marcin Jeske, Malgorzata Chrzanowska-Jeske and J. S. Zhang, "Substrate noise modeling in early floorplanning of MS-SOCs," *Proceedings of the ASP-DAC 2005. Asia and South Pacific Design Automation Conference*, Vol. 2, pp. 819-823, Shanghai, China, 2005.
- [62] Chung-Hsin Lin and Hung-Ming Chen, "On minimizing various sources of noise and meeting symmetry constraint in mixed-signal SoC floorplan design," *2009 1st Asia Symposium on Quality Electronic Design*, pp. 96-102, Kuala Lumpur, Malaysia, 2009.
- [63] Minsik Cho, Hongjoong Shin and David Z. Pan, "Fast substrate noise-aware floorplanning with preference directed graph for mixed-signal SOCs," *Asia and South Pacific Conference on Design Automation*, pp. 765-770, Yokohama, Japan, 2006.
- [64] Minsik Cho and David Z. Pan, "Fast Substrate Noise Aware Floorplanning for Mixed Signal SOC Designs," in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16, no. 12, pp. 1713-1717, December 2008.
- [65] Jiayi Liu, Sheqin Dong and Xianlong Hong, "Fast substrate noise driven floorplanning for mixed-signal circuits considering symmetry constraints," *2008 9th International Conference on Solid-State and Integrated-Circuit Technology*, pp. 2337-2340, Beijing, China, 2008.
- [66] John P. Z. Lee, Frank Wang, Abhijit Phanse and Linda C. Smith, "Substrate cross talk noise characterization and prevention in 0.35 μm CMOS technology," *Proceedings of the IEEE 1999 Custom Integrated Circuits Conference*, pp. 479-482, San Diego, CA, USA, 1999.
- [67] Hwan-Mei Chen, Ming-Hwei Wu, B. C. Liau, Laurence Chang and Ching-Fu Wu, "The study of substrate noise and noise-rejection-efficiency of guard-ring in monolithic integrated circuits," *IEEE International Symposium on Electromagnetic Compatibility. Symposium Record*, vol.1, pp. 123-128, Washington, DC, USA, 2000.
- [68] K. Rajendran, J. B. Johnson, S. Furkay, M. Kumar and S. Fischer, "Measurement and 3D simulations of substrate noise isolation and resistance for mixed signal applications," *Digest of Papers, 2004 Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 99-102, Atlanta, GA, USA, 2004.

- [69] Céline Barbier-Petot, Serge Bardy, Christèle Biard and Phillipe Descamps, "Substrate isolation in 90nm RF-CMOS technology," *2005 European Microwave Conference*, Paris, France, 2005.
- [70] Cheon Soo Kim, Piljae Park, Joung-Woo Park, Nam Hwang and Hyu Kyu Yu, "Deep trench guard technology to suppress coupling between inductors in silicon RF ICs," *2001 IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp. 1873-1876, Phoenix, AZ, USA, 2001.
- [71] David Szmyd, Laurent Gambus and William Wilbanks, "Strategies and test structures for improving isolation between circuit blocks," *Proceedings of the 2002 International Conference on Microelectronic Test Structures, 2002. ICMTS 2002.*, Cork, Ireland, 2002, pp. 89-93, doi: 10.1109/ICMTS.2002.1193177.
- [72] K. H. To, P. Welch, S. Bharatan, H. Lehning, T. L. Huynh, R. Thoma, D. Monk, W. M. Huang and V. Ilderem, "Comprehensive study of substrate noise isolation for mixed-signal circuits," *International Electron Devices Meeting. Technical Digest*, pp. 22.7.1-22.7.4, Washington, DC, USA, 2001.
- [73] K. Wai Chew, J. Zhang, K. Shao, W. Boon Loh and S.-F. Chu, "Impact of Deep N-well Implantation on Substrate Noise Coupling and RF Transistor Performance for Systems-on-a-Chip Integration," *32nd European Solid-State Device Research Conference*, pp. 251-254, Firenze, Italy, 2002.
- [74] Chih-Yuan Lee, Tung-Sheng Chen and Chin-Hsing Kao, "Methods for noise isolation in RFCMOS ICs," *IEEE Electron Device Letters*, vol. 24, no. 7, pp. 478-480, July 2003.
- [75] Tung-Sheng Chen, Chin-Yuan Lee, Chin-Hsin Kao, Der-Sheng Deng, Chung-Hsun Wu, Guo-Wei Huang and Kun-Ming Chen, "A highly efficient noise suppression technique for Si-based RFIC," *IEEE Radio Frequency Integrated Circuits (RFIC) Symposium*, pp. 343-346, Philadelphia, PA, USA, 2003.
- [76] Tung-Sheng Chen, Chih-Yuan Lee and Chin-Hsing Kao, "An efficient noise isolation technique for SOC application," in *IEEE Transactions on Electron Devices*, vol. 51, no. 2, pp. 255-260, Feb. 2004.
- [77] Wen-Kuan Yeh, Shuo-Mao Chen and Yean-Kuen Fang, "Substrate noise-coupling characterization and efficient suppression in CMOS technology," in *IEEE Transactions on Electron Devices*, vol. 51, no. 5, pp. 817-819, May 2004.
- [78] G. Ali Rezvani and Jon Tao, "Substrate isolation in 0.18um CMOS technology," *Proceedings of the 2005 International Conference on Microelectronic Test Structures (ICMTS)*, pp. 131-136, Leuven, Belgium, 2005.
- [79] Song Ye and Jun Li, "The effects of a deep n-well junction on RF circuit performanc," *2010 International Workshop on Junction Technology Extended Abstracts*, pp. 1-4 Shanghai, China, 2010.
- [80] Henry Lin, James Kuo, Robert Sobot and Marek Syrzycki, "Investigation of Substrate Noise Isolation Solutions in Deep Submicron (DSM) CMOS Technology," *2007 Canadian Conference on Electrical and Computer Engineering*, pp. 1106-1109, Vancouver, BC, Canada, 2007.
- [81] Ping-Chun Yeh, Hwann-Kaeo Chiou, Chwan-Ying Lee, John Yeh, Denny Tang and John Chern, "An Experimental Study on High-Frequency Substrate Noise Isolation in BiCMOS Technology," in *IEEE Electron Device Letters*, vol. 29, no. 3, pp. 255-258, March 2008.
- [82] Yun-Chen Wu, Shaen S. H. Hsu, Kevin K. W. Tan and Yu-Shiang Su, "Substrate Noise Coupling Reduction in β LC β Voltage-Controlled Oscillators," in *IEEE Electron Device Letters*, vol. 30, no. 4, pp. 383-385, April 2009.
- [83] Vincenzo Ferragina, Nicola Ghittori, Guido Torelli, Giorgio Boselli, Gabriella Trucco and Valentino Liberali, "Analysis of Crosstalk Effects on Mixed-Signal CMOS ICs with Different Mounting Technologies," *2005 IEEE Instrumentation and Measurement Technology Conference Proceedings*, pp. 1979-1984, Ottawa, ON, Canada, 2005.
- [84] Daisuke Kosaka, Makoto Nagata, Yukio Hiraoka, Ikuro Imanishi, Masakatsu Maeda, Yoshitaka Murasaka and Atsushi Iwata, "Isolation strategy against substrate coupling in CMOS mixed-signal/RF circuits," *Digest of Technical Papers. 2005 Symposium on VLSI Circuits*, pp. 276-279, Kyoto, Japan, 2005.
- [85] Mohammad Hekmat, Shahriar Mirabbasi and Majid Hashemi, "On the behaviour of passive guard-rings in lightly doped substrates," *2006 IEEE International Symposium on Circuits and Systems*, pp. 1279-1282, Kos, Greece, 2006.
- [86] Yongho Oh, Seungyong Lee, Hyungscheol Shin and Jae-Sung Rieh, "Trench-type deep N-well dual guard ring for the suppression of substrate noise coupling," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, pp. 36-44, 2011.
- [87] Tallis Blalack, Youri Leclercq and C. Patrick Yue, "On-chip RF isolation techniques," *Proceedings of the Bipolar/BiCMOS Circuits and Technology Meeting*, pp. 205-211, Minneapolis, MN, USA, 2002.
- [88] David Redmond, Morgan Fitzgibbon, Alan Bannon, Darren Hobbs, Chunhe Zhao, Kiyoshi Kase, Joseph Chan, Michael Priel, Kevin Traylor and Keith Tilley, "A GSM/GPRS mixed-signal baseband IC," *2002 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, San Francisco, CA, USA, 2002.
- [89] Stephane Bronckers, Gerd Vandersteen, Geert van der Plas and Yves Rolain, "On the P+ guard ring sizing strategy to shield against substrate noise," *2007 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium* pp. 753-756, Honolulu, HI, USA, 2007.
- [90] Piljae Park and C. Patrick Yue, "Modeling of triple-well isolation and the loading effects on circuits up to 50 GHz," *2008 IEEE Custom Integrated Circuits Conference*, pp. 217-220, San Jose, CA, USA, 2008.
- [91] Olivier Valorge, Francis Calmon, Cristian Andrei, Christian Gontrand and Pierre Dautriche, "Mixed-signal IC design guide to enhance substrate noise immunity in bulk silicon technology," *Analog Integrated Circuits and Signal Processing*, vol. 63, pp. 185-196, 2010.
- [92] Stephane Bronckers, Geert Van der Plas, Gerd Vandersteen and Yves Rolain, "Substrate Noise Coupling Mechanisms in Lightly Doped CMOS Transistors," in *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 6, pp. 1727-1733, June 2010.
- [93] L. Forbes, B. Ficq and S. Savage. "Resonant forward-biased guard-ring diodes for suppression of substrate noise in mixed-mode CMOS circuits," *Electronics Letters* 31, vol. 9, pp. 720-721, 1995.

- [94] L. Forbes, W. T. Lim and K. T. Yan, "Guard ring diodes for suppression of substrate noise and improved reliability in mixed-mode CMOS circuits," *Proceedings of 5th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, pp. 145-148, Singapore, 1995.
- [95] Le Zhang, Er-Ping Li and Xiao-Peng Yu, "Frequency-Response-Oriented Design and Optimization of N+ Diffusion Guard Ring in Lightly Doped CMOS Substrate," *IEEE Transactions on Electromagnetic Compatibility*, vol. 59, no. 2, pp. 481-487, April 2017.
- [96] Philipp Basedau and Qiuting Huang, "A post processing method for reducing substrate coupling in mixed-signal integrated circuits," *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 41-42, Kyoto, Japan, 1995.
- [97] Philipp Basedau, Qiuting Huang, Oliver Paul and Henry Baltes, "Isolating analog circuits from digital interferences," *Proceedings of 4th International Conference on Solid-State and IC Technology*, pp. 656-658, Beijing, China, 1995.
- [98] Han-Su Kim, Keith A. Jenkins and Ya-Hong Xie, "Effective crosstalk isolation through p⁺ Si substrates with semi-insulating porous Si," *IEEE Electron Device Letters*, vol. 23, no. 3, pp. 160-162, March 2002.
- [99] Han-Su Kim, Kyuchul Chong, Ya-Hong Xie, Marc Devincentis, Tatsuo Itoh, Andrew J. Becker and Keith A. Jenkins, "A porous Si based novel isolation technology for mixed-signal integrated circuits," *2002 Symposium on VLSI Technology, Digest of Technical Papers*, pp. 160-161, Honolulu, HI, USA, 2002.
- [100] Han-Su Kim, Kyuchul Chong, Ya-Hong Xie and Keith A. Jenkins, "The importance of distributed grounding in combination with porous Si trenches for the reduction of RF crosstalk through p⁺ Si substrate," *IEEE Electron Device Letters*, vol. 24, no. 10, pp. 640-642, October 2003.
- [101] Chen Li, Huailin Liao, Chuan Wang, Ru Huang and Yangyuan Wang, "Effective Crosstalk Isolation With Post-CMOS Selectively Grown Porous Silicon Technique for Radio Frequency System-on-Chip (SOC) Applications," *IEEE Electron Device Letters*, vol. 29, no. 9, pp. 994-997, September 2008.
- [102] C. P. Liao, K. C. Juang, T. H. Huang, D. S. Duh, T. T. Yang and M. N. Liu, "A new isolation technology for mixed-mode and general mixed-technology SOC chips," *2000 Semiconductor Manufacturing Technology Workshop*, pp. 124-132, Hsinchu, Taiwan, 2000.
- [103] Ning Li, Takeshi Inoue, Takuichi Hirano, Jian Pang, Rui Wu, Kenichi Okada, Hitoshi Sakane and Akira Matsuzawa, "Substrate noise isolation improvement by helium-3 ion irradiation technique in a triple-well CMOS process," *2015 45th European Solid State Device Research Conference (ESSDERC)*, pp. 254-257, Graz, Austria, 2015.
- [104] Joyce H. Wu, Jesús A. Del Alamo and Keith A. Jenkins, "A high aspect-ratio silicon substrate-via technology and applications: through-wafer interconnects for power and ground and Faraday cages for SOC isolation," *International Electron Devices Meeting 2000. Technical Digest (IEDM)*, pp. 477-480, 2000.
- [105] Joyce H. Wu, Jörg Scholvin, Jesús A. del Alamo and Keith A. Jenkins, "A Faraday cage isolation structure for substrate crosstalk suppression," *IEEE Microwave and Wireless Components Letters*, vol. 11, no. 10, pp. 410-412, October 2001.
- [106] Joyce H. Wu, Jörg Scholvin and Jesús A. del Alamo, "A through-wafer interconnect in silicon for RFICs," *IEEE Transactions on Electron Devices*, vol. 51, no. 11, pp. 1765-1771, November 2004.
- [107] Kyuchul Chong, Xi Zhang, King-Ning Tu, Daquan Huang, Mau-Chung Chang and Ya-Hong Xie, "Three-dimensional substrate impedance engineering based on p/p⁺ Si substrate for mixed-signal system-on-chip (SoC)," *IEEE Transactions on Electron Devices*, vol. 52, no. 11, pp. 2440-2446, November 2005.
- [108] Hasan Sharifi and Saeed Mohammadi, "Substrate Noise Rejection in a New Mixed-Signal Integration Technology," *2008 IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems*, pp. 147-150, 2008.
- [109] Hasan Sharifi and Saeed Mohammadi, "Self-Aligned Wafer-Level Integration Technology with an embedded faraday cage for substrate crosstalk suppression," *Microwave and Optical Technology Letters*, vol. 50, pp. 829-832, 2008.
- [110] Shinichiro Uemura, Yukio Hiraoka, Takayuki Kai and Shiro Dosho, "Isolation Techniques Against Substrate Noise Coupling Utilizing Through Silicon Via (TSV) Process for RF/Mixed-Signal SoCs," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 810-816, April 2012.
- [111] Takashi Hashimoto, Hidenori Satoh, Hiroaki Fujiwara and Mitsuru Arai, "A Study on Suppressing Crosstalk Through a Thick SOI Substrate and Deep Trench Isolation," *IEEE Journal of the Electron Devices Society*, vol. 1, no. 7, pp. 155-161, July 2013.
- [112] R. B. Merrill, W. M. Young and Kevin Brehmer, "Effect of substrate material on crosstalk in mixed analog/digital integrated circuit," *Proceedings of 1994 IEEE International Electron Devices Meeting*, pp. 433-436, 1994.
- [113] Irfan Rahim, Bor-Yuan Hwang and Juergen Foerstner, "Comparison of Soi Versus Bulk Silicon Substrate Crosstalk Properties for Mixed-Mode IC's," *1992 IEEE International SOI Conference*, pp. 170-171, 1992.
- [114] Phillipe Descamps, Céline Barbier-Petot, Cristèle Biard and Serge Bardy, "Performance comparison of coupling effect between silicon and SOI substrates in RF-CMOS technology," *Electronics Letters*, vol. 42, issue 20, pp. 1151-1152, September 2006.
- [115] Kuntal Joardar, "A simple approach to modeling cross-talk in integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 10, pp. 1212-1219, October 1994.
- [116] Kuntal Joardar, "Signal isolation in BiCMOS mixed mode integrated circuits," *Proceedings of Bipolar/Bicmos Circuits and Technology Meeting*, pp. 178-181, 1995.
- [117] Kuntal Joardar, "Substrate crosstalk in BiCMOS mixed mode integrated circuits," *Solid-State Electronics*, vol. 39, no. 4, pp. 511-516, 1996.

- [118] Kuntal Joardar, "Comparison of SOI and junction isolation for substrate crosstalk suppression in mixed mode integrated circuits," *Electronics Letters*, vol. 31, no. 15, pp. 1230-1231, 1995.
- [119] Alberto Viviani, Jean-Pierre Raskin, Denis Flandre, Jean-Pierre Colinge and Danielle Vanhoenacker, "Extended study of crosstalk in SOI-SIMOX substrates," *Proceedings of International Electron Devices Meeting*, pp. 713-716, 1995.
- [120] Jean-Pierre Raskin, Alberto Viviani, Denis Flandre and Jean-Pierre Colinge, "Substrate crosstalk reduction using SOI technology," *IEEE Transactions on Electron Devices*, vol. 44, no. 12, pp. 2252-2261, December 1997.
- [121] Junichi Kodate, Mitsuru Harada and Tsuneo Tsukahara, "Suppression of substrate crosstalk in mixed-signal complementary MOS circuits using high-resistivity SIMOX (Separation by IMplanted OXygen) wafers," *Japanese Journal of Applied Physics* vol. 39, no. 4S, pp. 2256-2260, 2000.
- [122] Khaled Ben Ali, Cesar Roda Neve, Ali Gharsallah and Jean-Pierre Raskin, "RF Performance of SOI CMOS Technology on Commercial 200-mm Enhanced Signal Integrity High Resistivity SOI Substrate," *IEEE Transactions on Electron Devices*, vol. 61, no. 3, pp. 722-728, March 2014.
- [123] John S. Hamel, Stefanos Stefanou, Michael Bain, B. Mervyn Armstrong and Harold S. Gamble, "Substrate crosstalk suppression capability of silicon-on-insulator substrates with buried ground planes (GPSOI)," in *IEEE Microwave and Guided Wave Letters*, vol. 10, no. 4, pp. 134-135, April 2000.
- [124] Stefanos Stefanou, John S. Hamel, Michael Bain, P. Baine, B. Mervyn Armstrong, Harold S. Gamble, Rick Manuel and Margaret Huang, "Physics and compact modeling of SOI substrates with buried ground plane (GPSOI) for substrate noise suppression," *2001 IEEE MTT-S International Microwave Symposium Digest*, vol.3, pp. 1877-1880, 2001.
- [125] Stefanos Stefanou, John S. Hamel, P. Baine, Michael Bain, B. Mervyn Armstrong, Harold S. Gamble, Michael Kraft, H. A. Kemhadjian and K. Osman, "Cross-talk suppression Faraday cage structure in silicon-on-insulator," *2002 IEEE International SOI Conference*, pp. 181-182, 2002.
- [126] Stefanos Stefanou, John S. Hamel, P. Baine, Michael Bain, B. Mervyn Armstrong, Harold S. Gamble, Michael Kraft and H. A. Kemhadjian, "Ultralow silicon substrate noise crosstalk using metal Faraday cages in an SOI technology," *IEEE Transactions on Electron Devices*, vol. 51, no. 3, pp. 486-491, March 2004.
- [127] Stefanos Stefanou, John S. Hamel, P. Baine, Michael Bain, B. Mervyn Armstrong, Harold S. Gamble, Michael Kraft and H. A. Kemhadjian, "Impact of buried oxide thickness and ground plane resistivity on substrate cross-talk in ground plane silicon-on-insulator (GPSOI) cross-talk suppression technology," *2004 IEEE International SOI Conference*, pp. 84-85, 2004.
- [128] Yasushi Hiraoka, Satoshi Matsumoto and Tatsuo Sakai, "New substrate-crosstalk reduction structure using SOI substrate [for one-chip transceiver IC]," *2001 IEEE International SOI Conference, Proceedings*, pp. 107-108, 2001.
- [129] Zhang Guoyan, Liao Huailin, Huang Ru, Zhang Xing and Wang Yangyuan, "The simulation analysis of cross-talk behavior in SOI mixed-mode integrated circuits," *2001 6th International Conference on Solid-State and Integrated Circuit Technology, Proceedings*, vol.2, pp. 916-919, 2001.
- [130] Wai Leng Cheong, Brian Owens, Hui En Pham, Christopher Hanken, Jim Le, Terri Fiez and Karthikeya Mayaram, "Comparison of supply noise and substrate noise reduction in SiGe BiCMOS and FDSOI processes," *2009 10th International Symposium on Quality Electronic Design*, pp. 112-115, 2009.
- [131] Fei Lu, Qi Chen, Chenkun Wang, Feilong Zhang, Cheng Li, Rui Ma, X. Shawn Wang and Albert Wang, "In-Die Through-BEOL Metal Wall for Noise Isolation in 180-nm FD-SOI CMOS," *IEEE Electron Device Letters*, vol. 38, no. 5, pp. 630-632, May 2017.
- [132] Sayfe Kiaei, San-Hwa Chee and Dave Allstot, "CMOS source-coupled logic for mixed-mode VLSI," *IEEE International Symposium on Circuits and Systems*, vol.2, pp. 1608-1611, 1990.
- [133] David J. Allstot, Sayfe Kiaei and Rajesh H. Zele, "Analog logic techniques steer around the noise," *IEEE Circuits and Devices Magazine*, vol. 9, no. 5, pp. 18-21, September 1993.
- [134] David J. Allstot, Guojin Liang and Howard C. Yang, "Current-mode logic techniques for CMOS mixed-mode ASICs," *Proceedings of the IEEE 1991 Custom Integrated Circuits Conference*, pp. 25.2/1-25.2/4, 1991.
- [135] David J. Allstot, "Low noise digital logic techniques," *Proceedings Fourth Annual IEEE International ASIC Conference and Exhibit*, pp. T13-1/1, 1991.
- [136] R. T. L. Saez, Maher Kayal, Michel Declercq and Márcio Cherem Schneider, "Digital circuit techniques for mixed analog/digital circuits applications," *Proceedings of Third International Conference on Electronics, Circuits, and Systems*, vol. 2, pp. 956-959, 1996.
- [137] Edgar F. M. Albuquerque and Manuel M. Silva, "Current-balanced logic for mixed-signal IC's," *1999 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 274-277, 1999.
- [138] Edgar Albuquerque and Manuel Silva, "A new low-noise logic family for mixed-signal integrated circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 46, no. 12, pp. 1498-1500, December 1999.
- [139] Edgar F. M. Albuquerque and Manuel M. Silva, "Evaluation of substrate noise in CMOS and low-noise logic cells," *The 2001 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 4, pp. 750-753, 2001.
- [140] Edgar F. M. Albuquerque and Manuel M. Silva, "An experimental comparison of substrate noise generated by CMOS and by low-noise digital circuits," *2004 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. II-481-484, 2004.
- [141] Edgar F. M. Albuquerque and Manuel M. Silva, "A comparison by simulation and by measurement of the substrate noise generated by CMOS, CSL, and CBL digital circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 4, pp. 734-741, April 2005.

- [142] Kirti Gupta, Neeta Pandey and Maneesha Gupta, "Dynamic positive feedback source-coupled logic (D-PFSCCL)," *International Journal of Electronics*, vol. 103, no. 10, pp 1626-1638, 2016.
- [143] Keiko Makie-Fukuda, Satoshi Maeda, Toshiro Tsukada and Tatsuji Matsuura, "Substrate noise reduction using active guard band filters in mixed-signal integrated circuits," *Digest of Technical Papers, Symposium on VLSI Circuits*, pp. 33-34, 1995.
- [144] Keiko Makie-Fukuda and Toshiro Tsukada, "On-chip active guard band filters to suppress substrate-coupling noise in analog and digital mixed-signal integrated circuits," *1999 Symposium on VLSI Circuits. Digest of Papers*, pp. 57-60, 1999.
- [145] Wolfgang Winkler and Frank Herzel, "Active substrate noise suppression in mixed-signal circuits using on-chip driven guard rings," *Proceedings of the IEEE 2000 Custom Integrated Circuits Conference*, pp. 357-360, 2000.
- [146] Mark Shane Peng and Hae-Seung Lee, "Study of substrate noise and techniques for minimization," *2003 Symposium on VLSI Circuits. Digest of Technical Papers*, pp. 197-200, 2003.
- [147] Mark Shane Peng and Hae-Seung Lee, "Study of substrate noise and techniques for minimization," *IEEE Journal of Solid-State Circuits*, vol. 39, no. 11, pp. 2080-2086, November 2004.
- [148] Toshiro Tsukada, Yasuyuki Hashimoto, Kohji Sakata, Hiroyuki Okada and Koichiro Ishibashi, "An on-chip active decoupling circuit to suppress crosstalk in deep sub-micron CMOS mixed-signal SoCs," *2004 IEEE International Solid-State Circuits Conference*, vol. 1, pp. 160-519, 2004.
- [149] Toshiro Tsukada, Yasuyuki Hashimoto, Kohji Sakata, Hiroyuki Okada and Koichiro Ishibashi, "An on-chip active decoupling circuit to suppress crosstalk in deep-submicron CMOS mixed-signal SoCs," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 67-79, January 2005.
- [150] Po-Jen Yang, Po-Tsang Huang and Wei Hwang, "Substrate noise suppression technique for power integrity of TSV 3D integration," *2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 3274-3277, 2012.
- [151] Ramesh Senthinathan, John L. Prince and S. Nimmagadda, "Effects of skewing CMOS output driver switching on the 'simultaneous' switching noise," *Proceedings of the Eleventh IEEE/CHMT International Electronics Manufacturing Technology Symposium*, pp. 342-345, 1991.
- [152] Patrick Vuillod, Luca Benini, Alessandro Bogliolo and Giovanni De Micheli, "Clock-skew optimization for peak current reduction," *Proceedings of 1996 International Symposium on Low Power Electronics and Design*, pp. 265-270, 1996.
- [153] Ashok Vittal, Hien Ha, Forrest Brewer and Malgorzata Marek-Sadowska, "Clock skew optimization for ground bounce control," *Proceedings of International Conference on Computer Aided Design*, pp. 395-399, 1996.
- [154] Erik Backenius, Mark Vesterbacka and Robert Häggglund, "A strategy for reducing clock noise in mixed-signal circuits," *The 2002 45th Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 1-29, 2002.
- [155] Erik Backenius and Mark Vesterbacka, "Reduction of Simultaneous Switching Noise in Digital Circuits," *2006 NORCHIP*, pp. 187-190, 2006.
- [156] Erik Backenius, Mark Vesterbacka and Robert Häggglund, "Effect of Simultaneous Switching Noise on an Analog Filter," *2006 13th IEEE International Conference on Electronics, Circuits and Systems*, pp. 898-901, 2006.
- [157] Mustafa Badaroglu, Kris Tiri, Stéphane Donnay, Piet Wambacq, Ingrid Verbauwhede, Georges Gielen and Hugo De Man, "Clock tree optimization in synchronous CMOS digital circuits for substrate noise reduction using folding of supply current transients," *Proceedings 2002 Design Automation Conference*, pp. 399-404, 2002.
- [158] Mustafa Badaroglu, Kris Tiri, Geert Van der Plas, Piet Wambacq, Ingrid Verbauwhede, Stéphane Donnay, Georges G. E. Gielen and Hugo J. De Man, "Clock-skew-optimization methodology for substrate-noise reduction with supply-current folding," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 6, pp. 1146-1154, June 2006.
- [159] Xin Fan, Mikkel B. Stegmann, Oliver Schrape, Steffen Zeidler, Isac G. Jensen, Jannich Thorsen, Tobias Bjerregaard and Miloš Krstić, "Frequency-Domain Optimization of Digital Switching Noise Based on Clock Scheduling," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, no. 7, pp. 982-993, July 2016.
- [160] Xin Fan, "GALS Design Methodology Based on Pausible Clocking," *PhD Thesis*, Faculty of Mathematics and Natural Sciences, Humboldt Universität zu Berlin, Berlin, Germany, 2013.
- [161] Fabio Campi, Davide Pandini, Tobias Bjerregaard and Mikkel Stensgaard, "A power shaping methodology for supply noise and EMI reduction," *Proceedings of the IEEE Design Automation Conference (DAC'10)*, User Track presentation, 2010.
- [162] Tetsuro Tanaka, Tamotsu Ninomiya and Koosuke Harada, "Random-switching control in DC-to-DC converters," *20th Annual IEEE Power Electronics Specialists Conference*, vol. 1, pp. 500-507, 1989.
- [163] Feng Lin and Dan Y. Chen, "Reduction of power supply EMI emission by switching frequency modulation," *Proceedings of IEEE Power Electronics Specialist Conference - PESC '93*, pp. 127-133, 1993.
- [164] Feng Lin and Dan Y. Chen, "Reduction of power supply EMI emission by switching frequency modulation," *IEEE Transactions on Power Electronics*, vol. 9, no. 1, pp. 132-137, January 1994.
- [165] Keith B. Hardin, John T. Fessler and Donald R. Bush, "Spread spectrum clock generation for the reduction of radiated emissions," *Proceedings of IEEE Symposium on Electromagnetic Compatibility*, pp. 227-231, 1994.
- [166] Gianluca Setti, Michele Balestra and Riccardo Rovatti, "Experimental verification of enhanced electromagnetic compatibility in chaotic FM clock signals," *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol.3, pp. 229-232, 2000.

- [167] Stefano Santi, Riccardo Rovatti and Gianluca Setti, "Advanced chaos-based frequency modulations for clock signals EMC tuning," *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS)*, pp. III-116–119, 2003.
- [168] Sergio Callegari, Riccardo Rovatti and Gianluca Setti, "Chaos-based FM signals: application and implementation issues," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 8, pp. 1141-1147, August 2003.
- [169] Sergio Callegari, Riccardo Rovatti and Gianluca Setti, "Spectral properties of chaos-based FM signals: theory and simulation results," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 1, pp. 3-15, January 2003.
- [170] Luca Antonio De Michele, Fabio Pareschi, Riccardo Rovatti and Gianluca Setti, "Chaos-based high-EMC spread-spectrum clock generator," *Proceedings of the 2005 European Conference on Circuit Theory and Design*, vol. 1, pp. 1/165-1/168, 2005.
- [171] Hsiang-Hui Chang, I-Hui Hua and Shen-luan Liu, "A spread-spectrum clock generator with triangular modulation," *IEEE Journal of Solid-State Circuits*, vol. 38, no. 4, pp. 673-676, April 2003.
- [172] Minyoung Song, Sunghoon Ahn, Inhwa Jung, Yongtae Kim and Chulwoo Kim, "Piecewise Linear Modulation Technique for Spread Spectrum Clock Generation," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 21, no. 7, pp. 1234-1245, July 2013.
- [173] Yasushi Matsumoto, Katsumi Fujii and Akira Sugiura, "An analytical method for determining the optimal modulating waveform for dithered clock generation," *IEEE Transactions on Electromagnetic Compatibility*, vol. 47, no. 3, pp. 577-584, August 2005.
- [174] Davide De Caro, "Optimal Discontinuous Frequency Modulation for Spread-Spectrum Clocking," *IEEE Transactions on Electromagnetic Compatibility*, vol. 55, no. 5, pp. 891-900, October 2013.
- [175] Mustafa Badaroglu, Piet Wambacq, Geert Van der Plas, Stéphane Donnay, Georges Gielen and Hugo De Man, "Digital ground bounce reduction by phase modulation of the clock," *Proceedings Design, Automation and Test in Europe Conference and Exhibition*, Vol. 1, pp. 88-93, 2004.
- [176] John R. Carson, "Notes on the Theory of Modulation," *Proceedings of the Institute of Radio Engineers*, vol. 10, no. 1, pp. 57-64, February 1922.
- [177] Yuxin Wang and Zeljko Ignjatovic, "On-Chip Substrate Noise Suppression Using Clock Randomization Methodology," *2007 IEEE International Symposium on Circuits and Systems*, pp. 2176-2179, 2007.
- [178] Selçuk Köse, Emre Salman, Zeljko Ignjatovic and Eby G. Friedman, "Pseudo-random clocking to enhance signal integrity," *2008 IEEE International SOC Conference*, pp. 47-50, 2008.
- [179] Jens Sparsø and Steve Furber, "Principles of asynchronous circuit design – a systems perspective," Kluwer Academic Publishers, 2002.
- [180] Nigel C. Paver, Paul Day, Craig Farnsworth, D. L. Jackson, W. A. Lien and Jianwei Liu, "A low-power, low noise, configurable self-timed DSP," *Proceedings Fourth International Symposium on Advanced Research in Asynchronous Circuits and Systems*, pp. 32-42, 1998.
- [181] Stephen B. Furber, James D. Garside, Peter Riocreux, Steven Temple, Paul Day, Jianwei Liu and Nigel C. Paver, "AMULET2e: an asynchronous embedded controller," *Proceedings of the IEEE*, vol. 87, no. 2, pp. 243-256, February 1999.
- [182] Hans van Gageldonk, Kees van Berkel, Ad Peeters, Daniel Baumann, Daniel Gloor and Gerhard Stegmann, "An asynchronous low-power 80C51 microcontroller," *Proceedings Fourth International Symposium on Advanced Research in Asynchronous Circuits and Systems*, pp. 96-107, 1998.
- [183] C. H. (Kees) Van Berkel, Mark B. Josephs and Steven M. Nowick, "Applications of asynchronous circuits," *Proceedings of the IEEE*, vol. 87, no. 2, pp. 223-233, February 1999.
- [184] Antonio Jose Acosta, Raúl Jiménez, Jorge Juan, Manuel J. Bellido, and Manuel Valencia. "Influence of clocking strategies on the design of low switching-noise digital and mixed-signal VLSI circuits." *International Workshop on Power and Timing Modeling, Optimization and Simulation*, pp. 316-326. Springer, 2000.
- [185] Raúl Jiménez, Antonio J. Acosta, Eduardo J. Peralías, and Adoración Rueda. "An application of self-timed circuits to the reduction of switching noise in analog-digital circuits." *International Workshop on Power and Timing Modeling, Optimization and Simulation*, pp. 295-305. Springer, Berlin, Heidelberg, 2000.
- [186] Dhanistha Panyasak, Gilles Sicard, and Marc Renaudin, "A current shaping methodology for low EMI asynchronous circuits," *3rd International Workshop on Electromagnetic Compatibility of Integrated Circuits*, no. ISBN 2-87649-043-9, pp-43-48, Toulouse, France, November 2002.
- [187] Dhanistha Panyasak, Gilles Sicard, and Marc Renaudin, "A current shaping methodology for lowering EM disturbances in asynchronous circuits," *Microelectronics journal*, vol. 35, no. 6, pp. 531-540, 2004.
- [188] Rui A. L. de Cristo, Ricardo P. Jasinski and Volnei A. Pedroni, "Analysis and Preliminary Measurements of Radiated Emissions in an Asynchronous Circuit versus its Synchronous Counterpart," *2010 International Conference on Reconfigurable Computing and FPGAs*, pp. 127-131, 2010.
- [189] Jim Le, Christopher Hanken, Martin Held, Mike Hagedorn, Kartikeya Mayaram and Terri S. Fiez, "Comparison and Impact of Substrate Noise Generated by Clocked and Clockless Digital Circuitry," *IEEE Custom Integrated Circuits Conference 2006*, pp. 105-108, 2006.
- [190] Jim Le, Christopher Hanken, Martin Held, Michael S. Hagedorn, Kartikeya Mayaram and Terri S. Fiez, "Experimental Characterization and Analysis of an Asynchronous Approach for Reduction of Substrate Noise in Digital Circuitry," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 20, no. 2, pp. 344-356, February 2012.

- [191] Daniel M. Chapiro, "Globally-asynchronous locally-synchronous systems," *PhD Thesis*, Stanford University, 1984.
- [192] Paul Teehan, Mark Greenstreet and Guy Lemieux, "A Survey and Taxonomy of GALS Design Styles," *IEEE Design & Test of Computers*, vol. 24, no. 5, pp. 418-428, September-October 2007.
- [193] Miloš Krstić, Eckhard Grass, Frank K. Gürkaynak and Pascal Vivet, "Globally Asynchronous, Locally Synchronous Circuits: Overview and Outlook," *IEEE Design & Test of Computers*, vol. 24, no. 5, pp. 430-441, September-October 2007.
- [194] Xin Fan, Miloš Krstić, Christoph Wolf and Eckhard Grass, "A GALS FFT processor with clock modulation for low-EMI applications," *ASAP 2010 - 21st IEEE International Conference on Application-specific Systems, Architectures and Processors*, pp. 273-278, 2010.
- [195] Miloš Krstić, Xin Fan, Eckhard Grass, Luca Benini, Mohammad Reza Kakoei, Christoph Heer, Birgit Sanders, Alessandro Strano and Davide Bertozzi, "Evaluation of GALS methods in scaled CMOS technology: moonraker chip experience," *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*, vol. 3, no. 4, pp. 1-18, 2012.
- [196] Miloš Krstić, Xin Fan, Eckhard Grass, Luca Benini, Mohammad Reza Kakoei, Christoph Heer, Birgit Sanders, Alessandro Strano, Gabriele Miorandi and Alberto Ghiribaldi, "Silicon Validation of GALS Methods and Architectures in a State-of-the-Art CMOS Process," *Advancing Embedded Systems and Real-Time Communications with Emerging Technologies*, pp. 420-447, IGI Global, 2014.
- [197] <https://www.mathworks.com/help/matlab/sparse-matrices.html>
- [198] <https://www.intel.de/content/dam/www/public/us/en/documents/packaging-databooks/packaging-chapter-04-databook.pdf>
- [199] <https://www.ti.com/lit/an/snoa405a/snoa405a.pdf>
- [200] Steven S. Skiena, "The Algorithm Design Manual," Springer, 2008.
- [201] Steffen Zeidler, Xin Fan, Oliver Schrape and Miloš Krstić, "An Early Stage Design Flow for Switching Noise Attenuation," *Journal of Circuits, Systems and Computers*, vol. 25, no. 03, pp. 1640022 (1-21), 2016.
- [202] Jan Rabaey, "Low Power Design Essentials," Springer, 2009.

Acknowledgements

The research work presented in this thesis was mostly realized in scope of the project GASEBO, which was funded by DFG (Deutsche Forschungsgemeinschaft). Thus, I would like to thank DFG for financing the project.

I would like to thank my reviewers for their valuable and insightful comments. Especially I would like to thank my supervisor, Prof. Rolf Kraemer, who was very patient with me despite all the postponements and who motivated me when it was needed to finally finish the thesis. Also, I would like to express my deep gratitude to Prof. Miloš Krstić, who continuously supported me since the beginning of the GASEBO project with plenty of valuable advice and suggestions considering scientific, technical and organizational aspects of my work.

Further, I would like to thank Xin Fan and Steffen Zeidler, the colleagues who had previously researched on related topics at IHP, for all the useful insights, advice and support I got from them, as well as Vladimir Petrović for his practical lessons in digital ASIC design flow during my first months at IHP. I would also like to thank Oliver Schrape for his immense support regarding the back-end design of my test chip, Alexey Balashov and Goran Panić for their help regarding DRC and LVS issues, and the members of the test team at IHP - Irina Matthaei for her help regarding packaging, Peter Dähnert for his help regarding the test board, and Carsten Schulze and Silvia Hinrich for their support regarding testing.

Finally, I would like to thank my family - my parents Jefto and Verica, my grandfather Đorđo and my brother Marko - for their constant love and support.