PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

SLICE-BASED ANALOG DESIGN AND ITS APPLICATION TO PARTICLE PHYSICS INSTRUMENTATION

PABLO WALKER GALDAMES

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Advisor:
ANGEL ABUSLEME HOFFMAN
JUAN PEDRO OCHOA-RICOUX

Santiago de Chile, June 2021

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

SLICE-BASED ANALOG DESIGN AND ITS APPLICATION TO PARTICLE PHYSICS INSTRUMENTATION

PABLO WALKER GALDAMES

Members of the Committee:
ANGEL ABUSLEME HOFFMAN
JUAN PEDRO OCHOA-RICOUX
MATÍAS NEGRETE PINCETIC
KRZYSZTOF HERMAN
JUAN CARLOS HERRERA MALDONADO

Thesis submitted to the Office of Research and Graduate Studies
in partial fulfillment of the requirements for the degree of
Master of Science in Engineering

Santiago de Chile, June 2021

ACKNOWLEDGEMENTS

First, I would like to thank professor Angel Abusleme, one of my co-advisors. Angel taught me most of what I know about microelectronics, and has guided me throughout the research process with his knowledge, enthusiasm and his endless patience. Without him, this thesis would not have been possible, and quite likely I would have never designed or been involved with integrated circuits.

Second, I would like to thank professor Pedro Ochoa, my other co-advisor. Pedro introduced me to the area of experimental physics, and provided me with opportunities to participate in the design of particle physics instrumentation, beyond the present thesis, and has believed in me and my abilities as an engineer. His support has been invaluable over the years.

Finally, I would like to thank my parents for their love, support, understanding and patience throughout this long process, for which I am eternally grateful.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS iv

LIST OF FIGURES ix

LIST OF TABLES xvi

ABSTRACT xvii

RESUMEN xviii

1. INTRODUCTION 1
   1.1 Analog integrated circuit design 1
      1.1.1 The design process 1
      1.1.2 About systematization and automation 2
   1.2 Topics covered by the Thesis 3
   1.3 Particle physics experiments 4
   1.4 Electronics for particle physics experiments 6
      1.4.1 Detector 6
      1.4.2 Amplifier 7
      1.4.3 Pulse Shaper 8
      1.4.4 Digitizer 8
   1.5 Noise in electronics 9
   1.6 The g_m/I_D design methodology 11
      1.6.1 The g_m/I_D ratio as a design variable 11
      1.6.2 Methodology 13
   1.7 Thesis content 15

2. SLICE-BASED DESIGN METHODOLOGY 16
   2.1 Introduction 16

2.2 The effects of connecting circuits in parallel 17
2.2.1 The general case 17
2.2.2 Single-pole amplifier 18
2.2.3 Noise analysis 19
2.3 The design methodology 20
2.3.1 The pre-design stage 20
2.3.2 The design stage 22
2.3.3 Adjustable performance 23
2.3.4 Potential issues 23

3. NOISE ANALYSIS IN CHARGE-SENSITIVE AMPLIFIERS 25
3.1 Introduction 25
3.2 Equivalent noise charge 25
3.2.1 Front-end circuit model 26
3.2.2 Single-electron output power 27
3.2.3 Output noise power 29
3.2.4 The ENC equation 31
3.3 Filter design 31
3.3.1 Noise coefficients 31
3.3.2 Normalized noise coefficients 32
3.4 Charge-sensitive amplifier design 36
3.4.1 The g_m/I_D methodology and noise analysis 36
3.4.2 Charge amplifier noise using the g_m/I_D methodology 37
3.4.3 Capacitance matching for constant g_m/I_D 39
3.4.4 Flicker noise and corner frequency 40
3.4.5 Minimum noise versus g_m/I_D and peaking time 43
3.4.6 White-noise dominance – practical design guidelines 45
3.4.7 Noise analysis in slice-based design 48

4. EFFECTS OF DEVICE MISMATCH ON SLICE-BASED DESIGN 50

4.1 Introduction . . . 50
4.2 Mismatch model . . . 51
4.3 Mismatch parameters . . . 52
4.4 Physical interpretation of the distance term in Pelgrom's mismatch model . 54
4.5 Device mismatch in cell-based design . . . 55
4.5.1 Parameter variation model . . . 55
4.5.2 Parameter scaling . . . 57
4.6 A simple example of noise scaling in the presence of device mismatch . . . 60
4.7 Monte Carlo simulations . . . 62
4.7.1 Methodology . . . 63
4.7.2 Results – Single device noise scaling . . . 64
4.7.3 Results – Charge-sensitive amplifier . . . 68
4.7.4 Mismatch and switches – The charge-sensitive amplifier of the Heisenberg IC . . . 76

5. SYSTEM-LEVEL DESIGN . . . 77
5.1 Introduction . . . 77
5.2 System block diagram . . . 77
5.2.1 The Heisenberg chip block diagram . . . 79
5.3 System specifications . . . 80

6. CIRCUIT-LEVEL DESIGN . . . 82
6.1 Introduction . . . 82
6.2 Charge-sensitive amplifier design . . . 82
6.2.1 Amplifier static error and nonlinearity . . . 82
6.2.2 Frequency response . . . 83
6.2.3 The folded-cascode amplifier . . . 85
6.2.4 Equivalent transconductance, output resistance and open-loop gain . . 86
6.2.5 Input-referred noise . . . 88
6.2.6 Design parameters . . . 89

6.2.7 Folded-cascode bias circuit . . . 89
6.2.8 Slice parallel connection . . . 91
6.2.9 Feedback network . . . 92
6.3 Pre-charger circuit . . . 93
6.4 Pulse-shaping filter . . . 94

7. IMPLEMENTATION . . . 98
7.1 Introduction . . . 98
7.2 The Heisenberg chip layout . . . 98
7.2.1 Floorplan . . . 99
7.2.2 Amplifier slice and parallel connection . . . 100
7.2.3 Feedback network, pre-charger and output buffer . . . 102
7.3 Heisenberg Test PCB . . . 103
7.4 Firmware and data sampling . . . 105

8. TEST RESULTS . . . 109
8.1 Introduction . . . 109
8.2 Noise PSD and bandwidth scaling . . . 109
8.3 Input stimuli and noise measurements . . . 110
8.4 Pad leakage, feedback reset and baseline . . . 111
8.5 Test methodology . . . 112
8.6 CSA step response . . . 113
8.6.1 Setup . . . 113
8.6.2 Results . . . 114
8.6.3 The baseline . . . 115
8.6.4 Step Amplitude . . . 117
8.6.5 Bandwidth . . . 120
8.7 Filter output . . . 124
8.8 Noise measurements . . . 125
8.8.1 Setup . . . 126

8.8.2 Results 127
8.8.3 Noise magnitude and thermal noise 129
8.8.4 Flicker noise 131
8.8.5 Noise scaling 134
8.9 Summary of design flaws 135

9. CONCLUSION 137
9.1 Summary 137
9.2 Future work 138

REFERENCES 140

ANNEXES 146

A. Second-order transfer function of the CSA 147

LIST OF FIGURES

1.1 High-level view of the analog or mixed-signal design flow (Gielen and Rutenbar (2000)). 2

1.2 Cut-away view of the ATLAS detector (ATLAS Collaboration (2008)). 5

1.3 Block diagram for a single channel of a generic pulse processing circuit for particle physics experiments. 6

1.4 Front-end amplifier for particle physics instrumentation using a charge-sensitive amplifier. The charge generated by the detector is integrated on the feedback capacitor of the voltage amplifier to obtain a voltage step at the output. 7

1.5 Pulse shaping action done by the filter. The filter sets the peaking time τ_P, which in turn sets the conflicting requirements of noise bandwidth and system speed. 8

1.6 Equivalent representation of a noisy linear circuit as a noiseless linear circuit with external noise generators, referred to a single arbitrary port. 10

1.7 g_m/I_D $vs.$ normalized drain current $I_\square$ for different values of L on a 0.5-μm technology. Data points obtained from LTspice simulations using BSIM3 MOSFET models. 13

1.8 Relevant design ratios $v.s.$ g_m/I_D for different values of L on a 0.5-μm technology. Data points obtained from LTSpice simulations using BSIM3 MOSFET models 14

2.1 Illustration of the parallel connection of identical copies of a circuit. 17

2.2 Proposed layout design and parallel-connection scheme. 22

3.1 Schematic representation of a typical front-end circuit in particle physics experiments. The schematic includes the circuit models for the detector, the charge-sensitive amplifier, and the pulse-shaping filter. 26

3.2 Schematic for noise analysis. Noise sources from two devices are considered: shot noise from the detector; and amplifier noise, represented as both voltage and current noise. The amplifier noise includes both thermal and flicker noise processes. 28

3.3 Graphical representation of the pulse shaper output for a single electron of input charge, considering an arbitrary pulse shape. The maximum value of the signal is reached at time $t = \tau_P$. 29

3.4 Equivalent noise charge $vs.$ shaping time. The plot shows two equivalent ENC curves to illustrate the effect of the $1/f$ noise over the voltage noise. (Spieler (2005)) 35

3.5 Example of the normalized noise spectra for a single transistor including white and flicker noise, for different values of g_m/I_D. Relevant simulation parameters are $L = 0.6$ μm and HSPICE NMOS parameter $K_F = 10^{-30}$. 38

3.6 F_o as a function of I_D. Several values of g_m/I_D are plotted. A global minimum envelope curve, obtained numerically using g_m/I_D values ranging from 1 to 24 is also included. Both white and flicker noise sources are considered. Relevant simulation parameters include $L = 0.6$ μm and MOSFET SPICE model $K_F = 10^{-30}$. Other parameters include the use of a RU-2 filter, $\tau_P = 10$ μs and $C_K = 1$ pF. 41

3.7 Corner frequency and the reciprocal time constant for a flicker exponent of $A_F = 1$ in a 0.5-μm technology. 42

3.8 Relative effect of flicker noise on the total output noise. Flicker exponent $A_F = 1$ and RU-2 filter noise coefficients were considered. 43

3.9 Value of g_m/I_D for overall minimum noise when the capacitance matching condition is met. 44

3.10 Optimum condition for peaking time values of $\tau_P = 4$ μs and $\tau_P = 20$ μs. . . 45

3.11 F_o as a function of I_D. The same parameters as in Figure 3.6 were considered, with the exception of τ_P. 46

3.12 Normalized amplifier white noise PSD $vs.$ g_m/I_D. For strong inversion operation white noise is $\propto (g_m/I_D)^{-1}$ which is directly observable on the plot. This inverse relation with g_m/I_D translates into diminishing returns in noise performance with increasing values of g_m/I_D. 47

3.13 Width per unit current W/I_D, reciprocal of current density I_D/W. 48

4.1 Cell-based layout with horizontally aligned, vertically stacked cells. 56

4.2 Schematic for Monte Carlo mismatch simulations. A single equivalent device circuit is considered. Simulation parameters include $L_j = 0.6\mu\mathrm{m}$, $W_j = 127.5\mu\mathrm{m}$, $R_j = 1\Omega$, and $V_B = 1\mathrm{V}$. This bias voltage results in the device operating in strong inversion, at a $g_m/I_D \approx 4.5\ \frac{\mathrm{mS}}{\mathrm{mA}}$. 64

4.3 Mean input-referred noise PSD $vs.$ # of parallel connected cells. The results were obtained from simulation scenario #3, however, the resulting plots for the other simulation scenarios showed no measurable differences. 66

4.4 Absolute (A) and normalized (B) variance of the input-referred noise PSD $vs.$ # of parallel connected cells, for simulation scenario #1. 66

4.5 Absolute (A) and normalized (B) variance of the input-referred noise PSD $vs.$ # of parallel connected cells, for simulation scenario #2. 67

4.6 Absolute (A) and normalized (B) variance of the input-referred noise PSD $vs.$ # of parallel connected cells, for simulation scenario #3. 67

4.7 Mean output integrated noise $vs.$ # of parallel connected cells. The integrated bandwidth ranges from $0.1\ \mathrm{mHz}$ to $100\ \mathrm{kHz}$. The results were obtained from simulation scenario #1. The resulting plots for the other simulation scenarios showed no observable differences. 70

4.8 Absolute (A) and normalized (B) variance of the output integrated noise $vs.$ # of parallel connected cells, for simulation scenario #1. 70

4.9 Absolute (A) and normalized (B) variance of the output integrated noise $vs.$ # of parallel connected cells, for simulation scenario #2. 71

4.10 Absolute (A) and normalized (B) variance of the output integrated noise $vs.$ # of parallel connected cells, for simulation scenario #3. 71

4.11 Simulation scenario #1 – Histogram of the output integrated noise for different numbers of parallel connected cells. A Kernel curve estimation of the probability density function is also included for each case, computed using the MATLAB `histfit` command. 74

4.12 Simulation scenario #2 – Histogram of the output integrated noise for different numbers of parallel connected cells. A Kernel curve estimation of the probability density function is also included for each case. The histogram for $k = 1$ was omitted, given that there is no dispersion in the noise measurements in that particular case. 75

4.13 CSA Monte Carlo simulation – Normalized variance comparison: switches (A) $vs.$ no switches (B). 76

5.1 Testing system block diagram. 78

5.2 Heisenberg block diagram. 79

6.1 Generic CSA schematic for linearity analysis. 82

6.2 Schematic of a generic closed-loop CSA for frequency response analysis. . . 84

6.3 Schematic of a NMOS-input folded-cascode amplifier. 86

6.4 Schematic of the folded-cascode amplifier for small-signal analysis. 87

6.5 Schematic for the amplifier bias circuit. 90

6.6 Parallel connection scheme. 91

6.7 Schematic for the feedback network. 92

6.8 Schematic for the pre-charger circuit. 93

6.9 Signal diagram for pre-charging and feedback reset. 94

6.10 Schematic of a generic $CR - nRC$ filter. 95

6.11 Step response of a $CR - nRC$ filter for $n = \{1, 2, ..., 6\}$. 96

6.12 Schematic of the filter in the Heisenberg test board. 97

7.1 Heisenberg chip micrograph. 98

7.2 The Heisenberg chip layout. Die dimensions: $3\text{-mm} \times 3\text{-mm}$. 99

7.3 Heisenberg chip floorplan for a single PMOS based channel. 100

7.4 Amplifier slice layout for both PMOS and NMOS input device designs. . . . 101

7.5 Different approaches to slice stacking. 102

7.6 Feedback network and pre-charger circuit layout. 103

7.7 Operational amplifier cell layout. 103

7.8 Heisenberg test board. Dimensions: 124.9-mm $\times$ 64.4-mm. 104

7.9 3D model of the board stack between the Heisenberg test PCB and the FPGA board. 105

7.10 Numato Saturn - Spartan-6 FPGA Development Board. Dimensions: 62.8mm $\times$ 57.5mm. 105

7.11 Timing for the operation of the pre-charger circuit and data sampling. The timescale is distorted to better display the signal sequence. 107

8.1 Example of a noisy waveform ($k = 2$, $C_F = 8\,\text{pF}$). 110

8.2 Charge-sensitive amplifier output for different number of parallel-connected slices. Both measured waveforms and simulation waveforms are presented, for different values of C_F. The measured waveforms were obtained using an oscilloscope, and averaging 8192 identical events for each value of k to remove noise. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$. 114

8.3 Close-up of the baseline values just before the CSA step response. 116

8.4 Charge-sensitive amplifier output for different number of parallel-connected slices, without baseline. Both measured waveforms and simulation waveforms

are presented, for different values of C_F. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$. 117

8.5 Close-up of the amplitude values for the CSA step response. These values are limited to the AC amplitude, and do not include the baseline. Furthermore, these values were sampled when the waveform reached steady-state. 118

8.6 Plot for $V(k) = Q_I/(C_F + k \cdot C_{FP})$, where Q_I and C_{FP} were obtained through linear regression of the measured amplitude results. The estimated values are $Q_I = 1.635\,\mathrm{pC}$ and $21.34\,\mathrm{fF}$ for the $4\,\mathrm{pF}$ curve, and $Q_I = 3.235\,\mathrm{pC}$ and $21.69\,\mathrm{fF}$ for the $8\ \mathrm{pF}$ curve. 118

8.7 Charge-sensitive amplifier output for different number of parallel-connected slices, without baseline and normalized to unit amplitude. Both measured waveforms and simulation waveforms are presented, for different values of C_F. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$. 121

8.8 Schematic of the NMOS-input folded-cascode amplifier for small-signal analysis, with capacitive feedback. An explicit capacitor to account for the non-dominant pole was added to the node v_x to analyze the second-order response of the circuit. 123

8.9 $CR - 2RC$ filter output for different number of parallel-connected slices. Both measured waveforms and simulation waveforms are presented for $C_F = 4\ \mathrm{pF}$. 125

8.10 Noise measurements for all the combinations of k and C_F used in the testing of the Heisenberg chip. The plot is presented with a logarithmic y axis to be able to display all curves properly. 126

8.11 Noise measurements for an increasing number of parallel-connected slices, and two values of the feedback capacitance C_F. A nonlinear least-squares fit is also included using the model for expected behavior ($\propto 1/k$). 127

8.12 Histograms for the filter output voltage, being sampled at the peaking time, for different numbers of parallel-connected slices (k) using $C_F = 4$ pF. The variance of each histogram represents the noise for that particular configuration of variables. 128

8.13 Individual waveforms for the filter output for $C_F = 1$ pF, which is the smallest value available for the feedback capacitance on the Heisenberg chip. Two different values of k ($k = 1$ and $k = 8$) are presented. 133

8.14 Handpicked iterations of the Monte Carlo simulation of the CSA ($C_F = 8$ pF). 134

A.1 Schematic of the NMOS-input folded-cascode amplifier for small-signal analysis, with capacitive feedback. An explicit capacitor to account for the non-dominant pole was added to the v_x node to analyze the second-order response of the circuit. 147

LIST OF TABLES

2.1 Normalized MOSFET noise equations (Alvarez, Avila, Campillo, Dragone, and Abusleme (2012)). 20

3.1 Some commonly adopted time-invariant filters and the corresponding ENC noise coefficients. The ratio between the pulse width (τ_W) and the peaking time (τ_P) is also reported. Flicker exponent $A_F = 1$ was considered (De Geronimo and O'Connor (2005)). 34

3.2 Normalized noise equations (Alvarez et al. (2012)). 36

4.1 Examples of process and electrical parameters. 52

4.2 Single device Monte Carlo simulation – Random and gradient standard deviation values. 65

4.3 Charge-sensitive amplifier Monte Carlo simulation – Random and gradient standard deviation values. 68

5.1 Specifications summary. 81

6.1 Transistor design parameters for the charge-sensitive amplifier circuit. 89

6.2 Transistor design parameters for the amplifier bias circuit. The drain current for all transistors is $I_D = 25\ \mu$A. 90

6.3 Filter parameter values. 97

7.1 FPGA digital I/Os. 106

7.2 Timing parameters. 107

8.1 Parameters for noise calculations. 131

ABSTRACT

Advances over the last decades in electronic design automation (EDA) for the design of digital integrated circuits, have led to the development of a robust set of tools and methodologies that automate almost all low-level phases of the digital design workflow. In contrast, analog integrated circuit design remains a mostly handmade, time-consuming and knowledge-intensive process. The amount of design iterations can be heavily cut down by the use of realistic value tables through the g_m/I_D design technique; however, the process still remains time-consuming and error-prone, with an end result of limited applicability beyond the scope of the initial specifications.

The slice-based design methodology, first introduced in this thesis, is a new approach to analog integrated circuit design, suitable for implementation in EDA tools, that aims to help reduce the amount of time and expertise required from the user. This methodology, inspired by the g_m/I_D design technique, is based on the use of pre-designed circuit cells, which can be connected in parallel to scale important performance metrics.

This thesis serves as a practical exploration of the slice-based design methodology. Given the difficulty of assessing the applicability and practicality of the proposed design methodology to any arbitrary circuit topology, it was decided to limit the scope of the analysis to a particular target application: low-noise charge-sensitive amplifiers (CSA) used for instrumentation in particle physics experiments. Within this context, a custom application-specific integrated circuit (ASIC) was designed, fabricated and tested, which includes a CSA designed with the slice-based technique, to evaluate practical design considerations and measure real-world performance.

Keywords: Electronic design automation (EDA), Slice-based design, Particle physics instrumentation, Charge-sensitive amplifier (CSA), Electronic noise, Device mismatch

RESUMEN

Los avances durante las últimas décadas en la automatización del diseño electrónico (EDA) para el diseño de circuitos integrados digitales, han llevado al desarrollo de un conjunto de herramientas y metodologías robustas que automatizan casi todas las etapas de bajo nivel en el flujo de diseño digital. En contraste, el diseño de circuitos integrados analógicos sigue siendo un proceso hecho mayormente a mano, que requiere mucho tiempo y conocimientos. La cantidad de iteraciones de diseño puede reducirse con el uso de tablas de valores realistas mediante la técnica de diseño g_m/I_D, sin embargo, el proceso sigue siendo lento y propenso a errores, con un resultado final de aplicabilidad limitada más allá del alcance de las especificaciones iniciales.

La metodología de diseño basado en *slices*, introducida por primera vez en esta tesis, es un nuevo enfoque para el diseño de circuitos integrados analógicos, adecuado para la implementación en herramientas EDA, que tiene como objetivo ayudar a reducir la cantidad de tiempo y conocimiento requerido por el usuario. Esta metodología, inspirada en la técnica de diseño g_m/I_D, se basa en el uso de celdas de circuito pre-diseñadas, que pueden ser conectar en paralelo para escalar medidas de desempeño importantes.

Esta tesis sirve como una exploración práctica de la metodología de diseño basada en *slices*. Dada la dificultad de evaluar la aplicabilidad y practicalidad de la metodología de diseño propuesta a una topología de circuito arbitraria, se decidió limitar el alcance del análisis a una aplicación objetivo en particular: amplificadores sensibles a la carga (CSA) de bajo ruido utilizados en instrumentación para experimentos de física de partículas. En este contexto, un circuito integrado de aplicación específica (ASIC) *custom* fue diseñado, fabricado y probado, que incluye un CSA diseñado con la técnica basada en *slices*, para evaluar consideraciones prácticas de diseño y medir el desempeño real del circuito.

Palabras Claves: Automatización de diseño electrónico (EDA), Diseño basado en *slices*, Instrumentación para física de partículas, Amplificador sensible a la carga (CSA), Ruido electrónico, *Mismatch* en dispositivos

1. INTRODUCTION

1.1 Analog integrated circuit design

1.1.1 The design process

The proliferation of consumer electronics has been a driving factor in the advancement of integrated circuit (IC) design towards increasingly complex circuits and ever smaller process technologies. The move towards design complexity has been aided by a mature and widely available set of tools for Electronic Design Automation (EDA) in the digital domain. To take advantage of these tools, circuit functions (*e.g.* signal processing) are implemented in the digital domain whenever possible. In stark contrast, analog IC design lacks the automation tools that facilitate the design process, and remains essentially hand-crafted by analog designers, on technologies typically optimized for digital applications. Due to this comparative disadvantage, on the design of systems on a chip (SoC)[1], it is typically the development cycle of the analog blocks that bottleneck the design process, even though they comprise only a small area of the entirety of the chip (Martins, Lourenço, and Horta (2012)).

A typical electronic design flow for analog and mixed-signal integrated circuits is performed using a top-down approach, an example of which is shown graphically in Figure 1.1. Three levels of abstraction can be readily identified during the design process: the system-level, where system specifications are set and functional blocks are identified; the circuit level, where circuit schematics are designed for each functional block; and the layout-level, where the circuit layout for all the functional blocks is designed, followed by floorplanning, placement and global routing to generate the layout of the entire system. Simulation and verification steps are performed at each level to account for undesired effects (*e.g.* layout parasitics) and detect potential problems, and if the design fails to meet specifications at some point in the design flow, redesign iterations are performed.

[1] An SoC may include analog, mixed-signals and digital blocks implemented on the same die

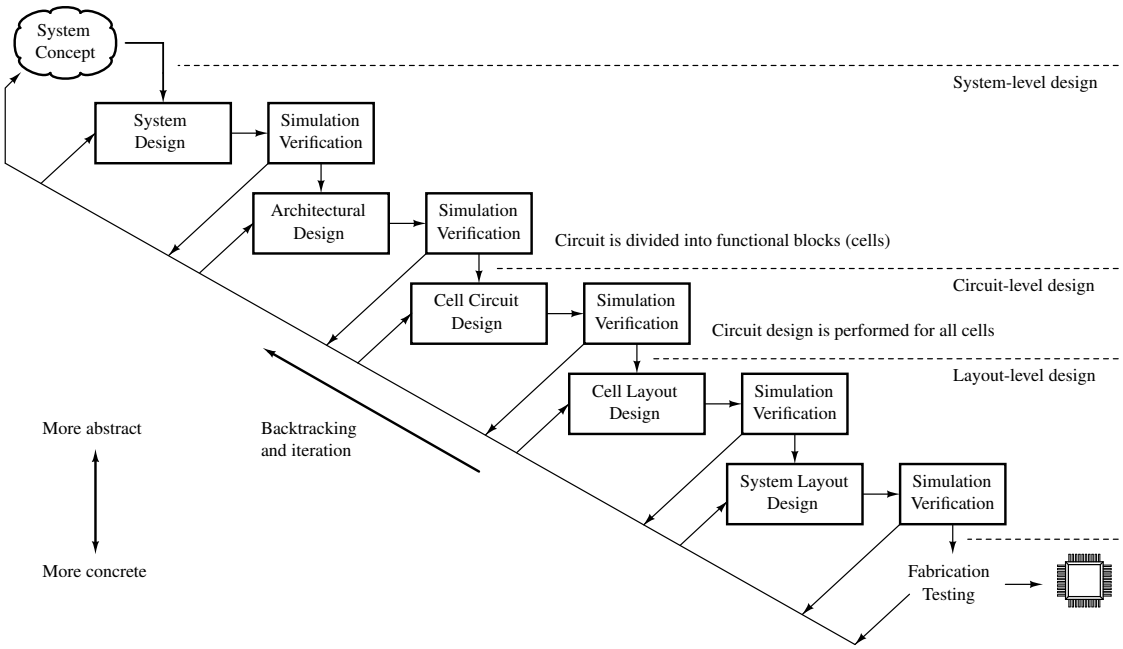


FIGURE 1.1. High-level view of the analog or mixed-signal design flow (Gielen and Rutenbar (2000)).

The circuit-level design is particularly challenging, as it often requires a custom optimized design, which is typically an underconstrained problem, with many degrees of freedom, and with many (often conflicting) performance requirements that must be taken into account (Gielen (2007)). To solve this problem effectively and produce an optimized design, an analog designer is required to have an advanced knowledge of device behavior, circuit topologies and design trade-offs. For these reasons, the analog design process is generally perceived to be less systematic, more heuristic, and much more knowledge-intensive than digital design (Gielen and Rutenbar (2000)).

1.1.2 About systematization and automation

While EDA tools for analog design have not reached the level of maturity to be widely adopted, computer-aided design (CAD) tools have been fundamental to tackle the design flow for decades. An analog designer will routinely use circuit simulators (*e.g.* LTspice (Analog-Devices (1999))), layout editing environments (*e.g.* Virtuoso (Cadence (2006))) and verification tools (*e.g.* CALIBRE (Mentor (2006))) to reach an optimized design.

The absence of mature EDA tools for analog design is not due to lack of trying, as research into the topic has been going on since the mid-1980s. Three distinct hierarchical levels are identified in the literature for analog design automation (ADA) (Gielen and Rutenbar (2000)): topology selection, where the most appropriate circuit topology is selected based on the given specifications; specification translation, where high-level specifications are mapped into sub-blocks, and at the lowest level, into device sizes; and layout generation, the creation of the geometrical layout of the low-level sub-blocks and the place and route of these sub-blocks at a higher level. A thorough and comprehensive review of the literature and the state of the art of these topics can be found on (Barros, Guilherme, and Horta (2010)) for topology selection and specification translation, and on (Martins et al. (2012)) for layout generation.

At the circuit level, analog designers rely heavily on hand analysis and circuit simulators to derive low-entropy expressions (Middlebrook (1991)) suitable for design. Among the techniques that allow some degree of systematization to the process, the g_m/I_D technique stands out (Silveira, Flandre, and Jespers (1996), Flandre, Viviani, Eggermont, Gentinne, and Jespers (1997)). The technique relies on the use of g_m/I_D as a design variable, which is a measure of the level of inversion of a transistor, and the use of tables for g_m/I_D-dependent parameters built from precise simulation results, both of which contribute to a more insightful approach to the design process. A detailed explanation of the importance of the g_m/I_D ratio and the homonymous design methodology is presented in Section 1.6.

1.2 Topics covered by the Thesis

The present thesis explores a technique for analog design, namely the slice-based design technique, suitable for implementation in EDA tools at the circuit and layout levels, but does not borrow concepts and techniques traditionally used in the literature of EDA, and was instead inspired in the g_m/I_D design technique.

In order to explore the proposed design technique and as a proof on concept, particle physics instrumentation was selected as the target application, and an integrated circuit,

namely the Heisenberg ASIC[2], was designed, fabricated and tested. The Heisenberg chip prominently includes a charge-sensitive amplifier (CSA) with configurable performance, and the main metric used to test the circuit was noise performance.

1.3 Particle physics experiments

Particle physics, also known as high-energy physics, is the field of natural science that studies the elementary particles that constitute matter, and the interactions between them (Nagashima (2013)). At the most fundamental level, that is, at the smallest scale, matter is constituted by a handful of different types of particles, replicated in astronomical quantities. Currently, the classification of all known elementary particles and their interactions is described by the Standard Model.

To peer deep into matter, particle physicist must first isolate elementary particles for study. Barring protons and electrons, isolated elementary particles do not occur naturally, but are instead created during high-energy collisions with other particles. There are three main particle sources that physicist use for experimentation: cosmic rays, nuclear reactors and particle accelerators (Griffiths (2008)).

In order to detect and measure the properties of elementary particles, particle physicist employ incredibly complex detector systems, in the context of large-scale experiments, which are among the most ambitious engineering projects mankind has ever built. A prime example of a type of particle physics experiments are particle colliders, the most prominent being the LHC[3] (Evans and Bryant (2008)), located at CERN[4], on the outskirts of Geneva, Switzerland.

Particle colliders are comprised of a particle accelerator, a detector system and a data processing system. The accelerator increases the energy of two particle beams traveling in opposing directions, and focuses them on a collision point. The resulting debris is

[2]Application-specific integrated circuit.

[3]Large Hadron Collider.

[4]European Organization for Nuclear Research.

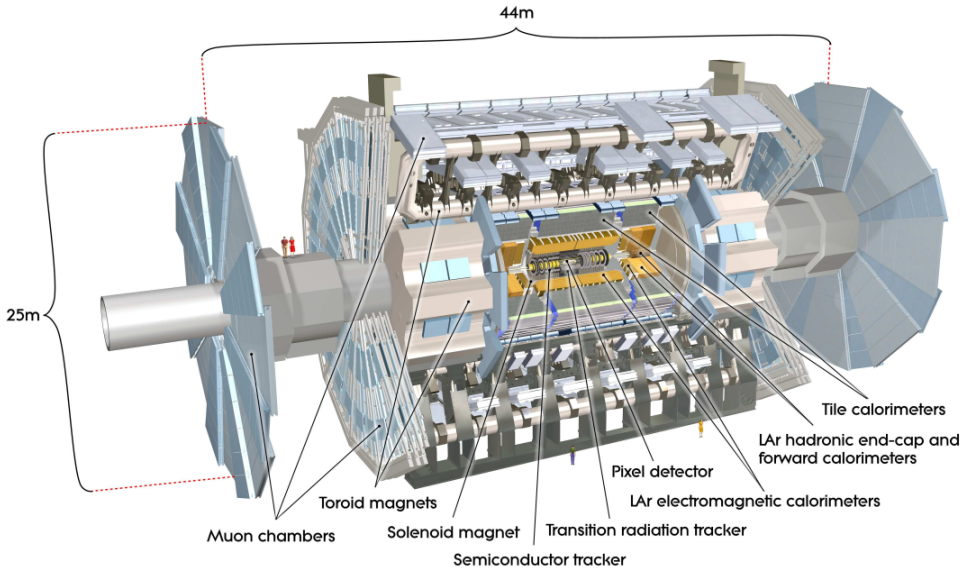


FIGURE 1.2. Cut-away view of the ATLAS detector (ATLAS Collaboration (2008)).

measured by the detector system, and the data stored for subsequent analysis. In modern particle accelerators, particles are made to collide at MHz rates during periods extending several months. The statistics of the large amount of accumulated data is then used to identify particles and measure their properties. An example of the results of particle physics experimental research is the recent of the Higgs Boson by the ATLAS (ATLAS Collaboration (2012)) and CMS (CMS Collaboration (2012)) experiments at the LHC.

Around the interaction point of a collider, a multi-layered detector system measures the collision debris, where each detector layer fulfills a different and specific purpose (*e.g.* particle tracking or calorimetry), with a multitude of individual detectors or detector channels on each layer. Figure 1.2 shows and example of one such detector system, the ATLAS detector (ATLAS Collaboration (2008)), centered at one of the collision points of the LHC.

As technology has advanced, and as particle physicist continue to pursue increasingly higher energies, the performance requirements of the individual detectors that comprise an experiment have grown ever more demanding. Modern finely segmented detectors can

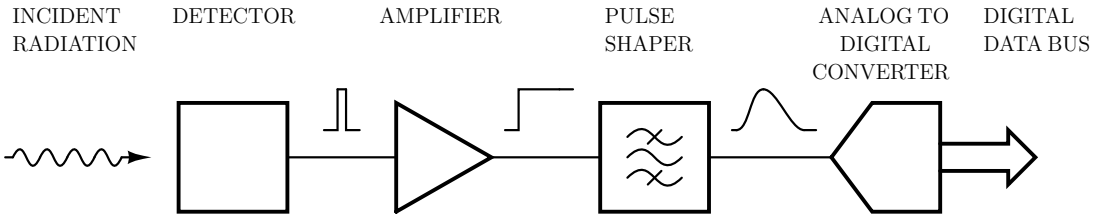


FIGURE 1.3. Block diagram for a single channel of a generic pulse processing circuit for particle physics experiments.

have hundreds of channels each, which are read by a handful of multichannel ASICs. Each one of these signal channels is read-out by an analog front-end, designed to comply with stringent speed, noise and power specifications.

The results of the technological advancements brought about by particle physics research is not limited to scientific understanding. From the World Wide Web, to modern medical imaging (*e.g.* MRI, PET scan), to the touchscreen that everyday consumer electronics use, many modern technologies owe their origin to the technological advancements brought about by particle physics experimentation.

1.4 Electronics for particle physics experiments

Although particle physics detector systems can take many different forms, their associated electronics perform the same basic functions (Grupen and Shwartz (2008)). The signal from the detector or detector channel in a multichannel array must be acquired, amplified, filtered and stored for subsequent analysis. A single channel of a generic particle physics detector system includes the detector, an amplifier, a filter, an analog-to-digital converter (ADC), and a readout circuit (Spieler (2005)). Figure 1.3 shows a simplified block diagram for a generic detector channel. A brief description of the function and operation of these functional blocks is presented below.

1.4.1 Detector

The detector converts the energy deposited by the particle into an electrical signal, typically in the form of a finite amount of electrical charge proportional to the absorbed

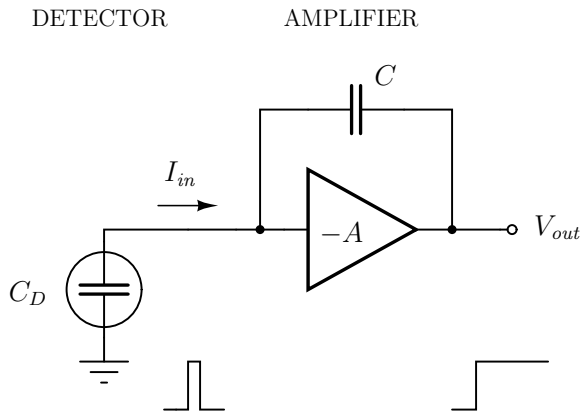


FIGURE 1.4. Front-end amplifier for particle physics instrumentation using a charge-sensitive amplifier. The charge generated by the detector is integrated on the feedback capacitor of the voltage amplifier to obtain a voltage step at the output.

energy. This can be achieved in a variety of ways, although the physical phenomena used for particle detection is out of the scope of this document. Common detector technologies include ionization chambers, scintillation counters, semiconductor detectors, and Cerenkov detectors, among others.

Detectors can also be classified according to their purpose, such as calorimeters and trackers. The former is used to measure the kinetic energy of particles by stopping the particle within its structure, the latter is used to detect particle trajectories, so they are typically highly segmented in order to obtain a good spatial resolution. More detailed information related to particle detectors can be found in (Grupen and Shwartz (2008)).

1.4.2 Amplifier

The front-end amplifier, also known as preamplifier, translates the electrical charge generated by the detector into a voltage signal. The charge-to-voltage translation is done by transfering the charge Q_{in} from the nonlinear capacitance of the detector C_D to a known capacitor C. The output voltage V_{out} of the amplifier is given by $V_{out} = Q_{in}/C$, and the gain of the amplifier is naturally measured in $[V/C]$ or $[F^{-1}]$. The most common preamplifier implementation consists of a voltage amplifier with a capacitor in negative

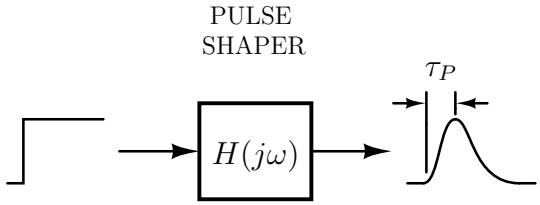


FIGURE 1.5. Pulse shaping action done by the filter. The filter sets the peaking time τ_P, which in turn sets the conflicting requirements of noise bandwidth and system speed.

feedback configuration, as shown in Figure 1.4. The resulting feedback circuit is a charge-sensitive amplifier (CSA), which has been extensively studied in the literature related to particle physics instrumentation (Snoeys et al. (2000), Aspell et al. (2001), De Geronimo and O'Connor (2005), O'Connor and De Geronimo (1999), Alvarez et al. (2012)).

1.4.3 Pulse Shaper

The primary function of the pulse shaper is to improve the signal-to-noise ratio (SNR). The frequency spectra of the signal and the noise differ, so it is possible to improve the SNR by applying a filter that tailors the frequency response to favor the signal, while attenuating the noise (Spieler (2005)). The filter also changes the time response of the input signal, as shown in Figure 1.5, reason why this functional block is referred to as pulse shaper. In this context, the terms pulse shaper and filter are used interchangeably.

The pulse shaper is typically an analog block, either time-invariant or time-varying, which sets both the speed and the total noise of the output signal before digitization. When designing a particle physics instrumentation system, it is necessary to find a balance between these two conflicting requirements.

1.4.4 Digitizer

The output of the pulse shaper is sampled at the peak value of the pulse shape and digitized using an analog-to-digital converter (ADC). On highly segmented multichannel detectors, the ADC is typically included on the front-end ASIC, and can be either dedicated or shared among multiple channels.

1.5 Noise in electronics

Noise is a generic term used to refer to any and all unwanted disturbances on a physical process. In electronics, noise is a current or voltage that is unwanted on an electrical circuit (Baker (2010)), which in turn manifests as unwanted disturbances on the desired signals of the circuit. Electronic noise sets the lower amplitude bound for the achievable resolution of a circuit, below which signal and noise are indistinguishable.

Types of noise can be broadly categorized as 1) electronic noise, which is the result of the discrete and random movement of electrical charge on an electrical device (*e.g.* thermal noise, shot noise, or flicker noise), 2) quantization noise, resulting from analog to digital domain conversions and vice versa, due to the discrete nature of digital signals, and 3) common-mode noise, resulting from extrinsic signals being picked up by the circuit and interfering with the desired signal.

Electronic noise is of particular importance to an analog circuit designer, and the main focus of the present document, since it is a direct consequence of the low level circuit design, unlike quantization noise and common-mode noise, which can be attributed to system level specifications (*e.g.* ADC resolution), and layout and implementation considerations (*e.g.* bad layout practices, poor EMI[5] shielding), respectively.

In integrated circuits, semiconductor devices (*e.g.* diodes, bipolar and MOS transistors) and resistors are the main sources of noise. A proper understanding of the underlying noise generating processes of MOS transistors is necessary for the design of low noise analog circuits in CMOS technologies. There are three uncorrelated noise generating processes in MOSFETs (Gray, Hurst, Lewis, and Meyer (2001)): 1) shot noise due to gate leakage current, 2) channel noise, which is a combination of thermal noise in strong inversion levels and shot noise in weak inversion levels, 3) and flicker or $1/f$ noise, also known as low-frequency noise. For moderate inversion levels, channel noise is modeled as a weighted average of both thermal and shot noise sources.

[5]Electromagnetic Interference.

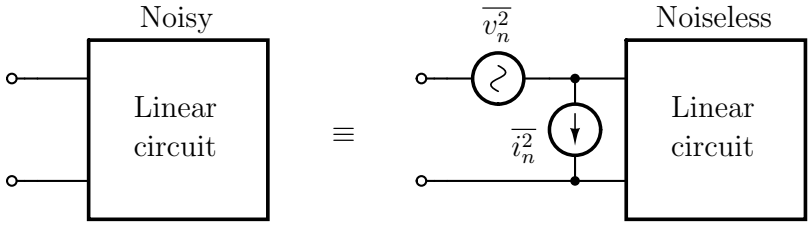


FIGURE 1.6. Equivalent representation of a noisy linear circuit as a noiseless linear circuit with external noise generators, referred to a single arbitrary port.

Since electronic noise is modeled as a stationary stochastic process, it is naturally characterized by its power spectral density (PSD), measured in $[\mathrm{V}^2/\mathrm{Hz}]$ or $[\mathrm{A}^2/\mathrm{Hz}]$. The integral of the PSD over the circuits bandwidth yields the total noise power, and its square root is the standard deviation or RMS[6] value of the voltage or current noise.

Any linear circuit with noisy components can be represented as an ideal noiseless linear circuit with external noise generators (Gray et al. (2001)). This situation is shown in Figure 1.6. These noise generators are characterized by their power spectral densities $\overline{v_n^2}(f)$ and $\overline{i_n^2}(f)$. Referring a noise source to one of the ports of a circuit is done by using the transfer function between the two.

For a fair comparison between the contribution of different noise sources on the same circuit, between the accumulated noise and the input signal, and even between different circuit topologies, is it often a good idea to refer the noise sources to the input port. When multiple noise sources are referred to the same port, the superposition of the noise contributions is applied in quadrature, considering correlation, as follows:

$$\sigma_{Tot}^2 = \sum_{i=1}^{N} \sigma_i^2 + \sum_{i \neq j} c_{ij} \cdot \sigma_i \cdot \sigma_j \tag{1.1}$$

where σ_i^2 represents the noise power of source i, N is the total number of noise sources, and c_{ij} is the correlation coefficient between sources i and j. The total noise of the circuit on a given port is the integral over frequency of all port-referred contributions.

[6]Root mean square.

1.6 The g_m/I_D design methodology

As CMOS technology has advanced, the behavior of the MOS transistor has changed enormously due to scaling and ever smaller operating voltages, making square-law equations obsolete (Razavi (2001)). In parallel, increasingly complex MOSFET models have been developed, *e.g.* the BSIM3 MOSFET model (Liu et al. (1998)), which accurately describe transistor behavior from weak to strong channel inversion, at the expense of complexity.

Design in moderate inversion, which is arguably the most complicated operating region to model, is increasingly important in modern, low voltage processes, due to high transconductance efficiency traded at a reasonable cost in bandwidth. In order for the analog designer to detach from overly complex low-level I-V transfer curves impractical for design, the g_m/I_D design methodology relies on the results of SPICE simulations, which are based on very accurate device models. Furthermore, the methodology employs the use of the inversion level of a transistor as a design variable, which leads to a more insightful approach to the design process. In this section, the importance of the g_m/I_D ratio and the basics of the design methodology of the same name are presented.

1.6.1 The g_m/I_D ratio as a design variable

Analog CMOS design is complicated by three degrees of design freedom present for every MOS device operating in the usual saturation region (Binkley (2007)). Traditionally, these design variables have been the drain current I_D, channel width W, and channel length L. However, selecting drain current, level of inversion, and channel length, provides better insight into device operation leading to a more optimized design, where channel width can be readily calculated as a consequence of these three design variables.

Among the figures of merit defined to quantify the level of inversion of MOS devices, the inversion coefficient IC (Vittoz (1994), Foty, Bucher, and Binkley (2002)), and the ratio between the transconductance g_m and the DC drain current I_D, hereafter referred to simply as g_m/I_D (Silveira et al. (1996), Flandre et al. (1997)), stand out. The latter will

be used for the remainder of the present document, and will serve as a design variable in the g_m/I_D design methodology.

The choice of g_m/I_D as a design variable is based on its relevance for three reasons:

1. It is directly related to the performance of MOS devices.
2. It provides an indication of the inversion level.
3. It can be used to calculate the dimensions of the transistor.

The first of these points is immediately apparent, as the g_m/I_D value is a measure of the efficiency to translate current into transconductance: if the transistor possesses a high g_m/I_D ratio, it means that it produces a large g_m per unit current. The unit for g_m/I_D is $[1/\mathrm{V}]$, or equivalently, $[\mathrm{mS}/\mathrm{mA}]$.

To better understand what information about the level of inversion of a device is provided by the g_m/I_D ratio, let us consider the following relation:

$$\frac{g_m}{I_D} = \frac{1}{I_D} \cdot \frac{\partial I_D}{\partial V_{GS}} = \frac{\partial \left[\ln(I_D)\right]}{\partial V_{GS}} = \frac{\partial \left\{ \ln\left[\frac{I_D}{\left(\frac{W}{L}\right)} \right] \right\}}{\partial V_{GS}} \tag{1.2}$$

In weak inversion, or subthreshold region, MOS devices behave as bipolar transistors (Gray et al. (2001)), and the dependence of the drain current I_D with the gate-to-source voltage V_{GS} is exponential. This means that the relationship between $\ln(I_D)$ and V_{GS} is linear, and the derivative is constant. This constant value is where the derivative (1.2) is maximized, and this maximum value is equal to $1/(nU_T)$, where n is the subthreshold slope factor, which is technology dependent, and U_T is the thermal voltage. As the g_m/I_D value decreases, the operating point moves towards strong inversion, or active region, where the dependence of I_D and V_{GS} is quadratic, which means that $\frac{\partial[\ln(I_D)]}{\partial V_{GS}} \propto \frac{1}{V_{GS}}$. Deeply into strong inversion, the relation between I_D and V_{GS} becomes almost linear due to velocity saturation (Silveira et al. (1996)).

The previously stated observations are better illustrated via a simple but powerful graphical method shown in Figure 1.7 (Foty et al. (2002)), where the g_m/I_D ratio is plotted against the normalized drain current $I_\square = I_D/(W/L)$, for a 0.5-$\mu\mathrm{m}$ technology. The

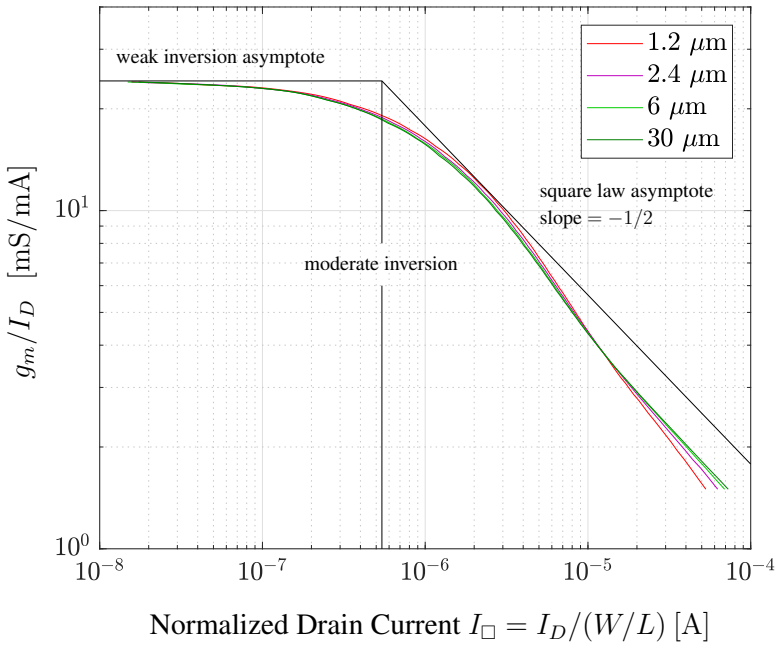
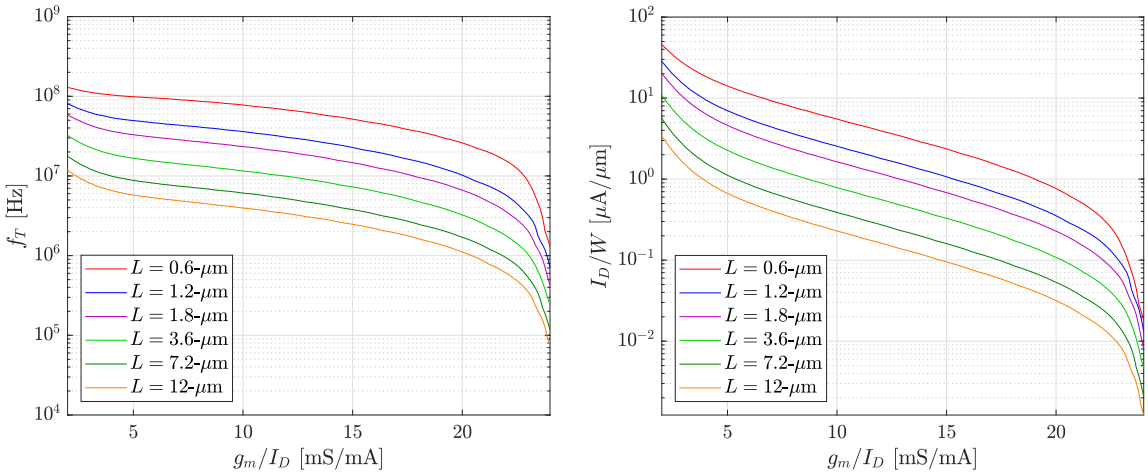


FIGURE 1.7. g_m/I_D *vs.* normalized drain current $I_\square$ for different values of L on a 0.5-μm technology. Data points obtained from LTspice simulations using BSIM3 MOSFET models.

simple limits of device behavior are shown as two asymptotes: the weak inversion asymptote, where g_m/I_D approaches the thermal voltage limit for transconductance efficiency, and the strong inversion asymptote, which illustrates the traditional square-law MOSFET model, in the form of a line plot with a slope of $-1/2$. Centered at the point where these two asymptotes intersect, lies the interpolation region of moderate inversion, which extends as a continuum between these two regions of operation.

1.6.2 Methodology

The basics of the g_m/I_D methodology are simple. Let us consider a transistor of width W and length L biased at a certain operating point, with a gate-to-source voltage V_{GS}, drain current I_D, transconductance g_m, and gate-to-source capacitance C_{gs}. If an identical transistor is connected in parallel, the equivalent device will have a width of $2 \cdot W$, drain current of $2 \cdot I_D$, transconductance of $2 \cdot g_m$, gate-to-source capacitance of $2 \cdot C_{gs}$, while the length L, the gate-to-source voltage V_{GS}, and the ratio g_m/I_D, a measure of the level



(A) Transit frequency f_T.

(B) Drain current density I_D/W.

FIGURE 1.8. Relevant design ratios $v.s.$ g_m/I_D for different values of L on a 0.5-μm technology. Data points obtained from LTSpice simulations using BSIM3 MOSFET models

of inversion of the channel, remain unchanged. There are several other ratios that are only dependent on the operating point of the transistor, independent of device width, including the transit frequency f_T (commonly defined as g_m/C_{gs}), drain current density I_D/W, and even normalized noise $\overline{I_n^2}/I_D$ (Alvarez and Abusleme (2012)).

The g_m/I_D methodology requires device characterization prior to the design stage, in a process that needs to be performed only once for a given technology. Through SPICE simulations, value tables for width-independent parameters can be computed and mapped as a function of g_m/I_D, for different values of L. Furthermore, value tables for parameters that are also a function of the drain-to-source voltage V_{DS}, such as the intrinsic gain of the transistor $g_m r_o$, can also be computed by adding another simulation axis for V_{DS}. After the simulations have been carried out and the values stored, interpolation functions in a numerical-analysis software (*e.g.* MATLAB, SciLab, Excel) can be used to retrieve specific values as a function of g_m/I_D and L. Figure 1.8 shows g_m/I_D curves for the drain current density I_D/W and transit frequency g_m/C_{gs} of a 0.5-μm technology, obtained from LTspice simulations while using BSIM3 MOSFET models. It can be inferred from

the drain current density plot I_D/W that, once I_D, g_m/I_D and L have been selected, the transistor width W can be calculated unambiguously.

With technology-specific value tables on hand, the design process can begin. The process involves the use of the three device design variables previously stated (I_D, g_m/I_D and L), low-entropy circuit equations expressed as functions of g_m/I_D-dependent parameters, simulations, and design scripts for iteration and optimization.

1.7 Thesis content

Chapter 2 provides an overview of the proposed slice-based design methodology for analog circuits, including the basic formulation, some specific examples, and an exploration of potential problems. Chapter 3 provides an in-depth look into the mathematical framework used in noise analysis of particle physics front-ends and the application of this framework to the specific problem of slice-based design, and also includes some novel ideas regarding flicker noise modeling using a g_m/I_D approach. Chapter 4 provides and analysis of the effects of device mismatch on the application of the proposed design methodology, in the form of both a mathematical model for mismatch as a function of a discrete number of parallel-connected circuit copies, and a Monte Carlo simulation to check the applicability of the model. In Chapter 5, the ASIC system-level design and specifications are presented. Chapter 6 shows the the circuit-level design of the ASIC, with a particular emphasis on the design of a single charge-sensitive amplifier cell and parallel connection scheme. Chapter 7 presents the layout-level implementation of the designed ASIC, together with the implementation of the test PCB and FPGA firmware. In Chapter 8, the test results are presented, and the potential implications of these results on the proposed design methodology are analyzed. Finally, Chapter 9 summarizes the results and contributions of this work, and presents ideas for future research.

2. SLICE-BASED DESIGN METHODOLOGY

2.1 Introduction

The inherent difficulties of the analog design process were presented in Section 1.1, and can be summarized as follows: despite great advances in EDA tools for digital design in the last decades, analog design remains a mostly handmade, time-consuming and knowledge-intensive process. The amount of design iterations can be heavily cut down by the use of realistic value tables through the g_m/I_D design technique, however, the process still remains time-consuming and error-prone, with an end result of limited applicability beyond the scope of the initial specifications.

The slice-based design methodology, object of study of the present thesis, aims to help reduce the amount of time and manual work required from the user, while at the same time lowering the barrier of entry to non-experts in the design analog integrated circuits. From the perspective of the user, the methodology involves the use of a library of optimized circuits to cover different corners of the design space. The circuits in this library are indivisible cells, hereafter referred to as slices, that can be connected in parallel in order to scale important performance metrics. Through the careful selection of the correct slice and number of parallel-connected circuits, a wide range of specifications can be met with minimal time investment from the user. Thus, the design of a library of optimized and fully characterized circuit slices is a precondition for this methodology to be of any use.

As it is hard to say whether this methodology will be usable or appropriate in any arbitrary application, it was decided to limit the scope of the present research to amplifiers, and more specifically, to charge-sensitive amplifiers used in particle physics instrumentation. The applicability of the proposed design methodology to other types of circuits and applications is left for future research.

The present chapter provides a detailed description of the proposed design methodology, and can be broadly divided into two sections: Section 2.2 explores the effects of connecting large circuits blocks in parallel, and Section 2.3 presents the methodology, its

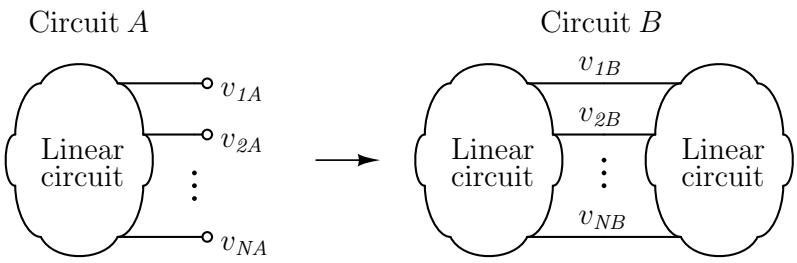


FIGURE 2.1. Illustration of the parallel connection of identical copies of a circuit.

implications at the circuit and layout levels, and potential problems with this approach to design.

2.2 The effects of connecting circuits in parallel

The basis of the slice-based design technique is in the parallel connection of complex circuit blocks, in order to meet load, noise and other relevant specifications that scale with parallel connection, at the expense of power consumption and die area. The idea of connecting circuits in parallel to increase drive capability, or the trade-off between power consumption and noise performance, are not new concepts in IC design (Harris et al. (1999), Razavi (2001)), however, the idea of connecting large and complex circuits in parallel in order to scale circuit performance as an approach to the design process has not been found on the literature.

2.2.1 The general case

When an arbitrary linear circuit (namely, circuit A) is connected in parallel to an identical copy of itself, *i.e.* each of the N nodes of the circuit is connected to the corresponding node of the identical copy, in the resulting circuit (namely, circuit B) some figures of merit and quantities change, whereas other stay the same. The concept is illustrated in Figure 2.1. For example, all the N node voltages remain, whereas all M branch currents are doubled:

$$v_{iB} = v_{iA} \tag{2.1}$$

$$i_{jB} = 2i_{jA} \tag{2.2}$$

As a consequence of this, all impedance elements Z_k are halved whereas all admittance elements Y_k are doubled. The latter includes transconductances as well:

$$Z_{kB} = \frac{Z_{kA}}{2} \tag{2.3}$$

$$Y_{kB} = 2Y_{kA} \tag{2.4}$$

This applies to both explicit passive elements (*e.g.* resistors, capacitors and inductors) and equivalent node impedances. Naturally, with the doubling of branch currents, power consumption is doubled as well. The operating point of all MOS devices stays the same, as the g_m/I_D ratio remains unchanged.

2.2.2 Single-pole amplifier

Some specific circuit blocks are also simple to analyze, such as an amplifier. Let us consider a single-pole amplifier with an open-loop gain of A_{OL} and a bandwidth of ω_c. The open-loop gain of the amplifier, which is a non-dimensional figure, can be expressed as the product of the circuit effective transconductance G_{meff} and the output resistance R_{Out}. On the parallel connection of identical circuits, the former increases and the latter decreases, while the open-loop gain remains constant:

$$A_B = G_{meffB} \cdot R_{OutB} = 2G_{meffA} \cdot \frac{R_{outA}}{2} = A_A \tag{2.5}$$

The same is true for the bandwidth of the amplifier. The equivalent capacitance of the dominant pole increases twofold, while the equivalent resistance seen by the capacitor decreases by the same quantity:

$$\omega_{cB} = \frac{1}{R_{eqB} \cdot C_{eqB}} = \frac{1}{\frac{R_{eqA}}{2} \cdot 2C_{eqA}} = \omega_{cA} \tag{2.6}$$

However, it is not uncommon for the bandwidth of a circuit to be set by an externally connected load. As long as the load is also parallel-connected, the bandwidth is maintained, otherwise the bandwidth would change. Nonetheless, the resulting parallel-connected amplifier has twice the drive capability of the single circuit.

2.2.3 Noise analysis

In terms of noise, the effects of connecting circuits in parallel are more involved. Let us consider an arbitrary linear circuit that has a single noise generator. The simplest example is a resistor, which generates thermal noise. The power spectral density (PSD) of the thermal noise generated by a resistor is directly proportional to its resistance when expressed as a voltage variance, whereas when expressed as a current variance, it is inversely proportional to its resistance:

$$\frac{\overline{V_n^2}(f)}{\Delta f} = 4kTR \left[\frac{\mathrm{V}^2}{\mathrm{Hz}}\right] \quad , \quad \frac{\overline{I_n^2}(f)}{\Delta f} = \frac{4kT}{R} \left[\frac{\mathrm{A}^2}{\mathrm{Hz}}\right] \tag{2.7}$$

When connecting two identical copies of a circuit in parallel, all equivalent resistance values are halved, which in turn means that voltage noise power is halved as well, while current noise power is increased by a factor of two. Since voltage signals remain unchanged, the signal-to-noise ratio (SNR), as expressed as a ratio of squared voltage signals, is increased by a factor of two. And since current signals double, current signal power quadruples, and the SNR, as expressed as a ratio of squared current signals, is increased by a factor of two as well.

Let us consider now the case of a single MOSFET transistor as the noise generator. It can be shown (Alvarez and Abusleme (2012)) that MOSFET voltage and current noise, including thermal noise (for strong inversion), shot noise (for weak inversion) and flicker noise, can be normalized and expressed as $\overline{V_n^2} = \overline{\hat{V}_n^2}/I_D$ for voltage noise, and as $\overline{I_n^2} = \overline{\hat{I}_n^2} \cdot I_D$ for current noise, where $\overline{\hat{V}_n^2}$ and $\overline{\hat{I}_n^2}$ are normalized voltage and current power spectral densities, which are solely a function of g_m/I_D, and I_D is the transistor drain current. An example of this can be seen in Table 2.1, which shows the normalized PSD for both thermal and shot noise of a MOSFET, expressed as functions of g_m/I_D.

In other words, for a constant g_m/I_D value, MOSFET voltage noise is inversely proportional to the drain current, while current noise is directly proportional to the drain current. When connecting two identical copies of a circuit in parallel, the equivalent transistor drain current doubles while the g_m/I_D value remains unchanged, which in turn means that

TABLE 2.1. Normalized MOSFET noise equations (Alvarez et al. (2012)).

	$\overline{V_n^2}/\Delta f$	$\overline{\hat{V}_n^2}/\Delta f$	$\overline{I_n^2}/\Delta f$	$\overline{\hat{I}_n^2}/\Delta f$
Thermal noise for strong inversion	$\frac{4k_BT\gamma}{g_m}$	$4k_BT\gamma\left(\frac{I_D}{g_m}\right)$	$4k_BT\gamma g_m$	$4k_BT\gamma\left(\frac{g_m}{I_D}\right)$
Shot noise for weak inversion	$\frac{2qI_D}{g_m^2}$	$2q\left(\frac{I_D}{g_m}\right)^2$	$2qI_D$	$2q$

voltage noise power is halved, while current noise power is increased by a factor of two. The same as in the resistor example, the SNR is increased by a factor of two, no matter which representation is used to reach the result.

It is straightforward to extend the previously stated observations to an arbitrary linear circuit. Let us consider an arbitrary CMOS circuit with many noise generators. The total noise on the output node of the circuit can be computed as follows:

$$\overline{V_{n,out}^2}(f) = \sum_{i=1}^{N} \overline{V_{n,i}^2}(f) \cdot |H_i(f)|^2 \tag{2.8}$$

where $\overline{V_{n,i}^2}(f)$ is the PSD of each individual noisy device, and $H_i(f)$ is the transfer function to the output for each individual noise source. Since the noise generators, transistors or resistors, are independent, their noise contributions are uncorrelated and are added in quadrature. For parallel connected circuits, it is immediately apparent that, since the transfer functions remain unchanged and the PSD of each individual noisy device is halved, the noise power measured on the output node of the circuit is also halved. The same is also true for the total integrated noise of the circuit, that is, the integral of the PSD over frequency, as long as the circuit bandwidth does not change.

2.3 The design methodology

2.3.1 The pre-design stage

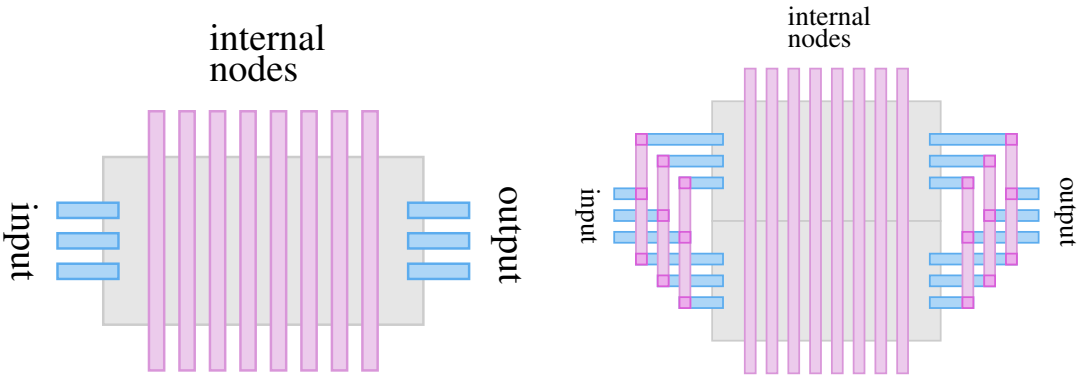
A pre-requisite to the application of the slice-based design methodology is the compilation of a library of optimized and fully characterized circuit designs. This task is done by an analog designer through the standard analog design workflow, with all the inherent

difficulties it has. The difference is that, once the design is done, it can be re-used in the future, as it was designed from the ground up for scalability.

At the circuit level, each amplifier cell will be optimized individually to meet a set of specifications, so that different cells cover different corners of the design space, *e.g.* maximum gain-bandwidth product, minimum noise, etc. Each device in the cell will have its operating point defined by its current and its g_m/I_D. The small-signal performance of the cell can be computed as a function of the resistances, transconductances and capacitances of individual devices. These equations can include node impedances, poles, effective transconductances, input-referred noise, among others, and will be part of the documentation for the cell. These equations can be re-computed and tabulated for increasing branch currents as the cells are connected in parallel. Currents can only increase in integer values of the unit cell, which is the indivisible unit. Any arbitrary scaling of a particular cell is still possible at a circuit level in order to achieve an optimized result for a particular application, however, this would require a custom layout.

For each slice, the corresponding bias circuit can be either integrated into the cell itself, or made into a separate, independent cell. Each approach has its own benefits and drawbacks: integrating the bias circuit into the cell allows for better device matching at the layout level, however, it might be power-inefficient when multiple cells are connected, as a single bias circuit could be enough to bias the resulting circuit. The use of an external bias circuit also allows for added granularity in the design, as it could be tweaked to modify the branch currents of the cell, and thus, the operating point of the different devices and the performance of the cell as a whole. While the latter is also possible for an integrated bias circuit, the process becomes more intrusive at the layout level, as the bias circuit might not be easily separable.

At the layout level, each amplifier cell needs to be designed from the ground up in a way that facilitates the parallel connection of multiple circuits. One such way is shown in Figure 2.2(A). In this scheme, each cell is implemented in a rectangular shape, with inputs and outputs on the sides, and all nodes running vertically. It could result convenient



(A) Single cell.

(B) Two cells connected in parallel.

FIGURE 2.2. Proposed layout design and parallel-connection scheme.

during the design process to reserve one or multiple metal layers for internal node traces for parallel connection, and focus on a compact design on the remaining layers. This approach favors functionality and simplicity, and serves as a proof of concept for the proposed design methodology. The optimal geometry for things such as intra-cell device matching or for minimum die area are out of the scope of the present document, and are left for future research.

2.3.2 The design stage

With a library of optimized circuits at hand, the design process can begin. The ASIC designer will pick a pre-designed cell according to the required specifications, and will scale it by connecting a number of copies in parallel, in order to meet load, noise and other relevant specifications that scale with parallel connection. Once the cell and number of parallel-connected circuits has been selected, the design can be validated through SPICE simulations. Should any incremental change be required on the cell, it is possible to tweak the biasing circuit to adjust all the currents. Through the use of g_m/I_D curves, the new current density I_D/W can be computed for all devices, from which the g_m/I_D value can be obtained, and all performance equations can be re-evaluated. This results in a circuit with a new set of specifications, and thus must be carefully evaluated by the designer.

Having the circuit-level design, the next step is layout. Figure 2.2(B) shows how two cells can be stacked and abutted, and the same scheme can be extended to any number of cells. Likewise, the biasing cell can be placed in the middle of the stack. This task can be efficiently automated through EDA tools, and by the use of a library of pre-designed and fully characterized circuit cells, the subsequent verification procedure for the resulting layout would become a mere sanity check. Through this design procedure, the time involved in the analog design blocks will be minimized, along with the associated uncertainties.

2.3.3 Adjustable performance

An unintended consequence of this approach to design is the added possibility of designing circuits that can scale dynamically according to real-time performance requirements. Through the use of switch banks to connect and power-on different numbers of parallel connected slices, the number of active copies of a circuit can be adjusted to meet changing specifications (*e.g.* load) while minimizing power consumption.

2.3.4 Potential issues

There are some concerns with the layout implementation that become apparent after careful analysis. First, there are some inherent parasitic components implied in the stackable layout due to the traces that connect the parallel cells, which might have an effect on performance depending on the circuit. Second, depending on the number of parallel connected circuits, the distance between cells might become large enough that mismatch related to process gradients becomes significant. Third, mismatch might also cause voltage differences between nominally identical nodes, which would translate into current flow through the wires that connect the parallel-connected circuits. The latter point is not exclusive to gradient-related mismatch, but can also occur due to size-related mismatch. Although a thorough analysis and understanding of each one these concerns is desirable, it was decided to leave them out of the scope of the present research document, with one notable exception, and are left for future research.

After the implementation of the Heisenberg chip, and to provide an explanation for some of the obtained results, the effects of device mismatch on the proposed design

methodology were further studied, and are presented in Chapter 4. It is possible that, depending on the specific technology, circuit topology and application, other issues might become dominant, such as wire parasitics.

3. NOISE ANALYSIS IN CHARGE-SENSITIVE AMPLIFIERS

3.1 Introduction

Particle physics detectors generate a finite amount of electrical charge in response to input stimuli, which is proportional to the amount of energy deposited by incident particles. One of the main purposes of the front-end electronics that process the output of the detector is to obtain an accurate measurement of this electrical charge in the form of a voltage signal, which is sampled and stored for subsequent analysis. In a properly designed front-end circuit, electronic noise sets the lower bound for the achievable resolution of the measurements. Thus, one of the main goals during the design phase of the front-end circuit is to maximize the signal-to-noise ratio (SNR) of the detector measurements.

The present chapter delves into the mathematical framework for noise analysis of particle physics front-end circuits in general, and the charge-sensitive amplifier in particular. Section 3.2 defines the equivalent noise charge (ENC), a figure of merit used to describe the noise performance of a front-end circuit, and presents a step-by-step derivation of a mathematical expression of the ENC useful for later sections. Section 3.3 further expands the ENC analysis by introducing a design-oriented methodology commonly used to simplify the front-end filter design. Finally, section 3.4 presents a design methodology for charge-sensitive amplifiers based on the g_m/I_D technique, and the application of this methodology to the slice-based design technique.

3.2 Equivalent noise charge

Since the output of a particle physics detector is electrical charge, one convenient way to compare system performance among different front-end circuits is to also express system noise in terms of charge. One common figure of merit used to describe the noise performance of a front-end circuit is the equivalent noise charge (ENC), measured in number of electrons. It is defined as the number of electrons of input charge necessary to produce an output signal-to-noise ratio (SNR) equal to 1. The current section presents a

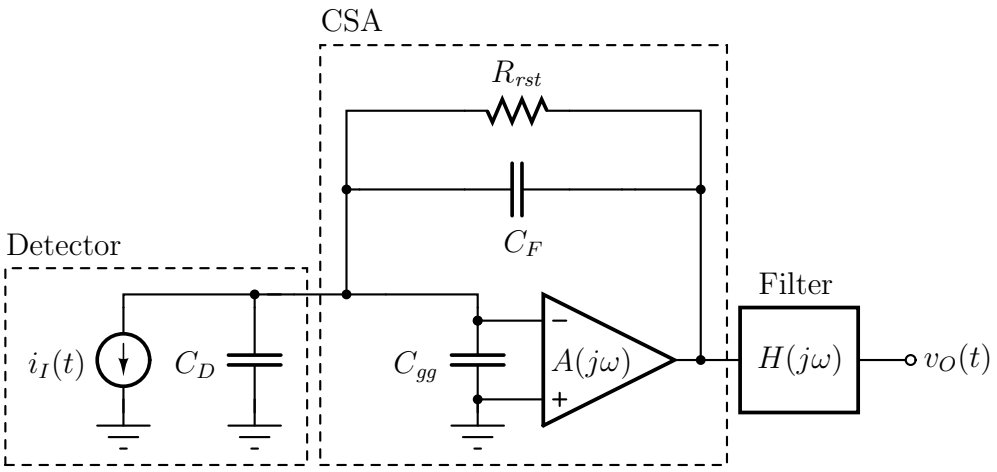


FIGURE 3.1. Schematic representation of a typical front-end circuit in particle physics experiments. The schematic includes the circuit models for the detector, the charge-sensitive amplifier, and the pulse-shaping filter.

step-by-step derivation of a mathematical expression of the ENC through design-oriented analysis.

3.2.1 Front-end circuit model

A typical front-end circuit in particle physics experiments includes a detector, a charge-sensitive amplifier (CSA), and a pulse-shaping filter on the analog-processing chain. Figure 3.1 shows a schematic representation of a typical front-end circuit. The detector is modeled as current signal source in parallel with a capacitance C_D. The charge-sensitive amplifier is modeled as a voltage amplifier of gain $A(j\omega)$, input capacitance C_{gg}, and feedback capacitor C_F. The pulse-shaping filter is modeled as a transfer function $H(j\omega)$.

The resistor R_{rst} shown in Figure 3.1 represents the reset element of the circuit, with the purpose of discharging the feedback capacitor. This reset element can be either a gate-controlled switch, for an instantaneous discharge; or a resistor, for a continuous-time discharge. In either case, the reset element is designed to have practically no effect during the relatively short duration of the output pulse, thus it can be neglected in the following analysis.

To further simplify the analysis, is it typically assumed that the amplifier has a very large voltage gain and infinite bandwidth. The first assumption is a desirable design trait, since having a large amplifier gain reduces the feedback error of the system. The second assumption is typically valid since, even though the amplifier's bandwidth is finite, it is usually fed to a lower bandwidth filter. In a properly designed system, both of these assumptions are reasonable.

There are mainly two noisy devices and three noise processes that interfere with the output measurements of a front-end circuit. The detector introduces shot-noise into the system, while the amplifier introduces both thermal and flicker noise. The detector's shot-noise and the amplifier's thermal noise both behave as white noise, meaning they have a constant power spectral density (PSD). The amplifier's flicker noise, on the other hand, has a PSD that behaves in a $1/f$ fashion, meaning that it has higher power at lower frequencies. The thermal noise and the flicker noise originate from different physical phenomena, so they are uncorrelated, and their powers can be added in quadrature.

Figure 3.2 shows the schematic of a typical front-end redrawn to include noise sources. The detector shot noise is modeled as a current noise source of PSD $\overline{I_D^2}$, while the amplifier noise is modeled as both voltage and current noise sources, with a PSD of $\overline{V_A^2}(j\omega)$ and $\overline{I_A^2}(j\omega)$, respectively.

From the circuit models presented in Figures 3.1 and 3.2, the ENC can be computed as the square root of the ratio between the total output noise power, and the output power due to a single electron of input charge in the absence of noise, as follows:

$$ENC \equiv \sqrt{\frac{\overline{V_{O,noise}^2}}{v_{O,electron}^2}} \tag{3.1}$$

3.2.2 Single-electron output power

To compute the denominator of the ENC expression shown in (3.1), it is necessary to compute the output power due to a single electron of input charge. Consider the circuit

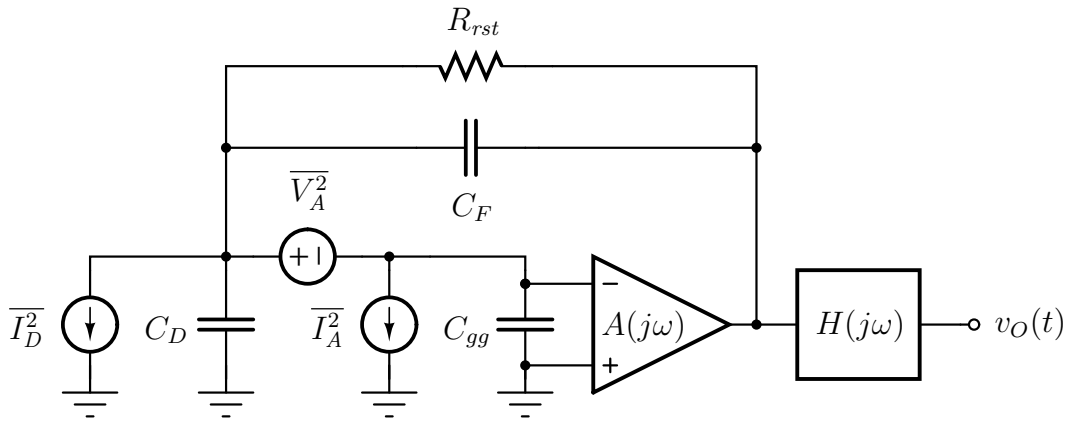


FIGURE 3.2. Schematic for noise analysis. Noise sources from two devices are considered: shot noise from the detector; and amplifier noise, represented as both voltage and current noise. The amplifier noise includes both thermal and flicker noise processes.

shown in Figure 3.1, with an input signal of

$$i_I(t) = q\delta(t) \tag{3.2}$$

where q is the electrical charge of a single electron, and $\delta(t)$ is the Dirac unit impulse.

The input signal shown in (3.2) is integrated by the CSA in the feedback capacitor, to produce an output voltage of value

$$v_{CSA}(t) = \frac{q}{C_F}u(t) \tag{3.3}$$

where $u(t)$ is the Heaviside unit step.

The output of the CSA shown in (3.3) is then processed by the pulse-shaping filter. Let $h(t)$ and $g(t)$ be the impulse response and the step response of the filter, respectively, related by definition by $g(t) = u(t) * h(t)$. The resulting expression at the output of the filter can be expressed as a function of $g(t)$, as follows

$$v_{O,electron}(t) = \frac{q}{C_F}g(t) \tag{3.4}$$

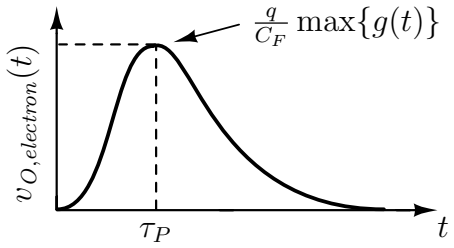


FIGURE 3.3. Graphical representation of the pulse shaper output for a single electron of input charge, considering an arbitrary pulse shape. The maximum value of the signal is reached at time $t = \tau_P$.

To maximize the SNR, the continuous-time output signal of the filter shown in (3.4) is sampled at the peak value of the pulse shape $g(t)$. The resulting time-independent expression can be written as

$$v_{O,electron} = \frac{q}{C_F}\max\{g(t)\}, \tag{3.5}$$

which represents the peak amplitude at the output of the front-end, for a single electron of input charge, in absence of noise sources. Figure 3.3 shows a graphical representation of the pulse shaper output for a single electron of input charge.

3.2.3 Output noise power

To compute the numerator of the ENC expression shown in (3.1), it is necessary to compute the total output noise power. To this end, let us consider the circuit shown in Figure 3.2. Since the detector noise and the amplifier noise originate from different physical phenomena, both are fully uncorrelated. Thus, detector and amplifier noise powers can be analyzed independently, and the results added by superposition to compute the total output noise power.

Let us consider the detector noise, modeled as a current noise source in Figure 3.2. Just like the input signal, the detector shot noise is fully integrated by the CSA in the feedback capacitor. Thus, the voltage noise power at the output of the CSA due to detector

shot noise can be computed as

$$\overline{V^2_{CSA,D}}(j\omega) = \frac{\overline{I^2_D}}{|j\omega C_F|^2} \tag{3.6}$$

The amplifier noise is modeled as two fully correlated noise sources in Figure 3.2. The input-referred voltage noise induces a noisy current due to capacitive coupling to ground via the input capacitor C_{gg}. The relation between the two is given by

$$\overline{I^2_A}(j\omega) = |j\omega C_{gg}|^2 \cdot \overline{V^2_A}(j\omega) \tag{3.7}$$

It can be shown (Sansen and Chang (1990)) that the voltage noise power at the output of the CSA due to amplifier input-referred noise is given by

$$\overline{V^2_{CSA,A}}(j\omega) = \left(\frac{C_{gg} + C_D + C_F}{C_F}\right)^2 \cdot \overline{V^2_A}(j\omega) \tag{3.8}$$

The results shown in (3.6) and (3.8) can be added by superposition to compute the noise power at the output of the CSA due to detector and amplifier noise, as follows

$$\overline{V^2_{CSA}}(j\omega) = \frac{\overline{I^2_D}}{|j\omega C_F|^2} + \left(\frac{C_{gg} + C_D + C_F}{C_F}\right)^2 \cdot \overline{V^2_A}(j\omega) \tag{3.9}$$

The output of the CSA is then processed by the filter, which is mathematically equivalent to multiplying (3.9) by $H(j\omega)$. Integrating the resulting expression over frequency gives the total output noise power of the front-end circuit due to detector and amplifier noise sources, as shown in the following expression:

$$\overline{V^2_{O,noise}} = \frac{1}{2\pi C_F^2} \int_0^\infty \left[\frac{\overline{I^2_D}}{|j\omega|^2} \cdot |H(j\omega)|^2 + (C_{gg} + C_D + C_F)^2 \cdot \overline{V^2_A}(j\omega) \cdot |H(j\omega)|^2\right] d\omega \tag{3.10}$$

Since the output of the CSA during normal operation is a voltage step, the pulse shaping filter is typically characterized by its step response $g(t)$, and its respective Fourier transform $G(j\omega)$. By definition, the relation between $H(j\omega)$ and $G(j\omega)$ is given by $H(j\omega) = j\omega \cdot G(j\omega)$. The expression shown in (3.10) can be rewritten as a function

of $G(j\omega)$ as follows

$$\overline{V_{O,noise}^2} = \frac{1}{2\pi C_F^2} \int_0^\infty \left[\overline{I_D^2} \cdot |G(j\omega)|^2 + (C_{gg} + C_D + C_F)^2 \cdot \overline{V_A^2}(j\omega) \cdot |j\omega \cdot G(j\omega)|^2 \right] d\omega \tag{3.11}$$

3.2.4 The ENC equation

To simplify notation, the total capacitance of the front-end will be referred to as

$$C_T = C_{gg} + C_D + C_F$$

Considering (3.1), (3.5) and (3.11), the ENC for the front-end circuit model presented in Section 3.2.1 can be computed as

$$ENC^2 = \frac{1}{2\pi} \frac{\int_0^\infty \left[\overline{I_D^2} \cdot |G(j\omega)|^2 + C_T^2 \cdot \overline{V_A^2}(j\omega) \cdot |j\omega \cdot G(j\omega)|^2 \right] d\omega}{q^2 |\max\{g(t)\}|^2} \tag{3.12}$$

Equation (3.12) shows that the detector noise $\overline{I_D^2}$, also known as parallel noise, can only be reduced by the action of the filter, since it has no interaction with any other component of the front-end circuit. On the other hand, the amplifier noise $\overline{V_A^2}(j\omega)$, also known as series noise, is affected by the total capacitance at the input node, the CSA design parameters, and the filter shape. The filter can effectively attenuate all noise contributions, but the extent of the attenuation and the interaction with both noise sources is not immediately apparent. A more insightful form of the ENC equation, useful for circuit design, is presented in the latter sections.

3.3 Filter design

3.3.1 Noise coefficients

The ENC expression shown in (3.12) can be simplified by separating the integral components by frequency dependence. The PSD of the detector's shot-noise is constant, so it can be removed from the integral. The amplifier's noise, comprised by thermal and flicker noise components, is frequency-dependent.

The PSD of the amplifier $\overline{V_A^2}(j\omega)$ can be decomposed into two independent contributions

$$\overline{V_A^2}(j\omega) = \overline{V_{A,W}^2} + \overline{V_{A,F}^2}(j\omega) \tag{3.13}$$

where $\overline{V_{A,W}^2}$ is the thermal noise component, which behaves as white noise in the sense that is frequency-independent, and $\overline{V_{A,F}^2}(j\omega)$ is the flicker, frequency-dependent component. The latter can be expressed as

$$\overline{V_{A,F}^2}(j\omega) = \left| \frac{K_F}{(j\omega)^{A_F}} \right| \tag{3.14}$$

where K_F is the flicker noise coefficient, and A_F is the flicker noise exponent.

With these considerations, it is possible to define the noise coefficients N_P, N_W and N_F, for parallel, white and flicker noise, respectively. The ENC expression shown in (3.12) can be rewritten as follows

$$ENC^2 = N_P \cdot \overline{I_D^2} + C_T^2 \left(N_W \cdot \overline{V_{A,W}^2} + N_F \cdot K_F \right) \tag{3.15}$$

where

$$N_P = \frac{\int_0^\infty |G(j\omega)|^2 \, d\omega}{2\pi q^2 \, |\max\{g(t)\}|^2} \tag{3.16}$$

$$N_W = \frac{\int_0^\infty |j\omega \cdot G(j\omega)|^2 \, d\omega}{2\pi q^2 \, |\max\{g(t)\}|^2} \tag{3.17}$$

$$N_F = \frac{\int_0^\infty \frac{|j\omega \cdot G(j\omega)|^2}{|j\omega|^{A_F}} d\omega}{2\pi q^2 \, |\max\{g(t)\}|^2} \tag{3.18}$$

3.3.2 Normalized noise coefficients

On a typical pulse-shaping filter, the pulse shape $g(t)$ has a clearly defined maximum value at τ_P, *i.e.* $\max\{g(t)\} = g(\tau_P)$, referred to as the peaking time. Let us consider a time-normalized version of the same function $g_n(t)$, such that the function peaks at $t = 1$. It is possible to write $g(t)$, and its Fourier transform, as a function of $g_n(t)$, as follows

$$g(t) = g_n(t/\tau_P) \quad \xrightarrow{\mathcal{F}\{\cdot\}} \quad G(j\omega) = \tau_P G_n(\tau_P \cdot j\omega) \tag{3.19}$$

Using this newly defined function and performing simple integral manipulation, the ENC integrals of the noise coefficients N_P, N_W and N_F can be rewritten in a more useful form

$$\begin{aligned}\int_0^\infty |G(j\omega)|^2\, d\omega &= \tau_P^2 \int_0^\infty |G_n(j\omega\cdot\tau_P)|^2\, d\omega \\ &= \tau_P \int_0^\infty |G_n(ju)|^2\, du \end{aligned} \tag{3.20}$$

$$\begin{aligned}\int_0^\infty |j\omega\cdot G(j\omega)|^2\, d\omega &= \tau_P^2 \int_0^\infty |j\omega\cdot G_n(j\omega\cdot\tau_P)|^2\, d\omega \\ &= \frac{1}{\tau_P} \int_0^\infty |ju\cdot G_n(ju)|^2\, du \end{aligned} \tag{3.21}$$

$$\begin{aligned}\int_0^\infty \frac{|j\omega\cdot G(j\omega)|^2}{|j\omega|^{A_F}} d\omega &= \tau_P^2 \int_0^\infty \frac{|j\omega\cdot G_n(j\omega\cdot\tau_P)|^2}{|j\omega|^{A_F}} d\omega \\ &= \tau_P^{A_F-1} \int_0^\infty \frac{|ju\cdot G_n(ju)|^2}{|ju|^{A_F}} du \end{aligned} \tag{3.22}$$

With these considerations, it is possible to define the normalized noise coefficients N_{Pn}, N_{Wn} and N_{Fn}, for parallel, white and flicker noise, respectively. Using (3.20), (3.21) and (3.22), the ENC can be rewritten as a function of $g_n(t)$, as follows

$$ENC^2 = \tau_P \cdot N_{Pn} \cdot \overline{I_D^2} + C_T^2 \left(\frac{N_{Wn}}{\tau_P} \cdot \overline{V_{A,W}^2} + \tau_P^{A_F-1} \cdot N_{Fn} \cdot K_F \right) \tag{3.23}$$

where

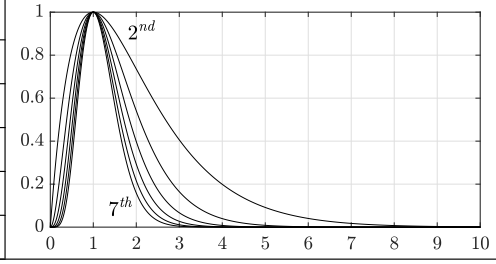
$$N_{Pn} = \frac{\int_0^\infty |G_n(j\omega)|^2\, d\omega}{2\pi q^2\, |\max\{g_n(t)\}|^2} \tag{3.24}$$

$$N_{Wn} = \frac{\int_0^\infty |j\omega\cdot G_n(j\omega)|^2\, d\omega}{2\pi q^2\, |\max\{g_n(t)\}|^2} \tag{3.25}$$

$$N_{Fn} = \frac{\int_0^\infty \frac{|j\omega\cdot G_n(j\omega)|^2}{|j\omega|^{A_F}} d\omega}{2\pi q^2\, |\max\{g_n(t)\}|^2} \tag{3.26}$$

The normalized noise coefficients N_{Pn} and N_{Wn} are only dependent on pulse shape, independent of time scale. The normalized noise coefficient N_{Fn} is dependent on both

TABLE 3.1. Some commonly adopted time-invariant filters and the corresponding ENC noise coefficients. The ratio between the pulse width (τ_W) and the peaking time (τ_P) is also reported. Flicker exponent $A_F = 1$ was considered (De Geronimo and O'Connor (2005)).

Filter	Shape	$q^2 N_{Wn}$	$q^2 N_{Fn}$	$q^2 N_{Pn}$	τ_W/τ_P
Triang.		1	0.44	0.33	2
RU-2		0.92	0.59	0.92	7.66
RU-3		0.82	0.54	0.66	5.04
RU-4		0.85	0.53	0.57	4.17
RU-5		0.89	0.52	0.52	3.73
RU-6		0.92	0.52	0.48	3.46
RU-7		0.94	0.51	0.46	3.27

the pulse shape and the flicker exponent A_F. For design purposes, it is typically assumed that $A_F = 1$, while in reality A_F can be in the range of $0.5 - 2$ (Spieler (2005)). For a given normalized step response $g_n(t)$, the integrals in (3.24), (3.25) and (3.26) can be conveniently integrated and tabulated. Table 3.1 shows an example of tabulated noise coefficients for RU-type filters, that is, filters with real coincident poles and unipolar pulse shape (De Geronimo and O'Connor (2005)).

It is also possible to express the noise coefficients in the time-domain response of the filter $g_n(t)$. Using Parseval's theorem and simple Fourier analysis, it can be shown that:

$$N_{Pn} = \frac{\int_{-\infty}^{\infty} |g_n(t)|^2 \, dt}{q^2 \, |\max\{g_n(t)\}|^2} \tag{3.27}$$

$$N_{Wn} = \frac{\int_{-\infty}^{\infty} |g_n'(t)|^2 \, dt}{q^2 \, |\max\{g_n(t)\}|^2} \tag{3.28}$$

$$N_{Fn} = \frac{\int_{-\infty}^{\infty} \left|g_n^{(\alpha)}(t)\right|^2 \, dt}{q^2 \, |\max\{g_n(t)\}|^2} \tag{3.29}$$

where $g_n^{(\alpha)}(t)$ is the derivative of order α of $g_n(t)$, and $\alpha = 1 - \frac{A_F}{2}$. For $A_F = 1$, N_{Fn} is a function of the half derivative of the filter step response.

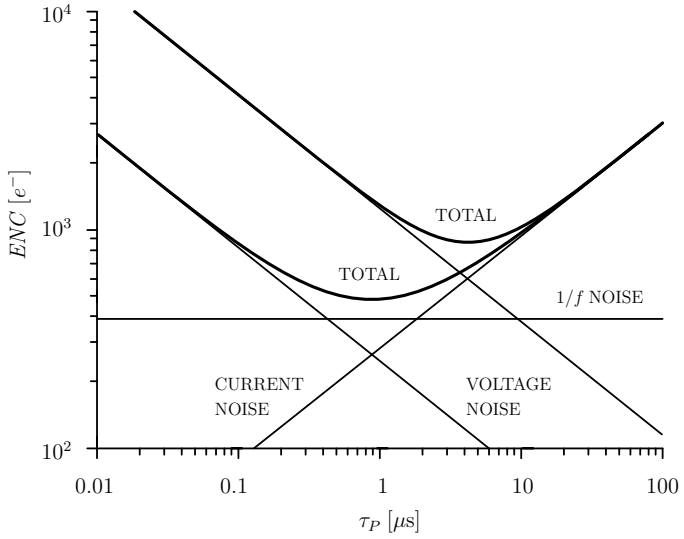


FIGURE 3.4. Equivalent noise charge $vs.$ shaping time. The plot shows two equivalent ENC curves to illustrate the effect of the $1/f$ noise over the voltage noise. (Spieler (2005))

The parallel white noise contribution is directly proportional to the filter time constant, while the series white noise contribution is inversely proportional to the filter time constant. As for the series low-frequency noise, for flicker exponents near unity, there is very little to no dependency on the time constant of the filter. Figure 3.4 shows the dependency of different noise components of the ENC on the filter time constant.

The ENC expression shown in (3.23) and the time and frequency domain representation of the filter coefficients offer a simple interpretation of the dependency of the noise contributions with the time scale. Larger values of the filter time constant imply a longer lasting pulse shape, which in turn increases the lingering effect of the noise charge integrated in the feedback capacitor due to parallel noise, and the cumulative effect is an increase on the output noise. Conversely, larger derivatives in the pulse shape imply a larger pulse bandwidth, thus the series white noise contribution increases with decreasing values of the filter time constant.

Without any system-level constraints, the minimum ENC is achieved by matching the parallel and series noise contributions, as shown in Figure 3.4. Optimizing the system

TABLE 3.2. Normalized noise equations (Alvarez et al. (2012)).

	$\overline{V_n^2}/\Delta f$	$\overline{\hat{V}_n^2}/\Delta f$
Thermal noise for strong inversion	$\frac{4k_BT\gamma}{g_m}$	$4k_BT\gamma\left(\frac{I_D}{g_m}\right)$
Shot noise for weak inversion	$\frac{2qI_D}{g_m^2}$	$2q\left(\frac{I_D}{g_m}\right)^2$
HSPICE $1/f$ noise (`NLEV=0`)	$\frac{K_F I_D^{A_F}}{g_m^2 C_{ox} L^2 f}$	$\frac{K_F I_D^{A_F-1}}{C_{ox} L^2 f}\left(\frac{I_D}{g_m}\right)^2$
HSPICE $1/f$ noise (`NLEV=2,3`)	$\frac{K_F}{C_{ox} W L f^{A_F}}$	$\frac{K_F}{C_{ox} L f^{A_F}}\left(\frac{I_D}{W}\right)$
BSIM3 thermal noise	$\frac{4k_BT}{g_m^2 R_{DS}+L^2 g_m^2/(\mu\lvert Q_{inv}\rvert)}$	$\frac{4k_BT}{g_m R_{DS}+L^2 g_m/(\mu\lvert Q_{inv}\rvert)}\left(\frac{I_D}{g_m}\right)$
BSIM3 $1/f$ noise for strong inversion	$\frac{k_BT}{L^2 f^{E_F}}\left(\frac{q^2\mu I_D W_A}{g_m^2 C_{ox}}+\frac{I_D^2 \Delta L_{clm} W_B}{g_m^2 qW}\right)$	$\frac{k_BT}{L^2 f^{E_F}}\left(\frac{q^2\mu W_A}{C_{ox}}+\frac{\Delta L_{clm} W_B}{q}\left(\frac{I_D}{W}\right)\right)\left(\frac{I_D}{g_m}\right)^2$

solely based on minimum noise would mean arbitrarily small or large values for the peaking time τ_P, specially if one contribution is far more dominant than the other. More often than not, the maximum allowed integration time is set by system-level specifications, such as collision frequency.

3.4 Charge-sensitive amplifier design

3.4.1 The g_m/I_D methodology and noise analysis

Traditional square-law MOSFET models have long since become obsolete for the accurate description transistor behavior in submicron technologies (Razavi (2001)). More complex models, impractical for hand analysis, are necessary to accurately describe device behavior. Circuit simulators such as SPICE use both complex device models and empirical data to provide very accurate results, excellent for numerical analysis. However, these tools do not provide the same insight into circuit operation as simpler models do.

The g_m/I_D design methodology (Silveira et al. (1996), Flandre et al. (1997)) overcomes this limitation by using accurate SPICE simulation results as data for hand analysis. This method was conceived without the inclusion of transistor noise models, but has since been extended for noise analysis (Alvarez and Abusleme (2012), Alvarez et al. (2012)).

MOSFET device noise can be written as a function of g_m/I_D simply by dividing the drain current noise PSD $\overline{I_n^2}$ by the drain current I_D, or by multiplying the transistor gate voltage noise PSD $\overline{V_n^2}$ by the drain current I_D. Table 3.2 presents some examples of noise

equations and their normalized versions. Although not immediately apparent, it can be shown that the noise equations presented in Table 3.2 are only dependent on g_m/I_D, at least in a first-order analysis (Alvarez et al. (2012)).

As an example, Figure 3.5 shows the normalized gate voltage noise power $\overline{\hat{V}_n^2}$ for different values of g_m/I_D, for a minimum-length NMOS transistor on a 0.5-$\mu\mathrm{m}$ CMOS technology, obtained through SPICE simulations.[1] These normalized noise values can be converted back to noise power simply by dividing the normalized noise by the transistor drain current.

Two distinct regions can be identified from the noise curves shown in Figure 3.5: a low-frequency region, dominated by flicker noise; and a high-frequency region, dominated by white noise. The frequency at which the flicker and white contributions are matched is known as the corner frequency f_c. Smaller values of g_m/I_D imply a higher current density I_D/W, and thus smaller value for W for the same current value. Consequently, smaller values of g_m/I_D imply higher values for the corner frequency f_c due to the reduced gate-area of the transistor, so the flicker noise dominance extends to higher frequencies.

For the noise models used for the generation of Figure 3.5, the overall level of the normalized noise spectra decreases with increasing values of g_m/I_D, and is minimum for weak inversion operation. However, this does not mean that the ENC is necessarily improved for increasing values of g_m/I_D, as it will be shown in the following sections.

3.4.2 Charge amplifier noise using the g_m/I_D methodology

In traditional noise analysis for particle physics instrumentation systems it is typically assumed that the charge-sensitive amplifier noise is dominated by the input transistor noise. Typically, noise minimization is achieved by following a simple recipe for the design parameters of the input device: maximum available current, capacitance matching at the input node (O'Connor and De Geronimo (1999)), and minimum-length (Sansen

[1] Unfortunately, the models provided by the manufacturer for this technology did not include flicker parameters. Thus, the HSPICE flicker coefficient (see Table 3.2) was arbitrarily selected to favor the clarity of the analysis.

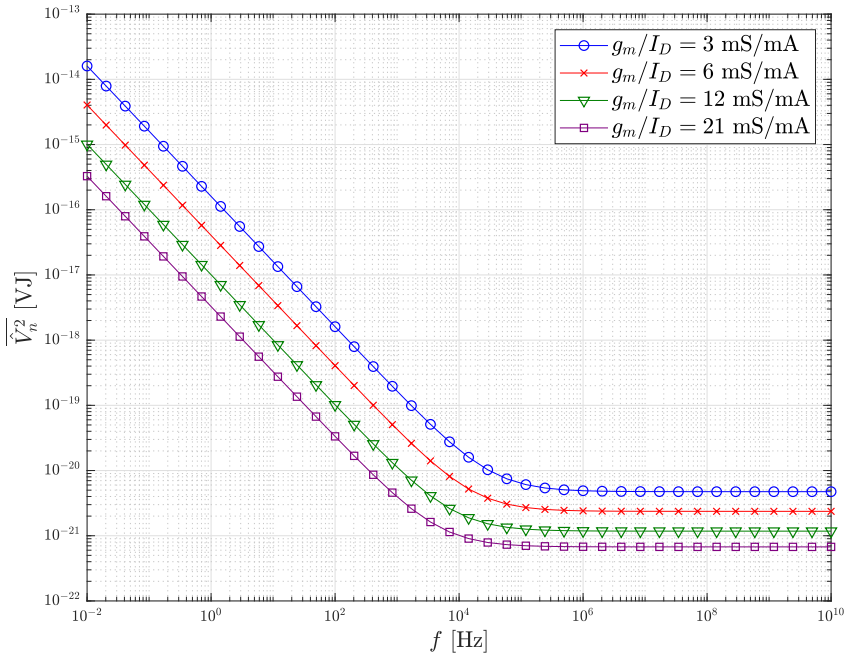


FIGURE 3.5. Example of the normalized noise spectra for a single transistor including white and flicker noise, for different values of g_m/I_D. Relevant simulation parameters are $L = 0.6$ μm and HSPICE NMOS parameter $K_F = 10^{-30}$.

and Chang (1990), Radeka (1984)). These guidelines, obtained through the use of simple transistor noise models and neglecting flicker noise, have proven to produce sub-optimal results through the use of more adequate device noise models for current technologies (De Geronimo and O'Connor (2005)).

An analysis of the noise in charge-sensitive amplifiers using a g_m/I_D approach was first performed in (Alvarez et al. (2012)). The current section presents some relevant conclusions obtained from this analysis, and further expands on these results in an attempt to gain insight into the design trade-offs of the CSA from a g_m/I_D perspective. Since the g_m/I_D methodology uses pre-computed curves, it is technology dependent. The following analysis considers curves obtained through SPICE simulations in a 0.5-μm CMOS technology, using arbitrarily selected HSPICE flicker parameters. It can be extended an applied to different technologies, but the curves could behave differently.

Without compromising the validity of the following analysis, only the CSA input device noise contribution will be considered. It will be assumed that the input transistor length L is set, so it is not a design variable. It can be shown that, although transistor noise is dependent on the length L, the dependence is not significant (Alvarez et al. (2012)). The analysis can be extended to include the noise contribution of more devices by referring their noise to the amplifier input node, and tabulating the results.

Consider that $\overline{\hat{V}_{A,W}^2}$ and $\hat{K}_F$ are the normalized white noise PSD and normalized flicker coefficient, respectively, defined as $\overline{\hat{V}_{A,W}^2} = \overline{V_{A,W}^2} \cdot I_D$ and $\hat{K}_F = K_F \cdot I_D$. Using this notation, the ENC expression shown in (3.23) can be rewritten as follows

$$ENC^2 = \tau_P \cdot N_{Pn} \cdot \overline{I_D^2} + \frac{(C_{gg} + C_K)^2}{I_D} \left(\frac{N_{Wn}}{\tau_P} \cdot \overline{\hat{V}_{A,W}^2} + N_{Fn} \cdot \hat{K}_F \right) \tag{3.30}$$

where $C_K = C_D + C_F$, the g_m/I_D-independent component of the input node capacitance.

Let us assume that the filter parameters are already set by system-level constraints, *e.g.*, by the maximum allowed integration time. In this scenario, the optimization of the ENC is solely dependent on minimizing the CSA contribution, as the detector contribution depends only on the pulse shaper parameters. Minimizing the ENC under this conditions is equivalent to minimizing the following objective function

$$F_o = \frac{\left(C_{gg}(x, I_D) + C_K\right)^2}{I_D} \left(\frac{N_{Wn}}{\tau_P} \cdot \overline{\hat{V}_{A,W}^2}(x) + N_{Fn} \cdot \hat{K}_F(x) \right) \tag{3.31}$$

where the dependences for transistor-dependent parameters have been made explicit as function arguments, and g_m/I_D has been written as x to simplify notation.

3.4.3 Capacitance matching for constant g_m/I_D

There are three input device variables that have an effect on the objective function: the transistor length L, the g_m/I_D value, and the drain current I_D. To gain insight into the design space of this objective function, let us consider the design corner in which the values of L and g_m/I_D are already set. For fixed values of L, g_m/I_D and filter parameters, the rightmost factor becomes constant, and thus minimizing F_o is equivalent to minimizing

the function

$$\begin{aligned} F_{o2} &= \frac{\left(C_{gg}(x, I_D) + C_K\right)^2}{I_D} \\ &= \frac{C_{gg}}{I_D}(x) \cdot C_{gg}(x, I_D) \left(1 + \frac{C_K}{C_{gg}(x, I_D)}\right)^2 \end{aligned} \tag{3.32}$$

Since C_{gg}/I_D is solely a function of g_m/I_D, it is constant under these assumptions. It is clear that the objective function F_{o2} is minimum for the condition $C_{gg} = C_K$, and thus the optimal current value is that for which this condition holds.

Consider Figure 3.6 as an example. This plot shows F_o as a function of the transistor drain current I_D for different values of g_m/I_D, and fixed values of L and filter parameters. For any given value of g_m/I_D, the minimum noise is achieved when the capacitance matching condition holds. A different value of g_m/I_D, however, implies a different current for minimum noise. Conversely, for a constant value of I_D, the capacitance matching condition no longer holds for the g_m/I_D value for which the total noise is minimized, with the exception of the global minimum.

3.4.4 Flicker noise and corner frequency

The corner frequency ω_c is defined as the frequency at which the flicker and white noise contributions are matched. In mathematical terms, this relation can be expressed as

$$\overline{\hat{V}^2_{A,W}}(x) = \frac{\hat{K}_F(x)}{\omega_c^{A_F}(x)} \tag{3.33}$$

The objective function (3.31) can be written in a more insightful form as a function of the corner frequency, as follows

$$F_o = \frac{\left(C_{gg}(x, I_D) + C_K\right)^2}{I_D} \cdot \left(N_{Wn} \cdot \omega_P + N_{Fn} \cdot \omega_c^{A_F}(x)\right) \cdot \overline{\hat{V}^2_{A,W}}(x) \tag{3.34}$$

where $\omega_P = 1/\tau_P$, a frequency value related to the peaking time. In this shape, two terms with frequency units, $N_{Wn}\omega_P$ and $N_{Fn}\omega_c^{A_F}$ are shown in addition. This allows for a direct magnitude comparison between white and flicker noise, dependent on the system speed constraints.

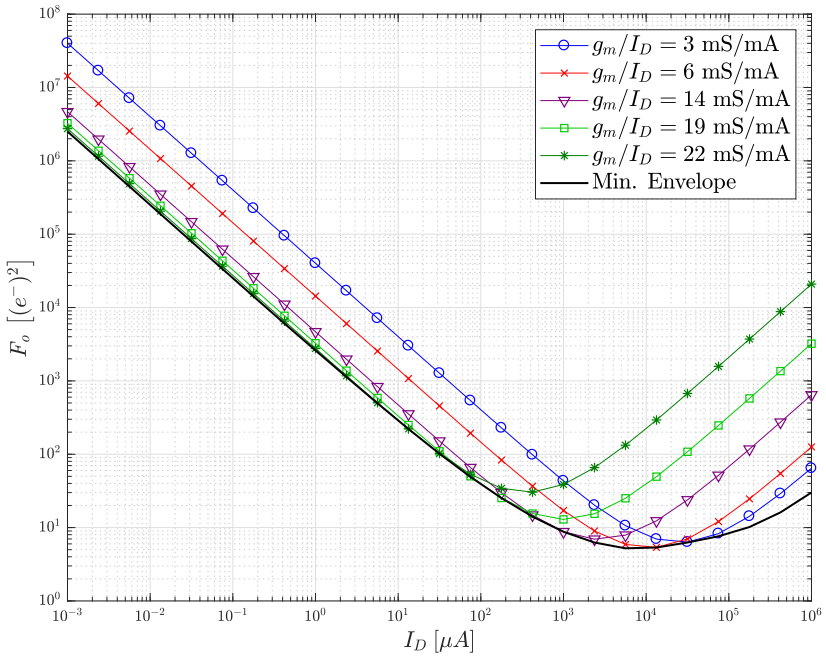


FIGURE 3.6. F_o as a function of I_D. Several values of g_m/I_D are plotted. A global minimum envelope curve, obtained numerically using g_m/I_D values ranging from 1 to 24 is also included. Both white and flicker noise sources are considered. Relevant simulation parameters include $L = 0.6\ \mu$m and MOSFET SPICE model $K_F = 10^{-30}$. Other parameters include the use of a RU-2 filter, $\tau_P = 10\,\mu$s and $C_K = 1$ pF.

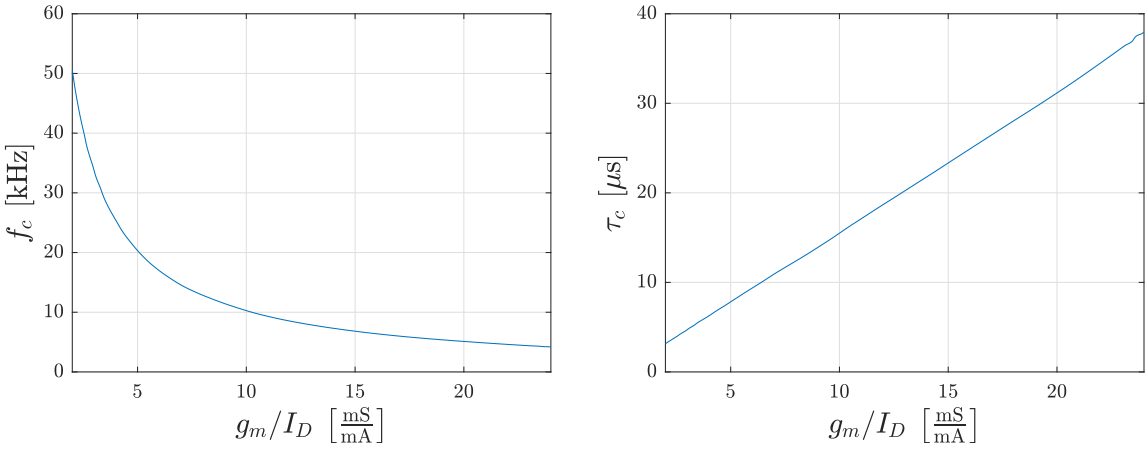
Alternatively, the effects of the flicker noise can be written as a multiplicative effect in the objective function (3.31), as follows

$$F_o = \frac{\left(C_{gg}(x, I_D) + C_K\right)^2}{I_D} \left(\Phi(x) \cdot \frac{N_{Wn}}{\tau_P} \cdot \overline{\hat{V}^2_{A,W}}(x) \right) \tag{3.35}$$

where

$$\Phi(x) = 1 + \frac{\frac{\tau_P}{N_{Wn}}}{\frac{\tau_c^{A_F}(x)}{N_{Fn}}} \tag{3.36}$$

is a dimensionless factor, and $\tau_c = 1/\omega_c$, a time constant related to the corner frequency. This form again allows for easy magnitude comparisons between τ_P/N_{Wn} and τ_c/N_{Fn}, to assess noise dominance. The term $\Phi(x)$ has a minimum value of unity, for white noise dominant systems, and has no defined upper bound. This factor is a measure of the relative effect of flicker noise with respect to white noise, for a given value of g_m/I_D. As an



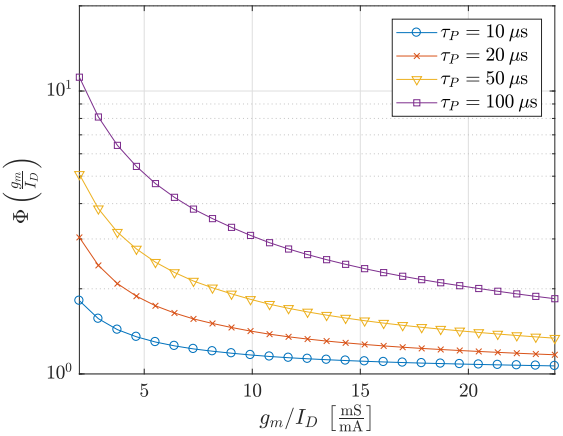
(A) Corner frequency.

(B) Corner frequency reciprocal.

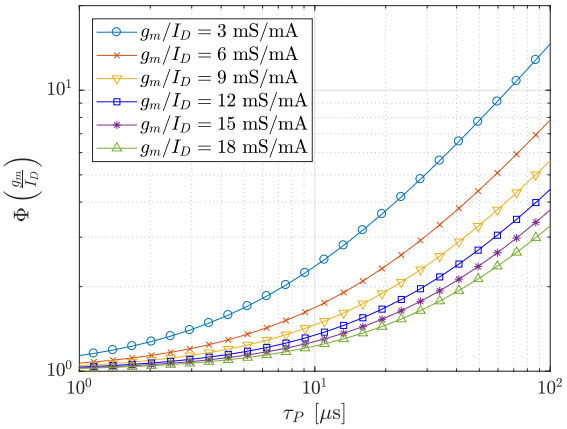
FIGURE 3.7. Corner frequency and the reciprocal time constant for a flicker exponent of $A_F = 1$ in a 0.5-μm technology.

example, a value of $\Phi(x) = 1.2$, means that the flicker noise total power is 20% that of white noise on the ENC, and accounts for 16.6% of the total output noise power.

The values of $\omega_c(x)$ and $\tau_c(x)$ can be tabulated as functions of g_m/I_D for a given technology, as shown in Figure 3.7. These curves allow the designer to perform easy hand calculation to assess early the effects of flicker noise on system noise performance. Typical values of N_{Wn}/N_{Fn} range between 1.3 and 2.3 (De Geronimo and O'Connor (2005)), so an intermediate value can be used as a rule of thumb even before the shaper is selected. As an example, consider a system with a peaking time of 100 ns and $N_{Wn}/N_{Fn} = 2$. For the worst-case scenario, that is, for low values of g_m/I_D, where $\tau_c \approx 5\ \mu$s, it can be computed that $\Phi = 1.01$, *i.e.* flicker noise accounts for approximately 1% of the total output noise, so it can be neglected from the analysis. Furthermore, the relative effects of flicker noise on the total output noise power can be directly tabulated as $\Phi(x)$ for a set of filter parameters, as shown in Figure 3.8. These curves allow the designer to better explore the design space of this technology when flicker noise becomes relevant. For relatively slower systems, with peaking time values ranging from 1μs to 10μs, it can be seen from Figure 3.8(B) that



(A) $\Phi(x)$ *vs.* g_m/I_D for different values of τ_P.



(B) $\Phi(x)$ *vs.* τ_P for different values of g_m/I_D.

FIGURE 3.8. Relative effect of flicker noise on the total output noise. Flicker exponent $A_F = 1$ and RU-2 filter noise coefficients were considered.

for the majority of the g_m/I_D design space, $\Phi(x) < 1.1$, which translates to approximately 9% of the total output noise power being due to flicker noise.

3.4.5 Minimum noise versus g_m/I_D and peaking time

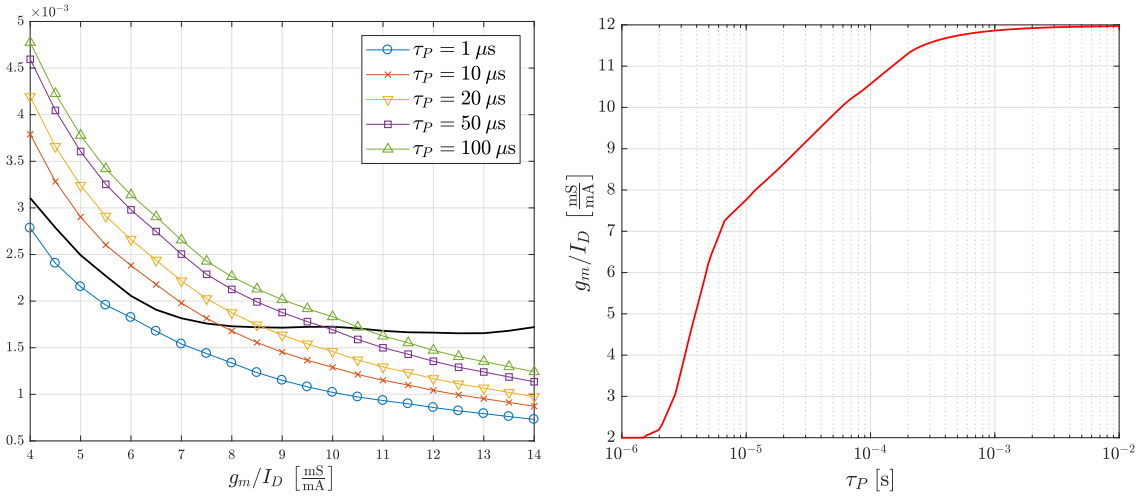
Let us consider Figure 3.6. The envelope curve is traced numerically by plotting the minimum value for each point of current for a large set of g_m/I_D values. All the g_m/I_D curves have a parabolic shape. It follows, by continuity, that the absolute minimum of the envelope curve corresponds to the minimum of a given g_m/I_D curve, and as such, it corresponds to the point where the capacitance matching condition holds.

Let us consider that the capacitance matching condition is met. From (3.31) and (3.32), and considering $C_{gg} = C_K$, the objective function can be written as follows

$$F_o = 4C_K \cdot \left(\frac{C_{gg}}{I_D}(x) \cdot N(x, \tau_P) \right) \tag{3.37}$$

where

$$N(x, \tau_P) = \left(\frac{N_{Wn}}{\tau_P} \cdot \overline{\hat{V}^2_{A,W}}(x) + N_{Fn} \cdot \hat{K}_F(x) \right) \tag{3.38}$$



(A) $-N'(x,\tau_P)/N(x,\tau_P)$ for different values of τ_P, and $\frac{C_{gg}}{I_D}'(x)/\frac{C_{gg}}{I_D}(x)$ *vs.* g_m/I_D.

(B) Optimum g_m/I_D *vs.* τ_P.

FIGURE 3.9. Value of g_m/I_D for overall minimum noise when the capacitance matching condition is met.

The filter coefficients were considered to be constant, so that the noise factor in (3.37) is only dependent on g_m/I_D and τ_P. The condition for the objective function (3.37) minimum can be written as:

$$\frac{\frac{C_{gg}}{I_D}'(x)}{\frac{C_{gg}}{I_D}(x)} = -\frac{N'(x,\tau_P)}{N(x,\tau_P)} \tag{3.39}$$

The optimum g_m/I_D for a given problem is only dependent on system speed. The left and right terms of the optimum condition shown in (3.39) are plotted in Figure 3.9(A). The intersection of the two curves corresponds to the optimum g_m/I_D for a given peaking time τ_P. Furthermore, the optimum g_m/I_D can be plotted directly as a function of τ_P for a given set of filter parameters, as shown in Figure 3.9(B). For smaller values of τ_P than the ones plotted in Figure 3.9(B) there is no clearly defined optimum g_m/I_D value, so the plot is saturated at an arbitrarily selected practical limit of $g_m/I_D = 2\ \frac{\text{mS}}{\text{mA}}$. For larger values of τ_P, the overall noise becomes fully dominated by flicker noise ($N_{Fn}\omega_c \gg N_{Wn}\omega_P$), so the optimum g_m/I_D becomes independent of the system peaking time τ_P, and converges to a value of $g_m/I_D \approx 12\ \frac{\text{mS}}{\text{mA}}$ for this particular technology.

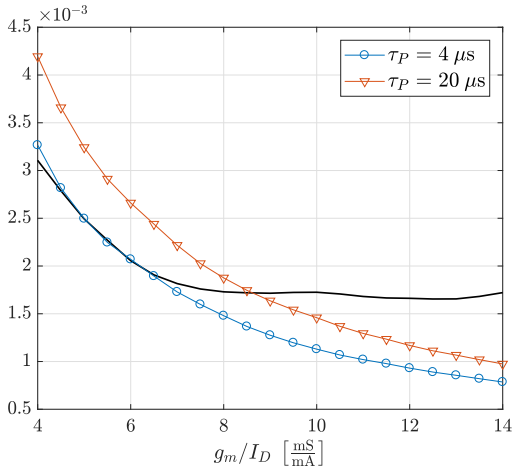


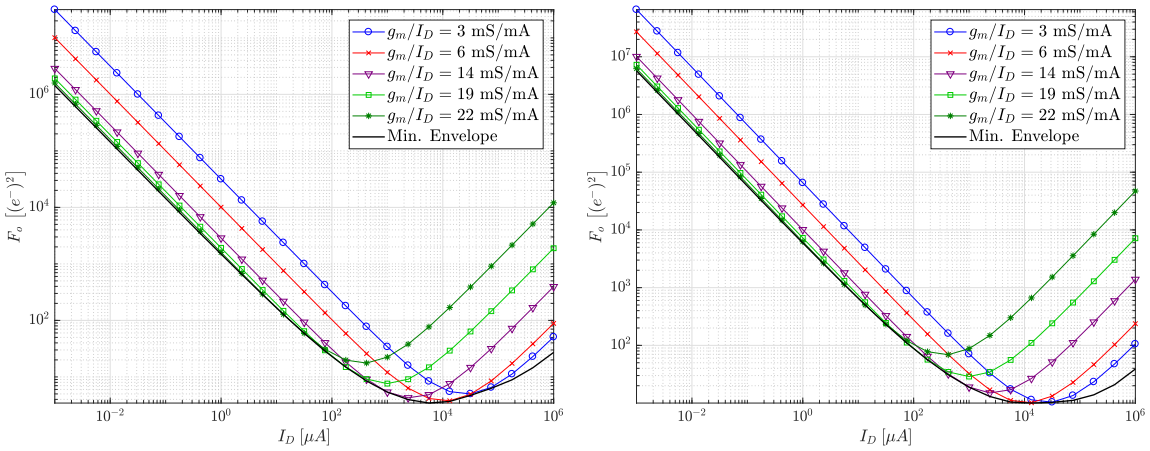
FIGURE 3.10. Optimum condition for peaking time values of $\tau_P = 4\ \mu$s and $\tau_P = 20\ \mu$s.

The large slope at the leftmost side in Figure 3.9(B) can be interpreted as a relative insensitivity to g_m/I_D, that is, values around the optimum produce similar results in terms of minimum noise. To exemplify this, consider values of $\tau_P = 4\mu$s and $\tau_P = 20\mu$s. Figure 3.10 shows the two curves that must intersect in order to reach the optimum condition for said peaking time values, and it can be observed that the two intersecting curves for $\tau_P = 4\mu$s seem to lie on top of each other on a small range of g_m/I_D values. The objective function F_o is plotted for different values of g_m/I_D on Figure 3.11 together with the global minimum envelope, for both $\tau_P = 4\ \mu$s and $\tau_P = 20\ \mu$s. It can be observed that the plot is relatively flat near the minimum noise point for $\tau_P = 4\ \mu$s when compared to the same curve for $\tau_P = 20\ \mu$s, which is consistent with the aforementioned observations.

3.4.6 White-noise dominance – practical design guidelines

The observations presented in the previous sections, although useful to better understand the g_m/I_D design space in terms of minimum noise conditions, are not necessarily practical when other design constraints are considered. For example, a large noise budget[2] might allow one to use a relatively small current compared to the optimum current for minimum noise, so any excess in current consumption might be seen as wasted power. As

[2]Noise budget: Specification of the maximum allowable noise in a system.



(A) $\tau_P = 20\ \mu$s. (B) $\tau_P = 4\ \mu$s.

FIGURE 3.11. F_o as a function of I_D. The same parameters as in Figure 3.6 were considered, with the exception of τ_P.

it was mentioned in Section 3.4.4, systems with a shorter peaking time are dominated by white noise. Additionally, when there is headroom on the noise budget, power or die area optimization might become more relevant. The current section explores the design space of white-noise dominated systems for small currents relative to the optimum current for minimum noise from a practical design perspective.

Let us consider the objective function (3.31). For white-noise dominated system, $i.e.$ $N_{Wn}\omega_P \gg N_{Fn}\omega_c$, and for small current values relative to the optimum condition, where $C_K \gg C_{gg}$, the objective function can be written as follows:

$$F_o = \frac{C_K^2}{I_D}\left(\frac{N_{Wn}}{\tau_P}\cdot\overline{\hat{V}_{A,W}^2}(x)\right) \tag{3.40}$$

In this shape (3.40) is a function of I_D, g_m/I_D, and the filter parameters. Overall noise is reduced by increasing the input-device current I_D, while normalized white noise is reduced by using higher g_m/I_D values. This means that, when the initial assumptions are met ($N_{Wn}\omega_P \gg N_{Fn}\omega_c$, $C_K \gg C_{gg}$) noise is monotonically reduced by using higher currents and larger values of g_m/I_D.

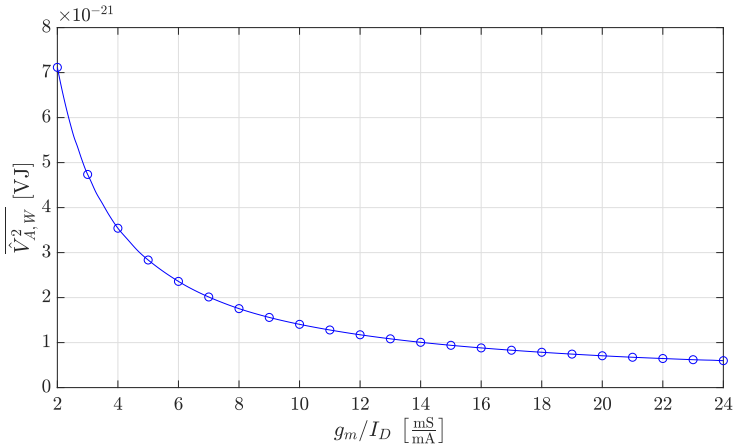


FIGURE 3.12. Normalized amplifier white noise PSD *vs.* g_m/I_D. For strong inversion operation white noise is $\propto (g_m/I_D)^{-1}$ which is directly observable on the plot. This inverse relation with g_m/I_D translates into diminishing returns in noise performance with increasing values of g_m/I_D.

Figure 3.12 shows a plot of $\overline{\hat{V}^2_{A,W}}$ as a function of g_m/I_D. As shown in Table 3.2, white noise for strong inversion operation, dominated by thermal noise, is inversely proportional to g_m/I_D. This inverse proportionality translates into diminishing returns in terms of noise performance for higher values of g_m/I_D. This means that, on a practical level, a value of $g_m/I_D = 14\ \frac{\text{mS}}{\text{mA}}$ and $g_m/I_D = 22\ \frac{\text{mS}}{\text{mA}}$ do not produce drastically different results. The same cannot be said for current density. Figure 3.13 shows the width per unit current W/I_D as a function of g_m/I_D. The plot reveals that, while $g_m/I_D = 14\ \frac{\text{mS}}{\text{mA}}$ and $g_m/I_D = 22\ \frac{\text{mS}}{\text{mA}}$ might produce similar and acceptable results in term of noise performance, there is approximately a $\times 10$ difference in terms of device width. This increase in device width for increasing values of g_m/I_D has two main drawbacks. First, it increases the die area used by the charge-sensitive amplifier. Second, it implies a higher capacitance per unit current, so a lower current value is needed to reach the capacitance matching condition. A larger headroom allows a given slice to have a wider applicability, as it can continue to scale for higher currents until it reaches the optimum.

An acceptable middle ground between scalability, die area, and noise performance lies around the middle values of the g_m/I_D design space. Values ranging from $14\ \frac{\text{mS}}{\text{mA}}$ to $16\ \frac{\text{mS}}{\text{mA}}$ have similar white noise performance in the available design space, while also

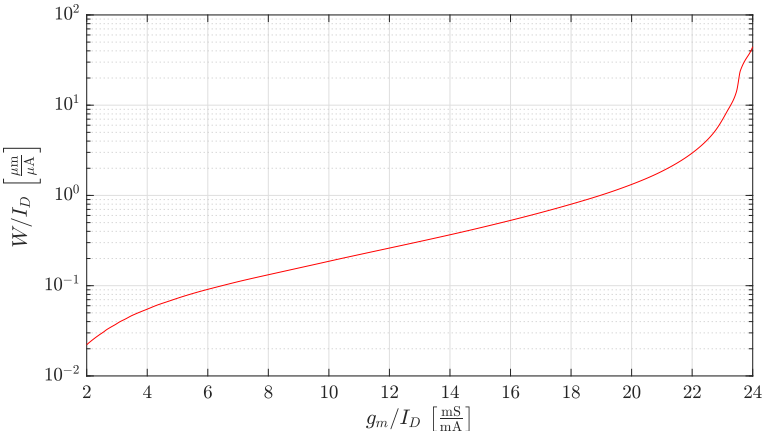


FIGURE 3.13. Width per unit current W/I_D, reciprocal of current density I_D/W.

having relatively smaller device sizes. This is where the design criteria becomes arbitrary and dependent on the priorities of the circuit designer.

3.4.7 Noise analysis in slice-based design

One of the questions that this thesis attempts to answer is whether it is possible to design a single charge-sensitive amplifier slice applicable to wide set of design constraints. From the g_m/I_D design space analysis presented in the previous sections, some observations can be made. First, it is not possible to design an optimum CSA cell independently of problem specifications. This is particularly true when secondary design constraints become relevant, such as power consumption and size optimization. Assuming that noise performance is paramount, and even assuming that current is a variable that can be scaled freely, it is still not possible to design an optimum slice to achieve minimum noise for every scenario, since the optimum g_m/I_D is dependent on the system peaking time.

Let us consider the specific scenario of the 0.5-μm CMOS technology used as an example in the previous sections. If the design goal is the most widely applicable single amplifier cell, in terms of speed constraints and size optimization, with adequate white performance near the minimum limits of the available design space of the technology, then a g_m/I_D near 12 $\frac{\text{mS}}{\text{mA}}$ could be a reasonable candidate. This value is the optimum for minimum noise in flicker noise dominant systems, as it can be seen from Figure 3.9(B);

produces a low white noise contribution when compared to the rest of the g_m/I_D design space in white noise dominant systems, as it can be seen from Figure 3.12; and it is on the lower end in terms of device size compared to the rest of the g_m/I_D design space, as shown in Figure 3.13. This particular observation cannot be generalized as the analysis is technology-specific. It is possible that for a different technology, the optimum g_m/I_D value for minimum flicker noise produces unacceptable results on other design dimensions.

Multiple charge-sensitive amplifier cells can be designed and compiled in a library, each of them designed to cover a narrower set of design constraints, in order to achieve close to optimum noise performance in a wider set of scenarios. Some examples include: a cell for white noise dominated systems, flicker noise dominated systems, with a good noise performance per area ratio, *etc*. A multitude of designs can be pre-made, in a process that needs to be performed only once for a given technology.

For a given pre-designed slice library, and considering that current scaling is done in integer multiples of the single-slice current, two simple guidelines guarantee the achievement of minimum noise for a given set of problem specifications:

1. Identify the optimum g_m/I_D for a given τ_P. This can be done using a precomputed table or plot, similar to the one shown in Figure 3.9(B). Select the amplifier slice with the closest value of g_m/I_D available that is greater or equal to the optimum value.

$$\min\{g_m/I_D\} \quad \text{s.t.} \quad g_m/I_D \geq (g_m/I_D)_{\text{Opt}} \tag{3.41}$$

2. Scale the amplifier current by connecting N slices in parallel. Select the current closest to the capacitance matching condition, while being smaller or equal. This can be done using g_m/I_D tables for C_{gg}/I_D.

$$\max\{I_D\} \quad \text{s.t.} \quad C_{gg}(I_D) \leq C_K \tag{3.42}$$

4. EFFECTS OF DEVICE MISMATCH ON SLICE-BASED DESIGN

4.1 Introduction

Modern circuit simulators use both complex device models and empirical data to produce very accurate results. The main reason that simulated circuit performance deviates from real circuit performance are variations on the manufacturing process, that manifest as device parameter variations with respect to nominal values. Parameter variations can occur between different batches, wafers on the same batch, dies on the same wafer, and between devices on the same die. These variations can be categorized as either systematic or random (Kinget (2005)).

Variations between batches and between wafers are common to all devices on the circuit, and are due to a systematic shift on the manufacturing process caused by electrical, lithographic or timing differences (Pelgrom, Duinmaijer, and Welbers (1989)). As an example, due to over-etching during photolithography, all transistors may have shorter-than-nominal channel length (Kinget (2005)). Additionally, gradients in the process along the surface of the wafers manifest as parameter variations between dies and even between devices on the same die, that are independent of device size.

Systematic parameter variations are typically accounted for at the design stage. The effects of systematic shifts can be minimized with the use of differential topologies or proper biasing techniques, making the circuit largely insensitive to parameters offsets. The effects of gradient variations can be minimized with the use of proper layout techniques, such as symmetry and common-centroid, to properly match critical transistors.

Random parameter variations are ones that manifests between devices that are in close proximity due to random variations on the manufacturing process, *e.g.* differences in the channel doping concentration between nominally identical transistors. These variations cannot be predicted during the design stage and are dependent on device size. Qualitatively, these kind of variations decrease for larger devices, as the parameter gets averaged over a larger area (Drennan and McAndrew (2003)).

The proposed slice-based analog design technique, object of study of this thesis, is particularly susceptible to gradient variations along the surface of the die. By using pre-existing cells, the application of layout techniques such as symmetry or common-centroid between critical transistors on different cells becomes impossible. Additionally, if a large number of cells are connected in parallel, nominally identical transistors can be separated by large distances.

The effects of parameter variations on the performance of circuits using the proposed design technique is not immediately apparent, and is the focus of the present chapter. The chapter can be broadly divided into two sections. The first half presents the mathematical framework to properly model parameter variations applied to the proposed design technique. The second half presents the results of a Monte Carlo simulation of the scalable charge-sensitive amplifier used in the design of the Heisenberg chip, under the influence of device mismatch.

4.2 Mismatch model

Mismatch is the performance difference between two or more devices on a single integrated circuit (Drennan and McAndrew (2003)). The term is also used to refer to performance differences between real and ideal devices. Differences in device performance can be attributed to parameter variations due to imperfections on the manufacturing process. Intradie parameter variation can be categorized into two: systematic variations due to parameter gradients, that are dependent on the distance between devices; and random variations that are dependent on device size.

The mismatch model presented by Pelgrom *et al.* (Pelgrom et al. (1989)) models the normalized standard deviation of parameter P between two nominally identical MOS transistors on the same die, of width W and length L, separated by a distance D from centroid to centroid, as

$$\sigma^2\left(\frac{\Delta P}{P}\right) = \frac{A_P^2}{WL} + S_P^2 D^2 \tag{4.1}$$

TABLE 4.1. Examples of process and electrical parameters.

Process Parameters	Electrical Parameters
Flatband Voltage (V_{fb})	Drain Current (I_D)
Mobility (μ)	Gate-source voltage (V_{GS})
Substrate Dopant Conc. (N_{sub})	Transconductance (g_m)
Gate Oxide Thickness (t_{ox})	Output conductance (g_o)

where A_P is the area proportionality constant for parameter P and accompanies the size-dependent term, while S_P describes the variation of parameter P with spacing and accompanies the distance-dependent term.

The values of A_P and S_P are typically provided by chip manufacturers for specific parameters, for the sake of mismatch calculations. Due to the random nature of mismatch, and the complexity of the analysis for large circuits, mismatch analysis is well suited for Monte Carlo simulations to assess circuit performance. The random size-dependent term in (4.1) is straightforward to be included in a circuit simulation tool (*e.g.* SPICE). Having identified the transistor mismatch parameters, a normally-distributed random variable needs to be added to each parameter on each transistor of the simulated circuit netlist, properly scaled by the standard deviation provided by the manufacturer and by device size. The systematic distance-dependent term in (4.1) is less straightforward to include in CAD[1] tools, and a physical interpretation of this term is presented in Section 4.4 to model it efficiently.

4.3 Mismatch parameters

For mismatch modeling, critical parameters can be categorized into two types: process and electrical. Process parameters are physically independent parameters that control the electrical behavior of a device. Electrical parameters are those that are of interest to a designer (Drennan and McAndrew (2003)). Table 4.1 shows some examples of process and electrical parameters.

[1]CAD: Computer-aided design

All electrical parameters are subject to mismatch, in the sense that they deviate from their nominal value. However, a limited number of independent process parameters, directly affected by the manufacturing process, are the underlying cause of variations of electrical parameters.

Let us consider an electrical parameter $e(\mathbf{p})$, that is a function of n independent process parameters $\mathbf{p} = \{p_1, p_2, \cdots, p_n\}$. The variance of the electrical parameter $e(\mathbf{p})$ is related to the variance of the independent parameters $\mathbf{p}$ through the propagation of uncertainty relationship (Drennan and McAndrew (2003)), as follows

$$\sigma^2(e) = \sum_{i=1}^{n} \left(\frac{\partial e}{\partial p_i} \right)^2 \sigma^2(p_i) \tag{4.2}$$

There are usually a small number of transistor mismatch parameters that are considered to be dominant. In (Pelgrom et al. (1989)) two main ones (V_{t0} and β) and a secondary one (γ) were suggested, derived from traditional square-law transistor models. With the evolution of transistor models, more accurate mismatch models have since been developed, and ones that utilize additional mismatch parameters, or different mismatch parameters altogether, have been proposed (Serrano-Gotarredona and Linares-Barranco (2003), Croon, Rosmeulen, Decoutere, Sansen, and Maes (2002), Drennan and McAndrew (2003)).

The purpose of the present chapter is not to produce incredibly accurate results, but to gain insight on the potential effects of device mismatch on circuit performance when using the proposed analog design technique. To favor simplicity, and due to the fact that the manufacturer only provides the values of A_β/β and $A_{V_{t0}}$ for the 0.5-μm CMOS technology used in the design of the Heisenberg chip, only the mismatch parameters β and V_{t0} will be considered.

4.4 Physical interpretation of the distance term in Pelgrom's mismatch model

The current section presents a simple physical interpretation and mathematical model for the distance-dependent term in (4.1), as presented in (Linares-Barranco and Serrano-Gotarredona (2007)).

Let us consider an arbitrary layout, where the central coordinates for each transistor M_i are known to be (x_i, y_i). Let us consider an arbitrary parameter subject to mismatch, namely P. Let us assume that, for a given die, it is possible to approximate the gradient of parameter P along the die by a plane, as follows:

$$P(x, y) = Ax + By + C \tag{4.3}$$

where

$$C = P_{Nom} + \Delta P_{Off} \tag{4.4}$$

and A and B are random numbers. The value P_{Nom} represents the nominal value of parameter P, while ΔP_{Off} represents a systematic offset for the given die with respect to the nominal value.

Consider two transistors, namely M_i and M_j, located in positions (x_i, y_i) and (x_j, y_j), respectively. The mismatch between transistors M_i and M_j caused by gradient effects is given by

$$(\Delta P_G)_{ij} = A(x_i - x_j) + B(y_i - y_j) \tag{4.5}$$

For a large number of realizations, the variance of the gradient-related mismatch between two transistors can be computed to be

$$\sigma^2 \left[(\Delta P_G)_{ij} \right] = \sigma^2(A)(x_i - x_j)^2 + \sigma^2(B)(y_i - y_j)^2 \tag{4.6}$$

Assuming symmetry along the axes in the random planes, *i.e.* no preferred direction, then it can be stated that $\sigma(A) = \sigma(B)$. Under this consideration, the expression shown in

(4.6) can be rewritten as

$$\sigma^2\left[(\Delta P_G)_{ij}\right] = \sigma^2(A)\left[(x_i - x_j)^2 + (y_i - y_j)^2\right] = \sigma^2(A)D_{ij}^2 \tag{4.7}$$

where D_{ij} is the distance between transistors M_i and M_j. Comparing equations (4.1) and (4.7) reveals that

$$S_P = \sigma(A) = \sigma(B) \tag{4.8}$$

So, for a given S_P provided by the manufacturer, two random variables can be readily computed to generate a random gradient plane to model the gradient-related systematic parameter variations of a die.

It can be shown (Linares-Barranco and Serrano-Gotarredona (2007)) that this gradient plane physical interpretation is the only valid interpretation of the distance term in Pelgrom's mismatch model.

4.5 Device mismatch in cell-based design

4.5.1 Parameter variation model

Let us consider a circuit layout consisting of k equidistant, horizontally aligned, vertically stacked, equally oriented analog cells, separated by a distance D_{cc}. Figure 4.1 shows a graphical representation of this cell configuration. Let us consider an arbitrary transistor in the first cell, namely M_1, and the corresponding transitors in the other cells, that will be referred to as M_j for cell j.

Let us consider an arbitrary parameter subject to mismatch between nominally identical transistors, namely P. For transistor M_j, the realization of parameter P will be

$$P_j = P_{Nom} + \Delta P_{Off} + (\Delta P_G)_j + (\Delta P_R)_j \tag{4.9}$$

where P_j has been decomposed into a nominal value P_{Nom} for the given transistor, a systematic offset ΔP_{Off} for the given die, and two mismatch components, $(\Delta P_G)_j$ and $(\Delta P_R)_j$. The two mismatch components $(\Delta P_G)_j$ and $(\Delta P_R)_j$ represent gradient, distance-dependent variations, and random, size-dependent variations, respectively.

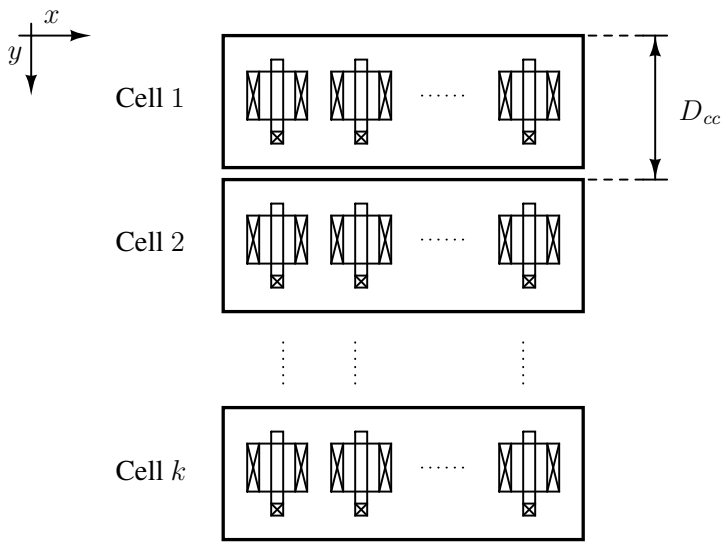


FIGURE 4.1. Cell-based layout with horizontally aligned, vertically stacked cells.

To compute distance-related mismatch between devices, transistor M_1 will be considered as reference. For simplicity, transistor M_1 will be considered to be located at coordinates $(0, 0)$. All gradient related variations for this transistor will be considered to be included in the systematic offset term ΔP_{Off}, common to all devices on the die.

As explained in Section 4.4, the distance-dependent term $(\Delta P_G)_j$ can be written as the product of a plane gradient and the distance between transistors, as follows

$$(\Delta P_G)_j = G_P \cdot D_{1j} \tag{4.10}$$

where G_P is the plane gradient for parameter P in the direction of transistor separation (direction vector $\hat{y}$ in figure 4.1), and D_{1j} is the distance between transistor M_1 and M_j. For this particular formulation, the value of G_P would correspond to B in (4.5), as the cells are horizontally aligned.

For the given assumptions, cell-to-cell distance and transistor to transistor distance are equivalent. Therefore, the distance between transistors M_1 and M_j can be written as a function of D_{cc}, as follows

$$D_{1j} = (j - 1) \cdot D_{cc} \tag{4.11}$$

With these considerations, mismatch parameter P_j can be rewritten as

$$P_j = P_{Die} + (j-1) \cdot D_{cc} \cdot G_P + (\Delta P_R)_j \tag{4.12}$$

where $P_{Die} = P_{Nom} + \Delta P_{Off}$ has a constant value for a given die and arbitrary origin selection.

4.5.2 Parameter scaling

Certain electrical parameters have a linear relationship with the device width, *e.g.* drain current I_D, long-channel current factor β, transconductance g_m, output conductance g_o, among others. When multiple transistors are connected in parallel, the individual instances of the parameter for each transistor can be added together directly to compute the resulting parameter of the equivalent transistor. In this section, this type of parameters will be considered.

Let us consider the equivalent transistor when k cells are connected together in parallel. The equivalent mismatch parameter P for this equivalent transistor can be written as

$$P_{eq}(k) = k \cdot P_{Die} + D_{cc} \cdot G_P \cdot \sum_{j=1}^{k}(j-1) + \sum_{j=1}^{k}(\Delta P_R)_j \tag{4.13}$$

In the middle term of (4.13), the sum of the first k positive integers is a well known summation known as the triangle number, given by the following explicit formula:

$$\sum_{j=1}^{k} j = \frac{k(k+1)}{2} \tag{4.14}$$

Thus, the expression shown in (4.13) can be rewritten as

$$P_{eq}(k) = k \cdot P_{Die} + \frac{k(k-1)}{2} \cdot D_{cc} \cdot G_P + \sum_{j=1}^{k}(\Delta P_R)_j \tag{4.15}$$

To simplify the analysis, systematic parameter offsets will be neglected. For instance, we will assume $P_{Die} = P_{Nom}$, that will be referred to as P for the sake of notation. Only intradie relative differences between individual devices will be considered. Under this

consideration, $P_{eq}(k)$ can be written in a normalized form, as follows

$$\frac{P_{eq}(k)}{k \cdot P} = 1 + \left(\frac{k-1}{2}\right)\left(\frac{D_{cc} \cdot G_P}{P}\right) + \frac{1}{k} \cdot \sum_{j=1}^{k} \frac{(\Delta P_R)_j}{P} \tag{4.16}$$

Further insight can be obtained from (4.16) by analyzing the variance of the expression for a large number of realizations. To compute the variance, since all the additive random effects are uncorrelated, the individual variances are added as follows:

$$\sigma^2\left[\frac{\Delta P}{P}(k)\right] = \left(\frac{k-1}{2}\right)^2 \cdot D_{cc}^2 \cdot \sigma^2\left(\frac{G_P}{P}\right) + \frac{1}{k^2} \cdot \sum_{j=1}^{k} \sigma^2\left[\frac{(\Delta P_R)_j}{P}\right] \tag{4.17}$$

The variance of the random variables can be written using the notation shown in (4.1), as follows

$$\sigma^2\left(\frac{G_P}{P}\right) = S_P^2 \tag{4.18}$$

$$\sigma^2\left[\frac{(\Delta P_R)_1}{P}\right] = \sigma^2\left[\frac{(\Delta P_R)_2}{P}\right] = \cdots = \sigma^2\left[\frac{(\Delta P_R)_k}{P}\right] = \frac{A_P^2}{2WL} \tag{4.19}$$

where the factor of 2 in the denominator of (4.19) accounts for the fact that each transistor deviates from a nominal mismatchless transistor. In contrast, the formulation presented in (4.1) models the relative variation between two devices.

Using this notation, the expression shown in (4.17) can be rewritten as

$$\sigma^2\left[\frac{\Delta P}{P}(k)\right] = \left(\frac{k-1}{2}\right)^2 \cdot D_{cc}^2 \cdot S_P^2 + \frac{1}{2k} \cdot \frac{A_P^2}{WL} \tag{4.20}$$

A close inspection of the expression shown in (4.20) reveals that it is a direct application of Pelgrom's mismatch model. An increase in k linearly increases the equivalent width of the device, so the size-dependent term is inversely proportional to k. At the same time, an increase in k increases the average distance between the individual devices in the array, so the distance-dependent term is proportional to k^2. For a given value of D_{cc}, this expression ties together both gradient and random variations using a single variable k.

From a design perspective, the expression shown in (4.20) reveals that the slice-based analog design technique is particularly susceptible to gradient-related variations, since for

a given cell library there is no way to reduce cell pitch. There is a trade-off between improvements on device performance (*e.g.* noise performance), and performance uncertainty due to mismatch effects with respect to the expected performance, both of which increase with k.

A single value of k can be calculated for which the variance of the two mismatch effects are matched. This is relevant to assess which is the dominant effect in a given design. The value of k for which the two additive terms in (4.20) are equal can be calculated by solving for k the following expression

$$k(k-1)^2 = C \tag{4.21}$$

where

$$C = \frac{2 \cdot \frac{A_P^2}{WL}}{D_{cc}^2 \cdot S_P^2} \tag{4.22}$$

There are three roots for the polynomial expression shown in (4.21), only one of which is strictly positive and real, and therefore of physical significance. The resulting expression is unintuitive and doesn't offer any insight into circuit operation by itself, thus is presented only for the sake of completeness.

$$k = \frac{1}{3}\left[\frac{\sqrt[3]{3\sqrt{3}\cdot\sqrt{27C^2-4C}+27C-2}}{\sqrt[3]{2}} + \frac{\sqrt[3]{2}}{\sqrt[3]{3\sqrt{3}\cdot\sqrt{27C^2-4C}+27C-2}} + 2\right] \tag{4.23}$$

The expression shown in (4.23) will most likely yield a non-integer solution. However, it can be numerically evaluated to compute a value of k, that will be referred to as k^*. For $k < k^*$, random size-dependent variations have higher mismatch variance. For $k > k^*$, gradient distance-dependent variations have higher mismatch variance.

4.6 A simple example of noise scaling in the presence of device mismatch

Noise performance is one of the most important metrics used to assess circuit performance of a particle physics front-end. In order analyze the slice-based design methodology applied to particle physics instrumentation, object of study of this thesis, a simple example of the effects of mismatch on noise performance is presented.

Let us consider an arbitrary circuit with a voltage input, for which the noise performance is dominated by the input device. Let us assume that the input device is operating in strong inversion and that thermal noise is the dominant noise component. The input referred power spectral density (PSD) of the circuit is given by

$$\overline{V_n^2} = \frac{4k_B T\gamma}{g_m} \tag{4.24}$$

Let us assume, for the sake of simplicity, that only the mismatch of the input transistor has an effect in noise performance. Mismatch analysis done by hand for a multiple-transistor circuit can become very impractical and unintuitive. Monte Carlo simulations are better suited to analyze circuit performance when multiple devices are subjected to mismatch.

Under these assumptions, and using the propagation of uncertainty relationship presented in (4.2), the mismatch-related variations of the nominal noise PSD can be computed

$$\begin{aligned}
\sigma^2\left(\Delta\overline{V_n^2}\right) &= \left(\frac{\partial\overline{V_n^2}}{\partial g_m}\right)^2 \sigma^2\left(\Delta g_m\right) \\
&= \left(\frac{4k_B T\gamma}{g_m^2}\right)^2 \sigma^2\left(\Delta g_m\right) \\
&= \left(\frac{4k_B T\gamma}{g_m}\right)^2 \cdot \sigma^2\left(\frac{\Delta g_m}{g_m}\right)
\end{aligned} \tag{4.25}$$

where the left term in the right side of the expression can be identified as $\left(\overline{V_n^2}\right)^2$. Thus

$$\sigma^2\left(\frac{\Delta\overline{V_n^2}}{\overline{V_n^2}}\right) = \sigma^2\left(\frac{\Delta g_m}{g_m}\right) \tag{4.26}$$

This equation can be expressed as a function of the mismatch parameters of interest β and V_{t0}. For the assumptions under consideration in the current example

$$g_m = \beta \cdot V_{OV} \tag{4.27}$$

where the overdrive voltage V_{OV} is defined as $V_{OV} = (V_{GS} - V_t)$.

For long-channel devices and strong inversion operation, the threshold voltage V_t can be modeled as (Muller and Kamins (2002))

$$V_t = V_{t0} + \gamma \left(\sqrt{|V_{SB}| + \Phi_s} - \sqrt{\Phi_s} \right) \tag{4.28}$$

where V_{t0} is the ideal threshold voltage for zero substrate bias, γ is the body bias coefficient, V_{SB} is the source to bulk voltage, and Φ_s is the surface potential. The exact meaning of this definition is irrelevant for the following analysis, but the relationship is mentioned because the variance $\sigma^2(\Delta V_{t0})$, measured in voltage squared, is one of the the values typically provided by a chip manufacturer.

The propagation of uncertainty relationship presented in (4.2) can be applied to (4.27) to compute the variance of g_m as a function of the mismatch parameters β and V_{t0}, resulting in the following relation

$$\sigma^2 \left(\frac{\Delta g_m}{g_m} \right) = \sigma^2 \left(\frac{\Delta \beta}{\beta} \right) + \left(\frac{V_{t0}}{V_{OV}} \right)^2 \cdot \sigma^2 \left(\frac{\Delta V_{t0}}{V_{t0}} \right) \tag{4.29}$$

thus, the values of S_{g_m} and A_{g_m} can be written as

$$S_{g_m}^2 = S_\beta^2 + \left(\frac{V_{t0}}{V_{OV}} \right)^2 \cdot S_{V_{t0}}^2 \tag{4.30}$$

$$A_{g_m}^2 = A_\beta^2 + \left(\frac{V_{t0}}{V_{OV}} \right)^2 \cdot A_{V_{t0}}^2 \tag{4.31}$$

For a circuit composed of k nominally identical copies connected in parallel, the value of $\sigma^2 \left[\frac{\Delta g_m}{g_m}(k) \right]$ for the equivalent circuit can be modeled using the expression shown in (4.20) applied to mismatch parameter $g_m(k)$. Thus, the noise variations for this circuit can

be modeled to be

$$\sigma^2\left[\frac{\Delta\overline{V_n^2}}{\overline{V_n^2}}(k)\right] = \left(\frac{k-1}{2}\right)^2 \cdot D_{cc}^2 \cdot S_{g_m}^2 + \frac{1}{2k} \cdot \frac{A_{g_m}^2}{WL} \tag{4.32}$$

The expression presented in (4.32) shows that the variance of the relative variations of the input-referred thermal noise PSD. Size-dependent variations are inversely proportional to k, due to the fact that the equivalent device increases in size directly proportional to k. Distance-dependent variations are proportional to k^2, due to the fact that the average distance between devices increases as more cells are connected in parallel.

For this example, it is more insightful to analyze the absolute variations of noise performance. The nominal thermal noise decreases with k, and can be expressed as $\overline{V_n^2}(k) = \overline{V_n^2}(1)/k$. Thus, the relationship shown in 4.32 can be rewritten as

$$\sigma^2\left[\Delta\overline{V_n^2}(k)\right] = \left[\overline{V_n^2}(1)\right]^2 \cdot \left[\left(\frac{1-\frac{1}{k}}{2}\right)^2 \cdot D_{cc}^2 \cdot S_{g_m}^2 + \frac{1}{2k^3} \cdot \frac{A_{g_m}^2}{WL}\right] \tag{4.33}$$

This expression shows that, for absolute noise variations, the size dependent term is inversely proportional to k^3. This means that size-dependent variations becomes increasingly less dominant at a fast rate for increasing values of k. On the other hand, distance dependent variations are proportional to $(1-1/k)^2$, which converges to 1 for large values of k. This means that, for a large number of parallel-connected cells, size-dependent variations can be potentially negligible, and distance-dependent variations are capped at a maximum variance value of $\left[\overline{V_n^2}(1) \cdot (D_{cc} \cdot S_{g_m})/2\right]^2$.

4.7 Monte Carlo simulations

Two different Monte Carlo simulations were carried out to evaluate the variations in circuit performance due to mismatch for a varying number of nominally identical parallel-connected cells. The first of the two, the results of which are presented in Section 4.7.2, is a single transistor simulation, intended to emulate the conditions of the example presented in Section 4.6, in order to evaluate the applicability of the mismatch model presented in the previous sections. The second simulation, the results of which are presented in Section

4.7.3, is of the charge-sensitive amplifier used in the design of the Heisenberg IC, to gain insight into the behavior of the circuit when mismatch becomes relevant.

4.7.1 Methodology

The Monte Carlo simulations were performed by means of multiple circuit simulations using LTspice, each of them representing a single instance of the random process. The circuit simulations were performed using the models of a commercial 0.5-$\mu\mathrm{m}$ CMOS process. Only intradie mismatch effects were considered, *i.e.* gradient and random variations. Mismatch parameters V_{t0} and β were considered.

A MATLAB script was used to run the Monte Carlo simulations. The script operations follow these steps: random parameter variations are computed; a unique circuit netlist for each simulation iteration is written; LTspice is called to simulate the netlist; the simulation results are read-out and compiled; and the process is repeated to gather statistics.

The effects of parameter variations were added to the simulated netlist at the level of the device model. Each transistor was instantiated with a slightly different model, subject to parameter variations. There is a limitation with this approach, however, since β is an electrical parameter, not present on the transistor model. Instead it was assumed, for the sake of simplicity, that variations in β were only due to variations in carrier mobility μ, and the effects were modeled by adjusting the value of μ on the device model.

The simulations were done using a varying number of parallel-connected cells, namely k, for $k = \{1, 2, ..., 8\}$. The distance between adjacent cells was considered to be the same in both simulations, with a value of $D_{cc} = 276.6\ \mu\mathrm{m}$. The values of both the number of parallel-connected cells and the distance between adjacent cells, are the same as the ones used in the Heisenberg chip.

To implement the effects of random, size-dependent variations, two normally-distributed random variables $\{\Delta V_{t0,R}, \Delta\mu_R\}$ were generated for each transistor. To implement the

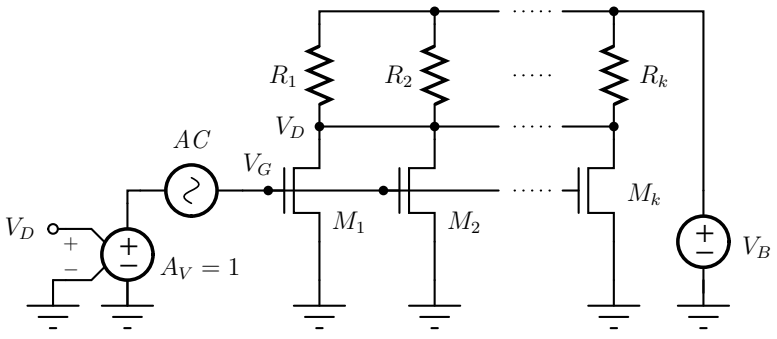


FIGURE 4.2. Schematic for Monte Carlo mismatch simulations. A single equivalent device circuit is considered. Simulation parameters include $L_j = 0.6\mu\text{m}$, $W_j = 127.5\mu\text{m}$, $R_j = 1\Omega$, and $V_B = 1\text{V}$. This bias voltage results in the device operating in strong inversion, at a $g_m/I_D \approx 4.5\ \frac{\text{mS}}{\text{mA}}$.

effects of gradient, distance-dependent variations, four normally distributed random variables $\{(G_{V_{t0}})_{\text{NMOS}}, (G_{V_{t0}})_{\text{PMOS}}, (G_\mu)_{\text{NMOS}}, (G_\mu)_{\text{PMOS}}\}$ were generated for the whole circuit, to serve as plane gradients. For each simulation realization, the resulting parameters were computed using the expression shown in (4.12) for each individual transistor.

4.7.2 Results – Single device noise scaling

A simple single-device Monte Carlo simulation, intended to emulate the conditions of the example presented in Section 4.6, was performed to test the applicability of the mismatch model presented in (4.20). Figure 4.2 shows a schematic representation of the simulated circuit, together with some relevant simulation parameters. The equivalent device is biased at $V_{GS} = V_{DS}$, to assure that it always operates in active region. For the selected simulation parameters, the equivalent device operates in strong inversion, at a g_m/I_D value of approximately $4.5\ \frac{\text{mS}}{\text{mA}}$.

Three simulation scenarios were considered:

1. Random variations only
2. Gradient variations only
3. A combination of the two effects

TABLE 4.2. Single device Monte Carlo simulation – Random and gradient standard deviation values.

Scenario	Type	$A_{V_{t0}}$ $[\% \cdot \mu\text{m}]$	A_μ $[\% \cdot \mu\text{m}]$	$S_{V_{t0}}$ $[\%/\mu\text{m}]$	S_μ $[\%/\mu\text{m}]$
$\#1$	NMOS	0.0338	1.3421	0	0
	PMOS	0.0290	1.2389	0	0
$\#2$	NMOS	0	0	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$
	PMOS	0	0	$3 \cdot 10^{-3}$	$3 \cdot 10^{-3}$
$\#3$	NMOS	0.0338	1.3421	$2.5 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$
	PMOS	0.0290	1.2389	$2.5 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$

The values of $A_{V_{t0}}$, A_μ, $S_{V_{t0}}$ and S_μ were arbitrarily selected to favor the clarity of the plotted results. The standard deviation values of the mismatch parameters are presented in Table 4.2. For each scenario, a total of 2000 points were simulated for each value of k.

Noise simulations were performed considering only the effects of thermal noise. The PSD of the equivalent device was measured at the output node, and referred to the input node of the equivalent device via the equivalent transconductance. For this single device simulation, only the mean and variance values of $\overline{V_n^2}(j\omega)$ were analyzed.

The mean noise value is shown in Figure 4.3 as a function of the number of parallel-connected cells. This plot was obtained from the results of simulation scenario $\#3$, however, the resulting plot for the other simulation scenarios showed no measurable difference. The noise values coincide with the ones obtained from a mismatchless simulation.

The variance values for simulation scenarios $\#1$, $\#2$ and $\#3$ are shown in Figures 4.4, 4.5 and 4.6, respectively. Both the absolute ($\sigma^2(\Delta\overline{V_n^2})$) and normalized ($\sigma^2(\Delta\overline{V_n^2}/\overline{V_n^2})$) variances of $\overline{V_n^2}$ are plotted. In each figure, together with the data points, a fitted curve is presented. The fitted curve was obtained using the MATLAB `fit` function. The curve was computed using the nonlinear least-squares method, for the fitting models shown in the legend of each plot, using the trust region algorithm. The mismatch models presented in (4.32) and (4.33) were used as the fitting models, expressed as a function of independent variable k.

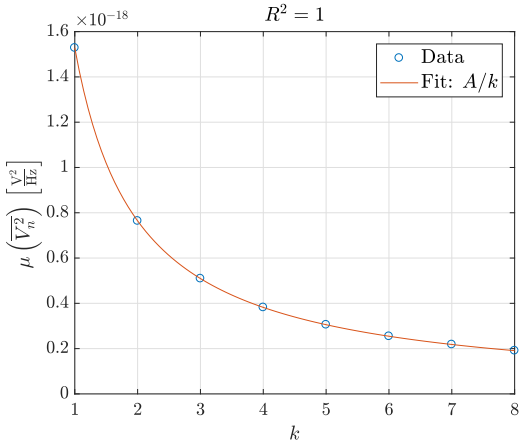
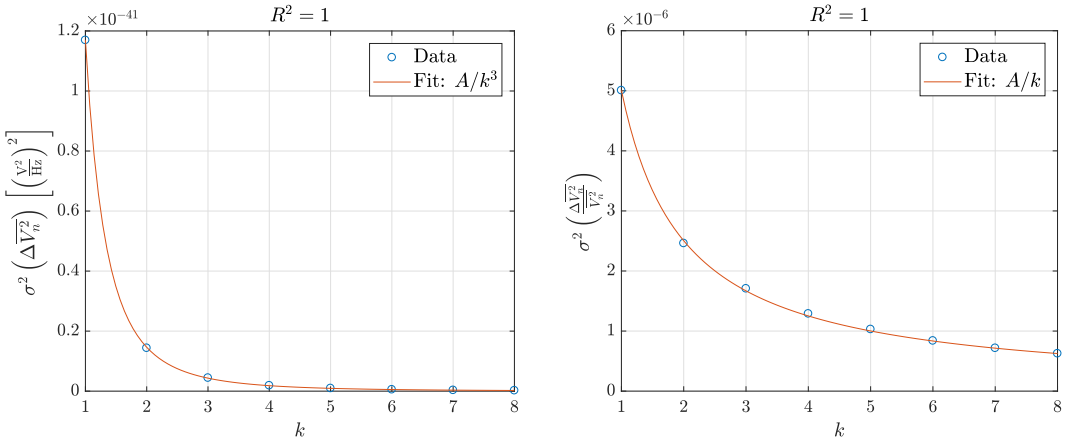


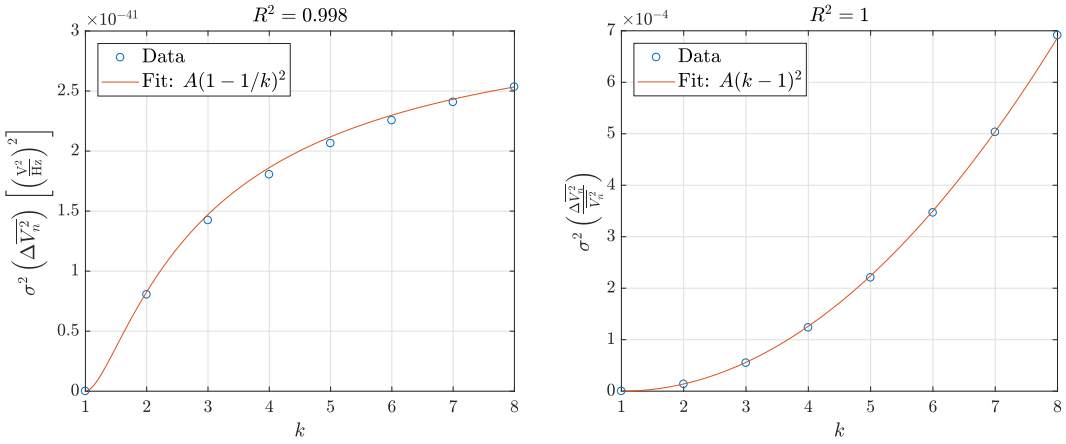
FIGURE 4.3. Mean input-referred noise PSD *vs.* # of parallel connected cells. The results were obtained from simulation scenario #3, however, the resulting plots for the other simulation scenarios showed no measurable differences.



(A) $\{A = 1.169 \times 10^{-41}\}$ (B) $\{A = 5.007 \times 10^{-6}\}$

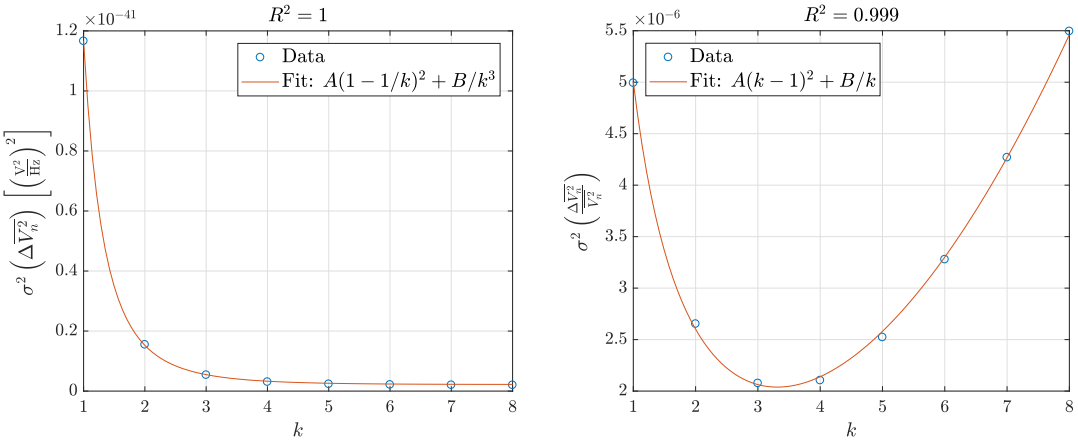
FIGURE 4.4. Absolute (A) and normalized (B) variance of the input-referred noise PSD *vs.* # of parallel connected cells, for simulation scenario #1.

Figures 4.4, 4.5 and 4.6 also show the calculated fit coefficients for each model (A, or A and B, depending on the number of fit coefficients). The exact coefficient values are not relevant to the current analysis, since this analysis is limited to the applicability of the mismatch model.



(A) $\{A = 3.31 \times 10^{-41}\}$ (B) $\{A = 1.401 \times 10^{-5}\}$

FIGURE 4.5. Absolute (A) and normalized (B) variance of the input-referred noise PSD $vs.$ # of parallel connected cells, for simulation scenario #2.



(A) $\{A = 2.619 \times 10^{-43}\}$, $\{B = 1.166 \times 10^{-41}\}$

(B) $\{A = 9.87 \times 10^{-8}\}$, $\{B = 5.004 \times 10^{-6}\}$

FIGURE 4.6. Absolute (A) and normalized (B) variance of the input-referred noise PSD $vs.$ # of parallel connected cells, for simulation scenario #3.

As a measure of the goodness of the fit, the value of the coefficient of determination R^2 is presented on each plot in Figures 4.4, 4.5 and 4.6. For all cases, the value of R^2 approximates very closely to 1, which indicates the data behavior can be almost completely predicted by the fitting models. For Figures 4.5(A) and 4.6(B), deviations from $R^2 = 1$

TABLE 4.3. Charge-sensitive amplifier Monte Carlo simulation – Random and gradient standard deviation values.

Scenario	Type	$A_{V_{t0}}$ $[\% \cdot \mu\mathrm{m}]$	A_{μ} $[\% \cdot \mu\mathrm{m}]$	$S_{V_{t0}}$ $[\%/\mu\mathrm{m}]$	S_{μ} $[\%/\mu\mathrm{m}]$
#1	NMOS	0.0338	1.3421	0	0
	PMOS	0.0290	1.2389	0	0
#2	NMOS	0	0	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$
	PMOS	0	0	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$
#3	NMOS	0.0338	1.3421	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$
	PMOS	0.0290	1.2389	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$

can potentially be attributed to the random nature of the Monte Carlo simulation, under the assumption that these values will likely converge to unity for more simulation points.

This single device Monte Carlo simulation shows that the mismatch models presented in (4.32) and (4.33), as derived from the example shown in Section 4.6, appear to be valid and applicable.

4.7.3 Results – Charge-sensitive amplifier

A Monte Carlo simulation of the charge-sensitive amplifier used in the design of the Heisenberg IC was performed to gain insight into the behavior of the full circuit under the influence of device mismatch. The details of the design of the CSA are presented in Chapter 6. The circuit schematic of the CSA and the bias circuit are presented in Figures 6.3 and 6.5, respectively. Both interconnected circuit blocks compose each individual simulated slice. The transistor design parameters for the amplifier and the bias circuit are presented in Tables 6.1 and 6.2, respectively.

Three simulation scenarios were considered:

1. Random variations only
2. Gradient variations only
3. A combination of the two effects

The values of $S_{V_{t0}}$ and S_{μ} were arbitrarily selected, given the manufacturer does not provide these values. The values of $A_{V_{t0}}$ and A_{μ} were computed using the values provided by the manufacturer as reference, namely $(A_{V_{t0}})_{\mathrm{Manu.}}$ and $(A_{\beta})_{\mathrm{Manu.}}$. Slightly larger values

than the ones provided, of

$$A_{V_{t0}} = \sqrt{2} \cdot (A_{V_{t0}})_{\text{Manu.}} \quad \text{and} \quad A_{\mu} = \sqrt{2} \cdot (A_{\beta})_{\text{Manu.}}$$

were used in the simulation to better observe edge cases, considering that a relatively small number of simulated points was used. For each scenario, 2000 iterations were simulated, each one corresponding to the connection of slices in sequence from 1 to 8. The values of $A_{V_{t0}}$, A_{μ}, $S_{V_{t0}}$ and S_{μ}, as used in the simulation, are presented in Table 4.3. These values, although arbitrarily selected, provide an acceptable point of reference, and the simulation results are intended to provide a reasonable explanation for the real world performance variations of the Heisenberg chip with respect to a mismatchless simulation. The measured results of the Heisenberg chip, together with an interpretation of these results based on this mismatch Monte Carlo simulation, are presented in Chapter 8.

Noise simulations were performed considering only the effects of thermal noise. The output total integrated noise $\overline{V_{O,noise}^2}$ was probed at the output node of the equivalent circuit. A bandwidth of integration ranging from $0.1\ \text{mHz}$ to $100\ \text{kHz}$ was considered.

Using the single parameter mismatch model presented in (4.12) it is potentially possible to compute an analytical expression for the statistical properties of the output integrated noise as a function of k considering the combined mismatch effect of all transistors in the circuit. However, this process is highly impractical and will most likely yield an unintuitive expression. Instead, the models derived for a single device will be used as a reference to analyze the results.

The mean noise value is shown in Figure 4.7 as a function of the number of parallel-connected cells. This plot was obtained from the results of simulation scenario $\#1$, however, the resulting plots for the other simulation scenarios showed no observable differences, and coincide with the noise values obtained from a mismatchless simulation.

The variance values for simulation scenarios $\#1$, $\#2$ and $\#3$ are shown in Figures 4.8, 4.9 and 4.10, respectively. Both the absolute and normalized[2] variances of $\overline{V_{O,noise}^2}$

[2] As defined on (4.16) and (4.17).

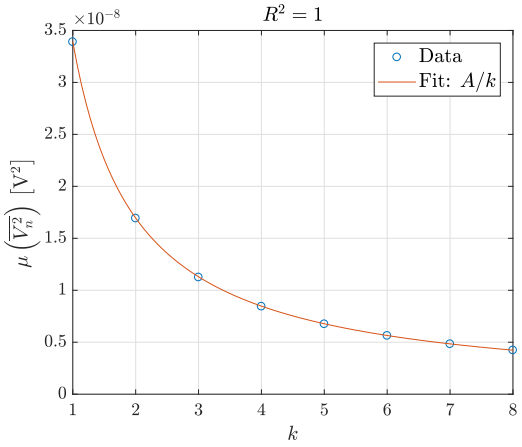
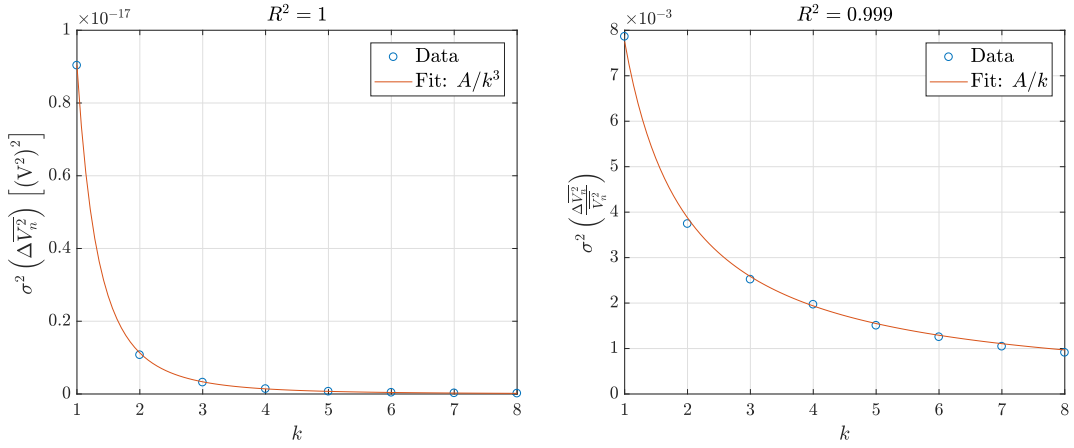


FIGURE 4.7. Mean output integrated noise $vs.$ # of parallel connected cells. The integrated bandwidth ranges from $0.1\,\mathrm{mHz}$ to $100\,\mathrm{kHz}$. The results were obtained from simulation scenario #1. The resulting plots for the other simulation scenarios showed no observable differences.

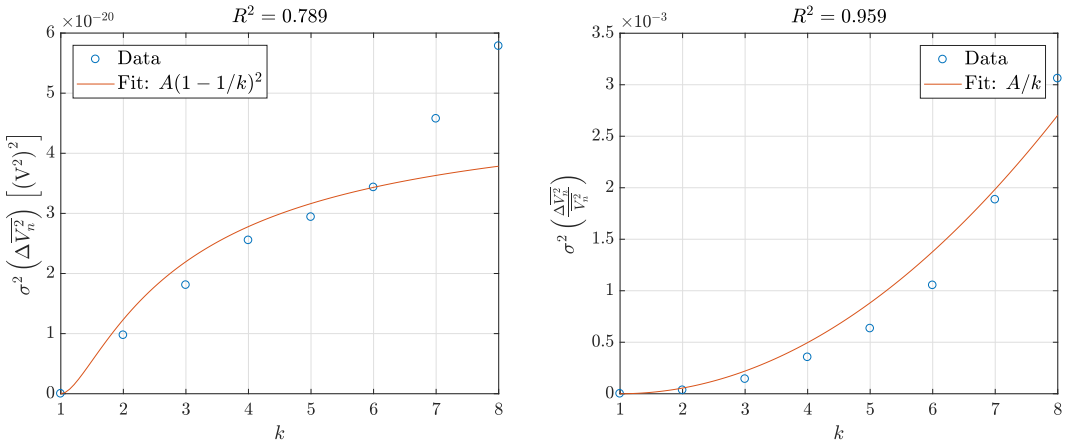


(A) $\{A = 9.033 \times 10^{-18}\}$ (B) $\{A = 7.75 \times 10^{-3}\}$

FIGURE 4.8. Absolute (A) and normalized (B) variance of the output integrated noise $vs.$ # of parallel connected cells, for simulation scenario #1.

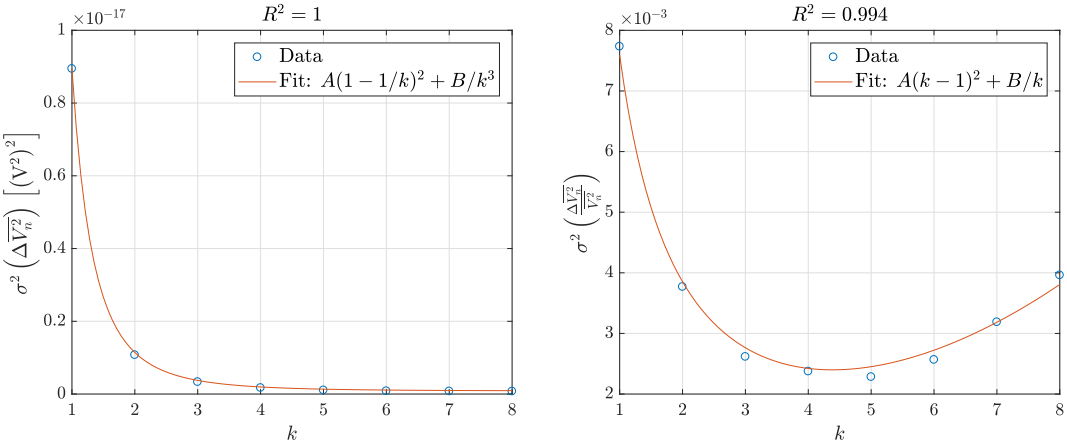
are plotted. A fitted curve is also included for each plot, computed using the single device mismatch model presented in (4.32) and (4.33) as a function of k, for reference.

The results of simulation scenario #1, shown in Figure 4.8, indicate that the mismatch model derived for a single device appears to properly model the behavior of the output integrated noise of the CSA when only size-dependent mismatch is considered. Potentially,



(A) $\{A = 4.944 \times 10^{-20}\}$ (B) $\{A = 5.519 \times 10^{-5}\}$

FIGURE 4.9. Absolute (A) and normalized (B) variance of the output integrated noise *vs.* # of parallel connected cells, for simulation scenario #2.



(A) $\{A = 9.801 \times 10^{-20}\}$, $\{B = 8.944 \times 10^{-18}\}$ (B) $\{A = 5.833 \times 10^{-5}\}$, $\{B = 7.591 \times 10^{-3}\}$

FIGURE 4.10. Absolute (A) and normalized (B) variance of the output integrated noise *vs.* # of parallel connected cells, for simulation scenario #3.

a more general mismatch model can be derived for an arbitrary circuit, or certain topologies. This is, however, beyond the scope of this thesis.

The results of simulation scenario #2, shown in Figure 4.9, indicate that the mismatch model derived for a single device, when only distance-dependent mismatch is considered,

is an unsatisfactory predictor of mismatch behavior of the CSA. This is particularly apparent in Figure 4.9(A). In this plot, for $k \geq 6$, it would appear that second-order effects, not accounted for in the proposed mismatch model, become dominant. It appears that the single-device model underestimates de amount of distance-related mismatch of a multiple-device circuit.

The results of simulation scenario $\#3$, shown in Figure 4.10, indicate that the proposed mismatch model adequately describes the mismatch behavior of this particular example. The discrepancy with the observations of the previous paragraph, where distance-related mismatch was considered in isolation, can be attributed to the fact that the power of the size-dependent mismatch is dominant in both plots.

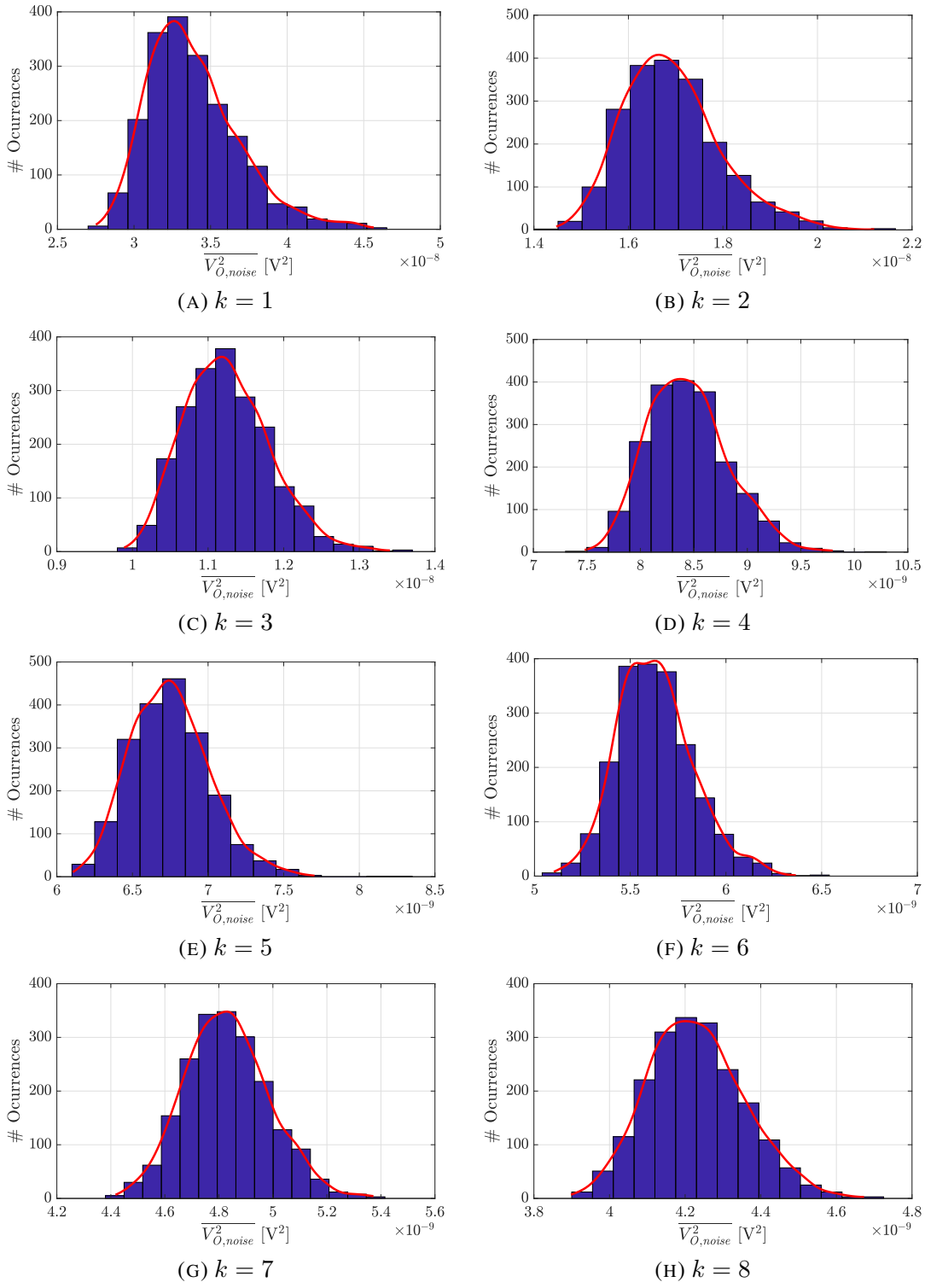
The histogram of the output integrated noise for different values of k is shown in Figures 4.11 and 4.12, for simulation scenarios $\#1$ and $\#2$, respectively. From the results of simulation scenario $\#1$, it can be observed that for smaller values of k, which have higher mismatch power, there appears to be a tendency of the probability distributions towards skewing to the right side of the plot. This can be attributed to the fact that, for higher mismatch power, the assumption that mismatch effects are small enough that they can be considered to be additive no longer holds (Pelgrom et al. (1989)).

From the results of simulation scenario $\#2$, shown in Figure 4.12, it can be observed that the probability distribution appears to be skewed toward the right side of the plot in all cases. This cannot be attributed to the magnitude of the mismatch power, given that the mismatch power of size-dependent variations is generally larger without displaying the same behavior. From Figure 4.10(B), it can be seen that the power of both mismatch effects is matched at a value close to $k = 5$, which is the discrete minimum of the data points. Considering $k = 5$ as an example, and observing Figures 4.11(E) and 4.12(D), reveals that both mismatch effects have significantly different probability distributions even when they have similar power.

The difference in the effect that size and distance-dependent variations have on the output noise of the circuit can potentially be attributed to the difference in nature of

both types of variations. The effect of gradient-related mismatch is cumulative, *e.g.* all NMOS currents increase/decrease in the direction of gradient variations, while the effect of distance-dependent variations can be averaged, *e.g.* an increase in current of one transistor in one cell is averaged with a decrease in current of a transistor in a different cell. This fundamental difference in behavior is non-trivial, and a proper understanding of this difference and its effects on circuit behavior requires further analysis.

Regardless of the reason why the probability density functions of the obtained results do not appear to behave as a normal distribution, the fact that they do not, certainly contributes to the poor applicability of the proposed single-device mismatch model, particularly in the case of distance-dependent variations.



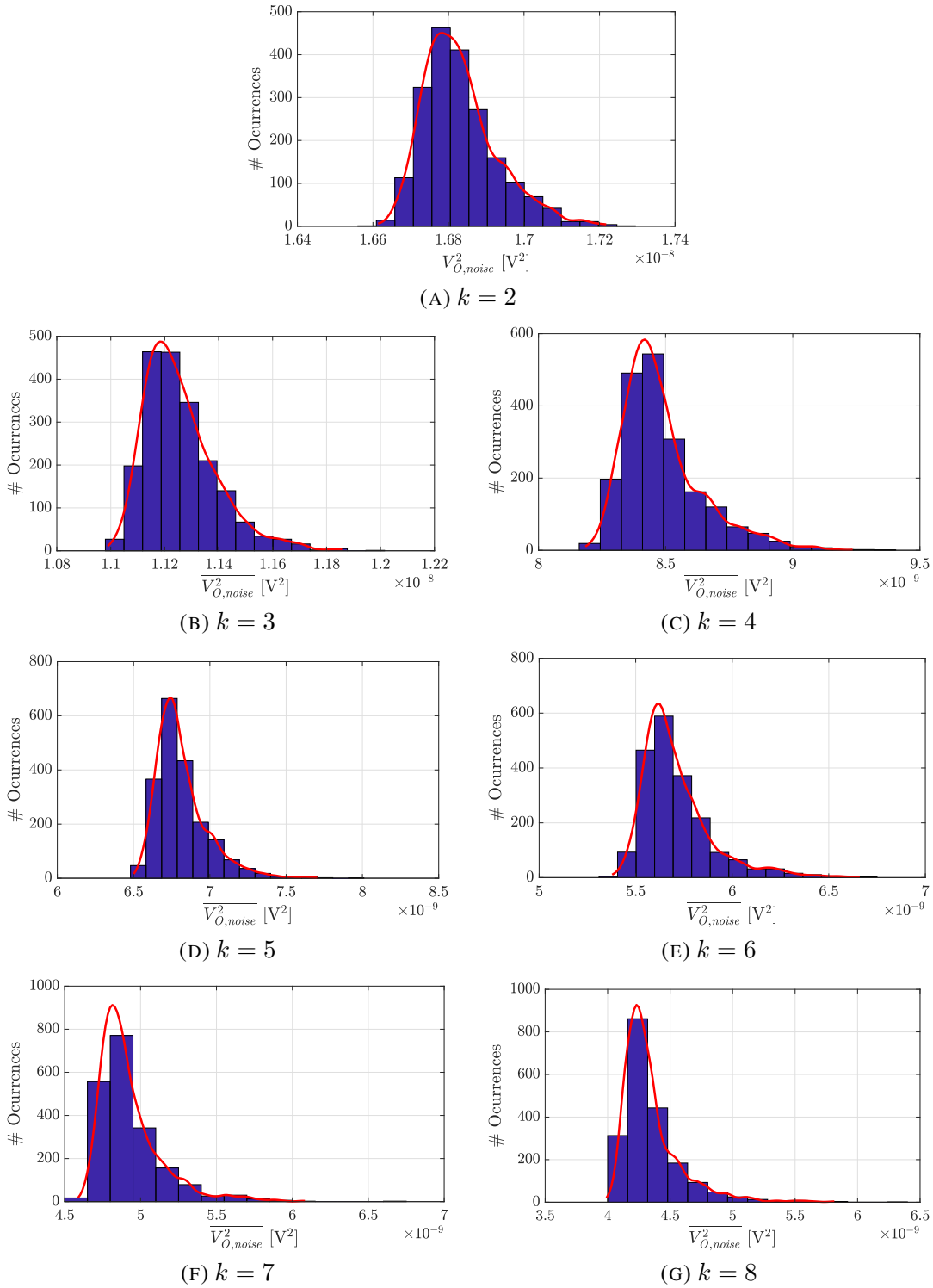
(A) $k = 1$ (B) $k = 2$

(C) $k = 3$ (D) $k = 4$

(E) $k = 5$ (F) $k = 6$

(G) $k = 7$ (H) $k = 8$

FIGURE 4.11. Simulation scenario #1 – Histogram of the output integrated noise for different numbers of parallel connected cells. A Kernel curve estimation of the probability density function is also included for each case, computed using the MATLAB `histfit` command.



(A) $k = 2$

(B) $k = 3$

(C) $k = 4$

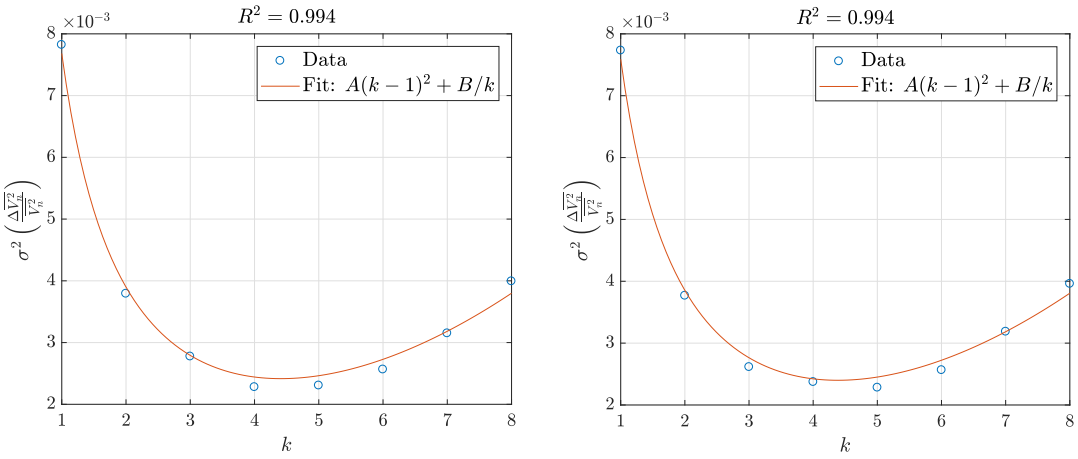
(D) $k = 5$

(E) $k = 6$

(F) $k = 7$

(G) $k = 8$

FIGURE 4.12. Simulation scenario #2 – Histogram of the output integrated noise for different numbers of parallel connected cells. A Kernel curve estimation of the probability density function is also included for each case. The histogram for $k = 1$ was omitted, given that there is no dispersion in the noise measurements in that particular case.



(A) $\{A = 5.794 \times 10^{-5}\}$, $\{B = 7.682 \times 10^{-3}\}$

(B) $\{A = 5.833 \times 10^{-5}\}$, $\{B = 7.591 \times 10^{-3}\}$

FIGURE 4.13. CSA Monte Carlo simulation – Normalized variance comparison: switches (A) *vs.* no switches (B).

4.7.4 Mismatch and switches – The charge-sensitive amplifier of the Heisenberg IC

The Monte Carlo simulation of the charge-sensitive amplifier design used in the Heisenberg IC, as presented in Section 4.7.3, assumed that the amplifier slices were connected in parallel using wires. In the implementation of the Heisenberg chip, the amplifier slices were connected in parallel using switch banks in the interface between slices. To assess whether the addition of switches in the circuit has a significant effect on the mismatch behavior of the circuit, the Monte Carlo simulation was performed again with the addition switches, of the same characteristics as the ones used in the Heisenberg chip. The obtained results showed no significant variation between the two simulations, so it was concluded that the addition of switches, particularly the relatively large switches used in the design of Heisenberg, has no significant effect in the mismatch behavior of the circuit. As an example, Figure 4.13 shows a comparison between the normalized variance of the circuit in the presence and absence of switches.

5. SYSTEM-LEVEL DESIGN

5.1 Introduction

The general goal of the present research document is to assess whether the slice-based analog design methodology can effectively be used to design a functional charge-sensitive amplifier to cover a wide range of technical specifications in particle physics instrumentation. Two specific goals were defined in order to fulfill this general goal:

- to achieve a thorough understanding of the limitations and benefits of designing a charge-sensitive amplifier using this methodology;
- and to implement a fully functional pulse processor channel using a configurable charge-sensitive amplifier to measure real-world performance.

To fulfill the latter, an ASIC[1] was designed and fabricated (hereafter, Heisenberg) and a testing system for this chip was implemented. The present chapter deals with the system-level design of the Heisenberg chip and the surrounding test system.

5.2 System block diagram

Figure 5.1 shows the block diagram of the testing system used to measure the performance of the Heisenberg chip. The block diagram illustrates, for the most part, a typical implementation of a pulse processing chain. Figure 5.1 also shows the naming convention to be used in the rest of this document: the Heisenberg IC, the Heisenberg test board, and the Heisenberg test system, in increasing hierarchical order.

The function of each block shown in Figure 5.1 can be better described in order by following the signal path, *i.e.* from left to right. In the leftmost side of the figure, two mutually-exclusive input sources can be distinguished: the particle detector and the pre-charger circuit. The particle detector is used under regular mode of operation on a typical pulse processor. It generates a finite amount of electrical charge in response to stimuli.

[1] ASIC: Application Specific Integrated Circuit

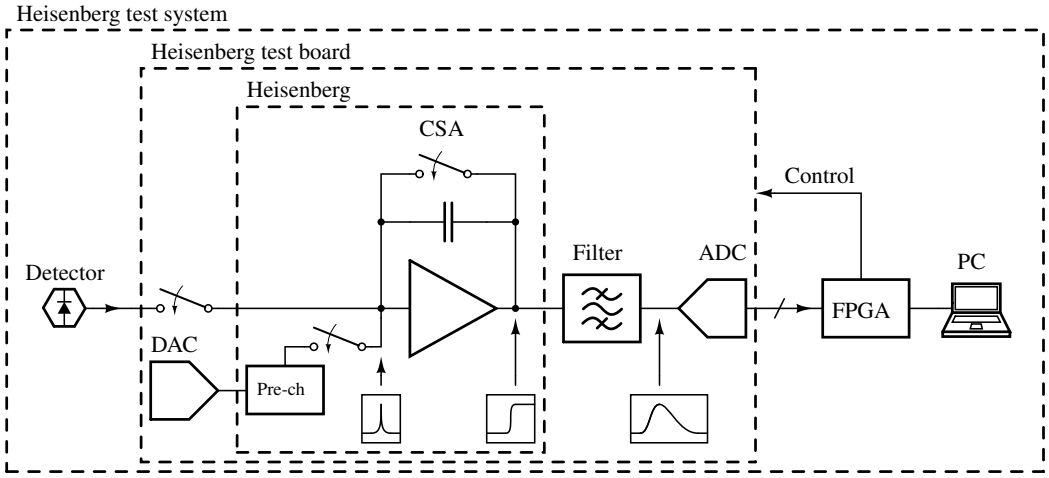


FIGURE 5.1. Testing system block diagram.

The Heisenberg test system, although intended to be a fully functional pulse processing channel, was not designed to work with any particular detector. The pre-charger circuit is used to inject a known amount of electrical charge into the circuit, to test the circuit performance under known conditions. The amount of charge deposited by the pre-charger is controlled using a DAC.

Independently of whether a particle detector or the pre-charger circuit is used to inject electrical charge, the end result is a step of charge at the input node of the Heisenberg IC, or analogously, a current impulse at the input branch. This current impulse is integrated by the charge-sensitive amplifier (CSA), generating an output voltage step of amplitude approximately equal to $V_{O,CSA} = Q_{in}/C_F$, where C_F is the CSA feedback capacitor.

Following the signal path, the output voltage of the CSA is then filtered by the pulse-shaping filter in order to increase the SNR, resulting in a semi-gaussian pulse at the filter output. The output is then sampled at the peak value of the semi-gaussian pulse, using an ADC.

The ADC output is a 16-bit bus that is connected to the FPGA in a development board. The FPGA communicates with a PC using the UART communication protocol, and passes down the signal-samples from the ADC to the PC for data processing.

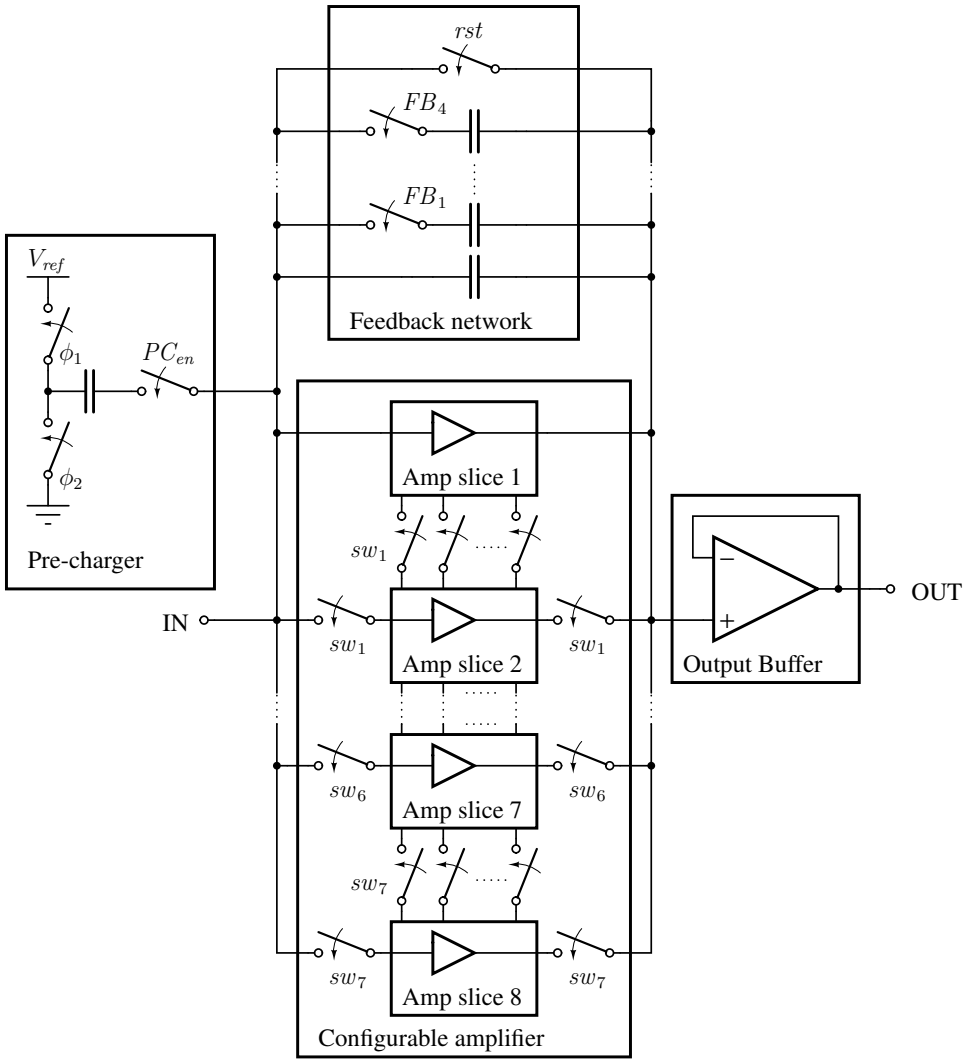


FIGURE 5.2. Heisenberg block diagram.

Besides performing data flow control between the ADC and the PC, the FPGA sets the control registers for the Heisenberg IC, DAC and ADC, while also controlling all the timing sequences for proper system operation.

5.2.1 The Heisenberg chip block diagram

Figure 5.2 shows the circuit block diagram of the Heisenberg chip. The circuit can be divided into four functional blocks: the pre-charger circuit, the configurable amplifier, the feedback network and the output buffer. The combination of the amplifier and the

feedback network is referred to as the charge-sensitive amplifier (CSA). Details of the circuit-level design of the CSA, the pre-charger circuit and the off-chip filter are presented in Chapter 6.

The configurable amplifier consists of eight identical amplifier slices that can be connected in parallel via switches. The connection of the different amplifier slices is done in thermometer mode, that is,

$$\{1\},\ \{1-2\},\ \{1-2-3\},\ ...,\ \{1-2-...-8\},$$

where the numbers corresponds to the numbering of the amplifier slices shown in Figure 5.2. Details of the parallel connection scheme are further expanded upon in Section 6.2.8.

The output node of the CSA is very sensitive to load capacitances, which can limit the system bandwidth, cause instability, and introduce output slewing. To avoid connecting directly the CSA output to an analog pad and wire bond on the chip, of relatively high ground-capacitance (without even considering other off chip parasitic effects), a voltage buffer is added on the signal path, as shown in Figure 5.2. For the output buffer, a pre-existing operational amplifier design was used, connected in buffer configuration.

5.3 System specifications

Table 5.1 shows a summary of the technical specifications of the Heisenberg test system. Since the system was tested without a particle detector, the negative effects that a large detector capacitance have on the noise performance of the system were emulated by using an explicit SMD capacitor soldered on the Heisenberg test board. A large capacitance value was selected in order to guarantee that the amplifier noise contribution is the dominant factor on the noise measurements of the filter output.

TABLE 5.1. Specifications summary.

Detector	None. Explicit capacitor soldered on board. Input capacitance: $C_D = 390\text{pF}$
Pre-charger	Reference capacitor: $C_{PC} = 1\text{pF}$ DAC: 12-bit, 3.3V reference Injected charge: 80.6fC min, 3.3pC max
CSA	Input signal: Up to 280pC Dynamic range: Open-loop gain $A_{ol} > 72\text{dB}$ over a 3.5V output swing Feedback capacitance: $C_F = \{1\text{pF}, 2\text{pF}, 3\text{pF}, 4\text{pF}, 5\text{pF}, 6\text{pF}, 7\text{pF}, 8\text{pF}\}$
Filter	Filter Topology: $CR - 2RC$ Peaking time: $\tau_P = 20\mu\text{s}$
ADC	Resolution: 16-bit Sampling rate: 2MSPS Reference: 4.96V (internal)

6. CIRCUIT-LEVEL DESIGN

6.1 Introduction

The present chapter delves into the circuit-level design of the Heisenberg chip, designed in a 0.5-$\mu\mathrm{m}$ CMOS technology, and test board. It details design considerations, equations and design parameters for the different functional blocks presented in the block diagram shown in Figure 5.1. Particularly, it details the design of the charge-sensitive amplifier (CSA), including the folded-cascode amplifier, the parallel connection scheme and the feedback network; the pre-charger circuit; and the pulse-shaping filter.

6.2 Charge-sensitive amplifier design

6.2.1 Amplifier static error and nonlinearity

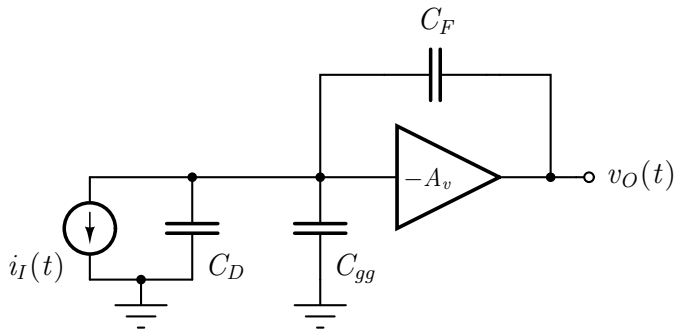


FIGURE 6.1. Generic CSA schematic for linearity analysis.

Consider the circuit schematic presented in Figure 6.1. This circuit shows a generic CSA of low-frequency open-loop gain A_v, feedback capacitance C_F, and input capacitance C_{gg}, connected to a detector represented by a signal source i_I and a detector capacitance C_D. Since this is a generic representation of a CSA, it is topology independent.

The input current is divided into three branches at the input node, flowing through capacitors C_D, C_{gg} and C_F. The current that flows through each capacitor is proportional to the apparent value of each capacitor seen from the input node. Capacitor C_F appears

to be larger due to the Miller effect, with a value of $(1 + A_v)C_F$. For very large values of A_v, the apparent value of C_F as seen from the input node is dominant over the other capacitors, so most of the current flows through capacitor C_F.

This same result, achieved through circuit inspection, is evidenced by proper circuit analysis. From the schematic in Figure 6.1, it can be shown that the relation between the output voltage and the input current is given by

$$\begin{aligned} V_O(s) &= \frac{I_I(s)}{sC_F}\left[\frac{\beta A_v}{1+\beta A_v}\right] \\ &= \frac{I_I(s)}{sC_F}\cdot\gamma_{ol} \end{aligned} \tag{6.1}$$

where

$$\beta = \frac{C_F}{C_D + C_{gg} + C_F} \tag{6.2}$$

Equation (6.1) shows that, for $A_v \to \infty$, $V_O(s) = I_I(s)/sC_F$, which can be interpreted as the circuit fully integrating the input current i_I in the feedback capacitor C_F. However, when the gain is finite, the current distributes between the different capacitors at the input node, reducing the amount of current integrated in the feedback capacitor and the amplitude of the output voltage.

Factor γ_{ol} in (6.1) represents the effects of the finite open-loop gain of the amplifier on the ideal transfer curve. For a limited output swing, where the amplifier gain remains relatively constant, the effect of γ_{ol} over the output value is a static error that can be corrected through post-processing. For a larger output swing, where the value A_v changes dynamically as the transistors instantaneous operating point changes, γ_{ol} can introduce nonlinearities. The effect of γ_{ol} can be mitigated by having a sufficiently large open-loop gain over the desired output dynamic range of the amplifier.

6.2.2 Frequency response

The schematic representation of the CSA in Figure 6.1 considered a voltage amplifier of infinite open-loop bandwidth and finite open-loop gain. Let us now consider an amplifier of finite bandwidth, represented by its effective transconductance G_{meff} and its output

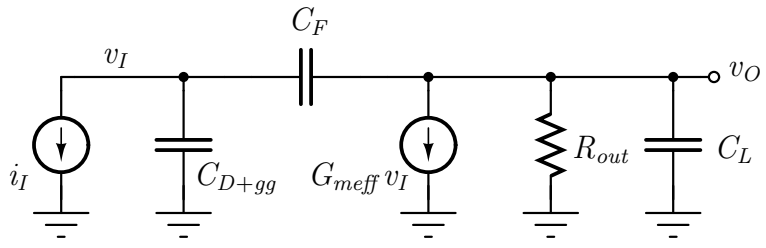


FIGURE 6.2. Schematic of a generic closed-loop CSA for frequency response analysis.

resistance R_{out}, both of finite value. Let us now assume that the frequency response of the open-loop amplifier is dominated by a single pole in the output, defined by the output resistance R_{out} and a load capacitance C_L. The resulting circuit is drawn in Figure 6.2. To simplify the notation, capacitors C_D and C_{gg} were added in parallel and represented as an equivalent capacitor C_{D+gg}.

The transfer function of the circuit in Figure 6.2 can be computed as

$$\frac{V_O(s)}{I_I(s)} = \frac{1}{sC_F} \cdot \gamma_{ol} \cdot \left[\frac{1 - s/z}{1 - s/p}\right] \tag{6.3}$$

where

$$\gamma_{ol} = \frac{G_{meff} R_{out} \cdot C_F}{(1 + G_{meff} R_{out}) \cdot C_F + C_{D+gg}} \tag{6.4}$$

$$z = \frac{G_{meff}}{C_F} \tag{6.5}$$

$$p = -\frac{(1 + G_{meff} R_{out}) \cdot C_F + C_{D+gg}}{R_{out}(C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg})} \tag{6.6}$$

The factor γ_{ol} is the same factor that appears in (6.1), and represents the effects of the finite open-loop gain of the amplifier on the output amplitude. Transfer function (6.3) shows that, besides the pole in the origin due to the integrating action of the closed-loop amplifier, the frequency response of the circuit in Figure 6.2 has a zero and a pole.

Considering a large open-loop gain $A_v = G_{meff} R_{out}$, (6.6) can be approximated as

$$p \approx -\frac{G_{meff} \cdot C_F}{C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg}} \tag{6.7}$$

Furthermore, for large capacitance detectors

$$p \approx -\frac{G_{meff} \cdot C_F}{(C_L + C_F) \cdot C_{D+gg}} \tag{6.8}$$

The closed-loop amplifier bandwidth, defined by the pole, sets the time constant of the step response of the amplifier. Since the filter sets the system peaking time, it is desirable that the CSA time constant is significantly smaller than the peaking time, so to not slow down the nominal speed of the system.

6.2.3 The folded-cascode amplifier

The charge-sensitive amplifier was designed around the folded-cascode topology due to its simplicity, in both circuit complexity and ease of analysis, and excellent gain and bandwidth. Furthermore, one of the main advantages that the folded-cascode configuration has over other amplifier topologies is the independence between the input and load branches, making it possible to set the transconductance of the input device to ensure that its noise contribution is dominant. This is particularly useful for low noise design in particle physics instrumentation where the ENC is the figure of merit.

Figure 6.3 shows the schematic of the single-ended, NMOS-input folded-cascode amplifier used in the design of the Heisenberg IC. It consists of 5 transistors: the input transistor M_I, the folding transistor M_F, the cascode for the input and folding transistors M_{CF}, the load transistor M_L, and the cascode for the load transistor M_{CL}.

Transistor M_F sets the total amplifier current, while transistor M_L sets the load branch current. The input transistor drain current is defined by the difference between the folding transistor drain current I_F and the load transistor drain current I_L. Current values of $I_F = 275\ \mu$A, $I_L = 25\ \mu$A and $I_I = 250\ \mu$A were used in the design of the NMOS-input amplifier slice in the Heisenberg IC.

In the folded-cascode topology, the output DC operating voltage, commonly referred to as signal baseline in particle physics instrumentation, is defined by the input device gate voltage when connected in a DC negative feedback configuration. This value is near V_{th}

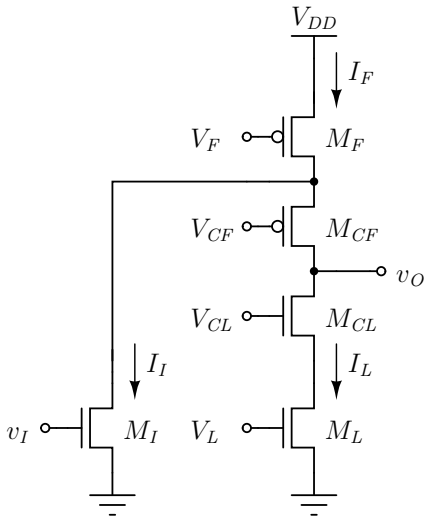


FIGURE 6.3. Schematic of a NMOS-input folded-cascode amplifier.

for an NMOS input device, and near $V_{DD} - |V_{th}|$ for a PMOS input device. A baseline near V_{th} for an NMOS-input CSA means it is better suited for applications that pull current from the feedback capacitor, given that this produces an output signal of positive polarity, and thus it has a larger output voltage swing. The opposite is true for a PMOS-input amplifier.

The schematic of the circuit shown in Figure 6.3 is simple enough to identify the small-signal operation of all devices by inspection. Transistor M_I is connected in a common-source (CS) configuration, so it is acting as a transconductor. Transistors M_F and M_L both have a constant gate-to-source voltage, so they operate simply as a load with a value defined by their output resistance. Transistor M_{CL} is the cascode for transistor M_L, so that both devices act together as an equivalent resistive load. Transistor M_{CF} acts as a cascode for the parallel between M_I and M_F. It acts as a current buffer, amplifying the apparent output resistance of the parallel between M_I and M_F, and collecting most of the current from the input transistor.

6.2.4 Equivalent transconductance, output resistance and open-loop gain

Figure 6.4 shows a schematic of the folded-cascode amplifier for small-signal analysis. The effective transconductance of the circuit G_{meff}, defined as the output current

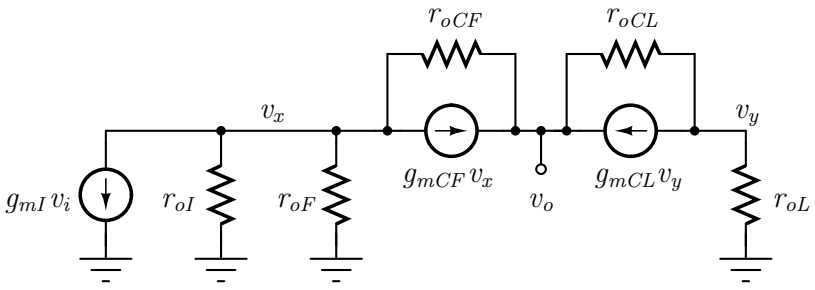


FIGURE 6.4. Schematic of the folded-cascode amplifier for small-signal analysis.

derivative $\partial I_O/\partial V_I$ when V_O is shorted to ground, is shown in (6.9). The output resistance of the circuit R_{out}, defined as the resistance seen from the output node of the amplifier when no input is applied, is shown in (6.10).

$$G_{meff} = g_{mI} \cdot \frac{(r_{oI} \parallel r_{oF})(1 + g_{mCF} \cdot r_{oCF})}{(r_{oI} \parallel r_{oF})(1 + g_{mCF} \cdot r_{oCF}) + r_{oCF}} \tag{6.9}$$

$$R_{out} = (r_{oL} + r_{oCL} + g_{mCL} \cdot r_{oCL} \cdot r_{oL}) \parallel ((r_{oI} \parallel r_{oF}) + r_{oCF} + g_{mCF} \cdot r_{oCF} \cdot (r_{oI} \parallel r_{oF})) \tag{6.10}$$

Assuming that all the resistance values are in the same order of magnitude, and that transistors M_{CF} and M_{CL} have a large intrinsic gain[1], the expressions for G_{meff} and R_{out} can be approximated by

$$G_{meff} \approx g_{mI} \tag{6.11}$$

$$R_{out} \approx (g_{mCL} \cdot r_{oCL} \cdot r_{oL}) \parallel (g_{mCF} \cdot r_{oCF} \cdot (r_{oI} \parallel r_{oF})) \tag{6.12}$$

Thus, the low-frequency, open-loop voltage gain of a folded-cascode amplifier can be approximated by

$$\begin{aligned} |A_v| &= G_{meff} \cdot R_{out} \\ &\approx g_{mI} \cdot [(g_{mCL} \cdot r_{oCL} \cdot r_{oL}) \parallel (g_{mCF} \cdot r_{oCF} \cdot (r_{oI} \parallel r_{oF}))] \end{aligned} \tag{6.13}$$

[1]The intrinsic gain of a transistor is defined as the product between the small-signal transconductance and the drain-to-source resistance of the transistor, $g_m \cdot r_o$.

6.2.5 Input-referred noise

The amplifier's input-referred noise can be expressed as a sum of the input-referred contributions of all individual transistors. The noise sources are referred to the input by their transfer functions, from noise source to input. The transfer function for all noise sources is calculated without considering parasitic capacitances for the different devices, assuming that the system frequency response is dominated by external factors. Thus, the total input-referred noise of the folded-cascode amplifier can be expressed as

$$\overline{V_N^2} = \overline{V_{NI}^2} \cdot N_I^2 + \overline{V_{NF}^2} \cdot N_F^2 + \overline{V_{NCF}^2} \cdot N_{CF}^2 + \overline{V_{NCL}^2} \cdot N_{CL}^2 + \overline{V_{NL}^2} \cdot N_L^2 \tag{6.14}$$

where $\overline{V_N^2}$ and N^2 are the gate-referred noise power and transfer function, respectively, of the different transistors in Figure 6.3. The gate-referred noise power $\overline{V_N^2}$ can include thermal and flicker noise processes.

The transfer function for each transistor on the CSA can be computed as

$$N_I = 1 \tag{6.15}$$

$$N_F \approx \frac{g_{mF}}{g_{mI}} \tag{6.16}$$

$$N_{CF} \approx \frac{1}{g_{mI} \cdot (r_{oI} \parallel r_{oF})} \tag{6.17}$$

$$N_{CL} \approx \frac{1}{g_{mI} \cdot r_{oL}} \tag{6.18}$$

$$N_L \approx \frac{g_{mL}}{g_{mI}} \tag{6.19}$$

Since g_{mI} is large by design, then

$$\frac{1}{g_{mI} \cdot (r_{oI} \parallel r_{oF})} \ll 1, \quad \frac{1}{g_{mI} \cdot r_{oL}} \ll 1 \quad \text{and} \quad \frac{g_{mL}}{g_{mI}} \ll 1,$$

so the noise contributions of M_F, M_{CL} and M_L can be ignored. The value of g_{mF} can be significant due to the large folding transistor current, so the dominant noise contributions are from M_I and M_F.

TABLE 6.1. Transistor design parameters for the charge-sensitive amplifier circuit.

Name	Type	L (μm)	W (μm)	g_m/I_D $\left(\frac{\text{mS}}{\text{mA}}\right)$	I_D (μA)
M_I	NMOS	0.6	127.5	16.4	250
M_F	PMOS	1.5	805.2	11.5	275
M_{CF}	PMOS	1.5	73.2	12.2	25
M_{CL}	NMOS	1.2	19.8	13.5	25
M_L	NMOS	1.2	19.8	12.7	25

6.2.6 Design parameters

The NMOS-input folded-cascode amplifier slice was designed using the g_m/I_D methodology. The design flow consisted on hand calculations assisted by g_m/I_D tables for coarse parameter values, an iterative optimization script for fine-tuning, and validation simulations. The design parameters for the folded-cascode amplifier are presented in Table 6.1.

The input-device g_m/I_D was selected based on the noise analysis presented in Chapter 3, particularly the observations made in Section 3.4.6, as a compromise between noise performance and power consumption in a white noise dominated system. The current ratio between the input and the load branches was selected so that the input-device noise contribution was dominant. With these constraints, the optimization was done in order to maximize the open-loop gain and bandwidth of the amplifier.

Circuit simulations were done using LTspice. Simulations show an open-loop gain of $77.85\ \text{dB}$ and an open-loop bandwidth of $762\ \text{kHz}$. Simulations also show that the output baseline, defined by the input-device gate-voltage V_{GS}, is set at $709\ \text{mV}$.

6.2.7 Folded-cascode bias circuit

Figure 6.5 shows a schematic of the bias circuit used to generate bias voltages $\{V_F, V_{CF}, V_{CL}, V_L\}$ for the folded-cascode amplifier. The design parameters for the bias circuit are presented in Table 6.2.

The bias circuit was included in each amplifier slice, instead of using a single bias circuit for the equivalent amplifier composed of all parallel-connected slices. This was

TABLE 6.2. Transistor design parameters for the amplifier bias circuit. The drain current for all transistors is $I_D = 25\ \mu$A.

Name	Type	L (μm)	W (μm)	Name	Type	L (μm)	W (μm)
M_1	PMOS	0.6	73.2	M_{11}	NMOS	0.6	19.8
M_2	PMOS	0.6	73.2	M_{12}	NMOS	0.6	19.8
M_3	NMOS	0.6	19.8	M_{13}	PMOS	0.6	73.2
M_4	NMOS	0.6	19.8	M_{14}	PMOS	0.6	73.2
M_5	PMOS	1.5	73.2	M_{15}	NMOS	1.2	3.9
M_6	PMOS	1.5	7.5	M_{16}	NMOS	1.2	19.8
M_7	NMOS	0.6	19.8	M_{17}	PMOS	0.6	73.2
M_8	NMOS	0.6	19.8	M_{18}	PMOS	0.6	73.2
M_9	PMOS	1.5	73.2	M_{19}	NMOS	1.2	19.8
M_{10}	PMOS	1.5	73.2	M_{20}	NMOS	1.2	19.8

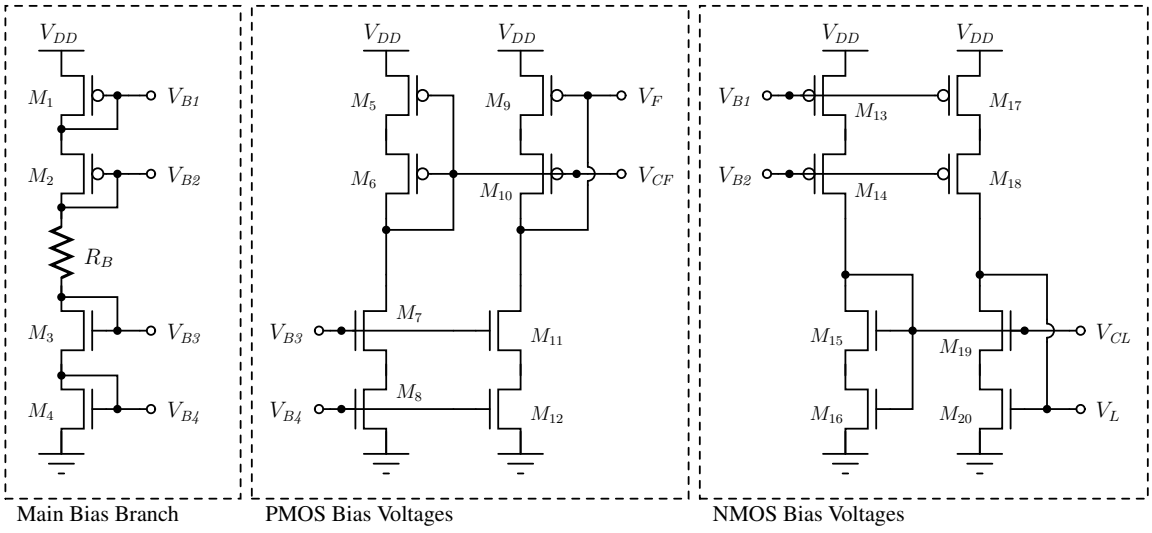


FIGURE 6.5. Schematic for the amplifier bias circuit.

mainly done to favor simplicity of implementation. A single bias resistor R_B was implemented off-chip, using a potentiometer, to bias all the amplifier slices at the same time. This means that all amplifier slices are always on, even when they are not connected in parallel.

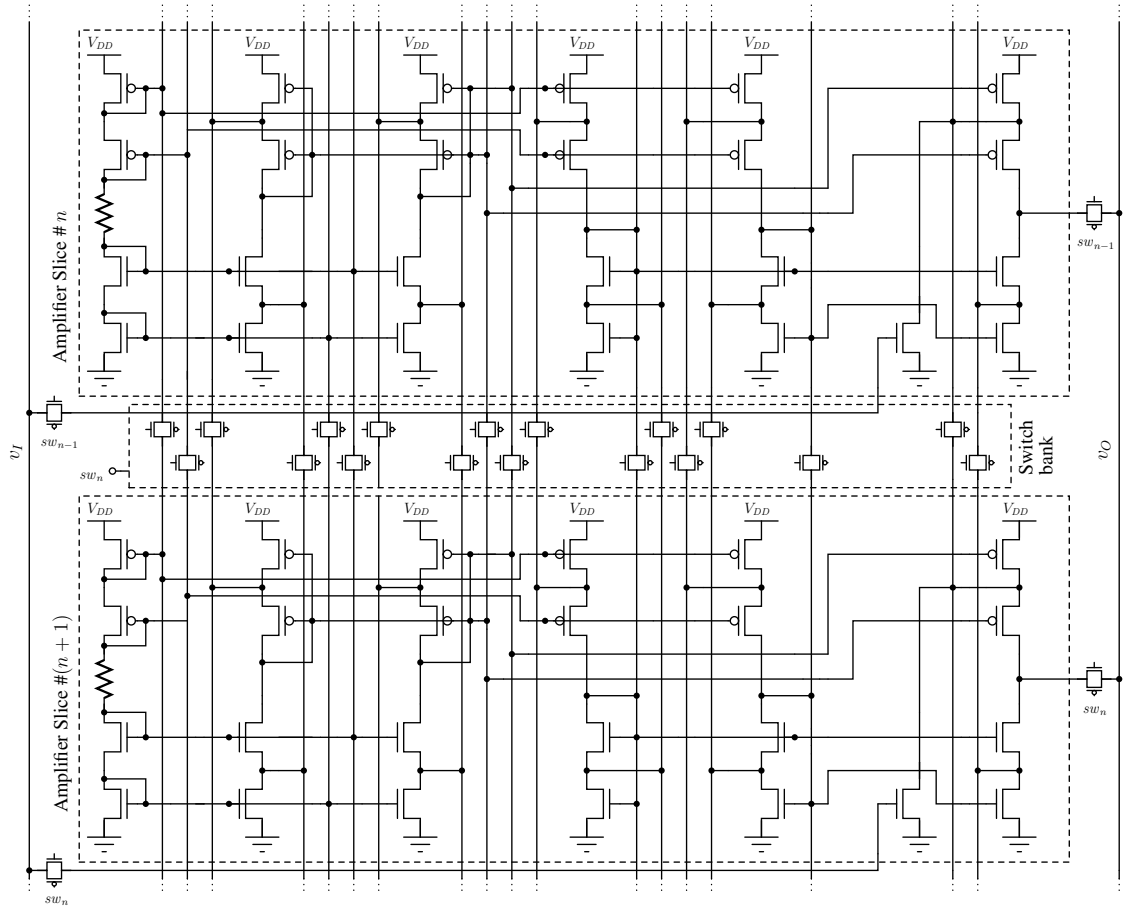


FIGURE 6.6. Parallel connection scheme.

6.2.8 Slice parallel connection

Figure 6.6 shows a schematic representation of the parallel-connection scheme used for the amplifier slices in the Heisenberg chip. All internal nodes of the amplifier, *i.e.* all nodes with the exception of the input and the output, are connected between adjacent slices using a switch bank of CMOS switches controlled via a single control signal sw_n, as shown in Figure 6.6. As for the input and output nodes, each slice is connected and disconnected from a common wire using CMOS switches, so that all amplifiers slices see the same signal path.

A single CMOS switch design of low series resistance was used for all the switches implemented in the Heisenberg chip, of widths $W_{NMOS} = 12$-μm and $W_{PMOS} = 36$-μm.

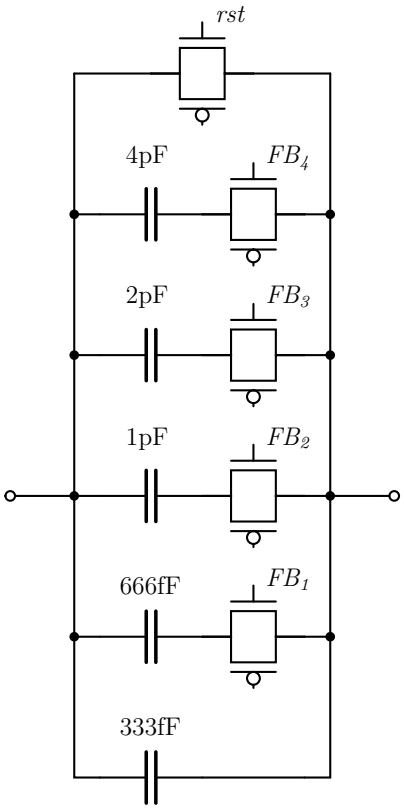


FIGURE 6.7. Schematic for the feedback network.

6.2.9 Feedback network

The feedback network, shown in Figure 6.7, includes the configurable feedback capacitor C_F and the reset switch M_{rst}. During CSA amplification, the reset switch remains open. Before each new pulse, the reset must be asserted, to discharge the feedback capacitor C_F and to bias the CSA output to signal baseline, of known value.

Control signals $\{FB_1, FB_2, FB_3, FB_4\}$ configure the value of the feedback capacitor C_F. All the capacitors in the network are individually controlled, hence it is possible to parallel-connect any combination to form an equivalent capacitor C_F of values ranging from a minimum of 333fF to a maximum of 8pF.

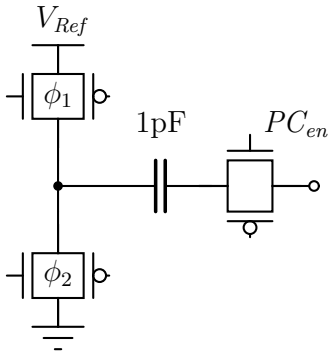


FIGURE 6.8. Schematic for the pre-charger circuit.

6.3 Pre-charger circuit

A pre-charger circuit is used to inject a known amount of electrical charge for testing or calibration purposes. In testing, it is used instead of a detector to characterize the circuit performance under known conditions.

Figure 6.8 shows the schematic of the pre-charger circuit. The circuit includes a 1-pF reference capacitor C_{PC}, and three switches, $M_{\phi 1}$, $M_{\phi 2}$ and M_{PCen}. Switch M_{PCen} can connect and disconnect the pre-charger from the signal path, in case it is not used.

In regular operation the pre-charger works in two distinct phases controlled by two non-overlapping clock signals, namely ϕ_1 and ϕ_2. During the first phase, switch $M_{\phi 1}$ is turned on, and the left plate of capacitor C_{PC} is tied to voltage V_{Ref}. At this stage, the capacitor is charged with an amount of charge equal to $Q = C_{PC}(V_{ref} - V_{GS})$, where V_{GS} is the input voltage of the CSA. Switch $M_{\phi 1}$ is then turned off, and the capacitor is left effectively floating, retaining the charge. During the second phase, the left plate of the capacitor is tied to ground, which causes a voltage drop in the right plate, given that the instantaneous charge in the capacitor is retained. Driven by the CSA feedback, the excess charge on the right side of the capacitor is removed as the voltage is driven back up to V_{GS}, discharging capacitor C_{PC} by an amount equal to $Q_{PC} = C_{PC}V_{Ref}$, and pulling charge Q_{PC} from the feedback capacitor. This produces a voltage variation on the output of the CSA of $C_{PC}V_{Ref}/C_F$.

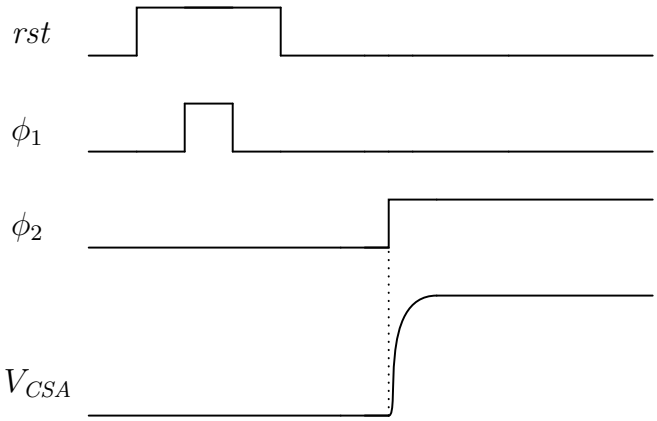


FIGURE 6.9. Signal diagram for pre-charging and feedback reset.

Figure 6.9 shows a signal diagram for clocks ϕ_1 and ϕ_2, feedback reset rst, and the CSA output voltage V_{CSA}. Phase ϕ_1 is asserted during feedback reset to avoid baseline shifts on the output.

It is possible to change the polarity of the injected charge, and thus that of the output step, simply by switching the phase order, *i.e.* first connecting the capacitor to ground, and then to V_{Ref}.

6.4 Pulse-shaping filter

The purpose of the filter in particle physics instrumentation is to increase the SNR by limiting the noise bandwidth, and to shape the output pulse to weight the parallel and series noise contributions in accordance to the application.

One commonly adopted type of linear time-invariant (LTI) filter used in particle physics instrumentation is the $CR - nRC$ bandpass filter. Figure 6.10 shows a generic schematic of one such filter, of order $n + 1$. This type of filters are simple in implementation and analysis, and have good noise performance. In fact, the most simple implementation, the $CR - RC$ filter, is only 36% worse in terms of SNR than the theoretical optimum filter, the *cusp* function (Spieler (2005)).

In a typical $CR - nRC$ filter, all the individual CR and RC filters have the same time constant τ. The transfer function of this filter can be described by the following

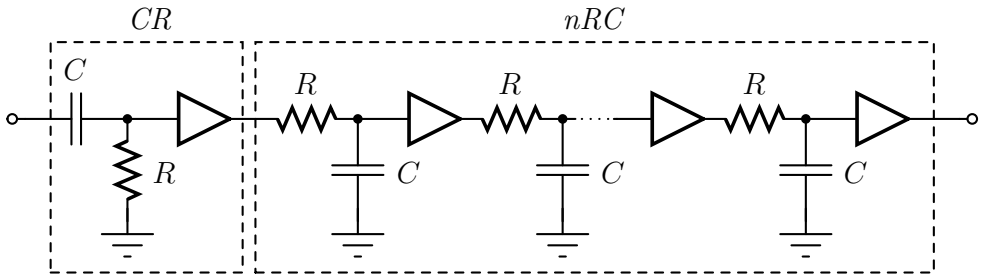


FIGURE 6.10. Schematic of a generic $CR - nRC$ filter.

expression:

$$H(s) = \left(\frac{s\tau}{1+s\tau}\right)\left(\frac{1}{1+s\tau}\right)^n \tag{6.20}$$

Since the input of the pulse shaper is typically a voltage step, the time response of the filter is better described by the step response, as follows:

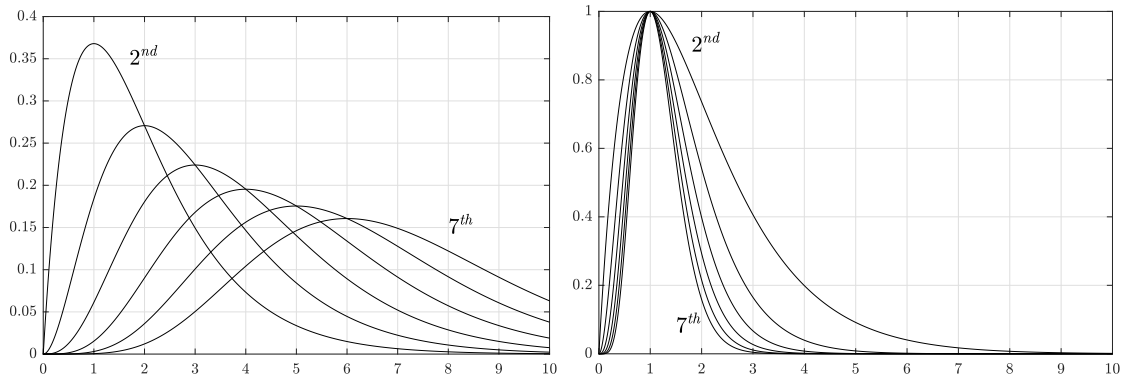
$$g(t) = \frac{1}{n!}\left(\frac{t}{\tau}\right)^n e^{-t/\tau} \tag{6.21}$$

The peaking time of a $CR - nRC$ filter is n times the time constant of the individual filters, *i.e.* $\tau_P = n\tau$. With this consideration, the peak value of the filter step response can be computed to be:

$$g(\tau_P) = \frac{n^n e^{-n}}{n!} \tag{6.22}$$

Figure 6.11(A) shows the step response of a $CR - nRC$ filter of time constant $\tau = 1$ for different values of n. The plot clearly illustrates both the proportionality of τ_P with n and the amplitude decrease with increasing values of n.

Figure 6.11(B) shows the time-normalized and amplitude-normalized step response of the filter for different values of n. From this plot it can be seen that there are other system-level benefits from using higher order filters other than potentially noise performance. In particular, higher order filters produce a narrower pulse response, which in turn produces a faster return to baseline. Return-to-baseline speed can be a critical system-level constraint depending on the system.



(A) Step response for $\tau = 1$.

(B) Step response with time and amplitude normalization.

FIGURE 6.11. Step response of a $CR - nRC$ filter for $n = \{1, 2, ..., 6\}$.

For the design of the Heisenberg test system a $CR - 2RC$ filter was used. Figure 6.12 shows the schematic of the filter, whereas Table 6.3 shows the values of the filter components. The circuit was implemented using discrete components on the Heisenberg test board.

The peaking time of the $CR - 2RC$ filter used is $20\mu s$. Two gain stages, of values $G_1 = 2.73$ and $G_2 = 1.36$, were added to compensate for the amplitude loss in the filter stages, by keeping the output peak of each RC filter stage near unity. These gain values were calculated directly using (6.21). The gain stages were implemented using non-inverting amplifiers, as shown in Figure 6.12. Amplifying the signal at this point in the signal path has no effect on the SNR, as it is dominated by the CSA noise performance. However, amplifying the signal allows to better use the dynamic range of the ADC that comes after the filter. Additionally, keeping the filters output peak of the same amplitude as the CSA output step allows for easy hand calculations during the testing stage.

TABLE 6.3. Filter parameter values.

Parameter	Value
C_{HP}	1-nF
C_{LP1}	1-nF
C_{LP2}	1-nF
R_{HP}	10-kΩ
R_{LP1}	10-kΩ
R_{LP2}	10-kΩ
R_{Gf1}	13-kΩ
R_{Gg1}	7.5-kΩ
R_{Gf2}	7.5-kΩ
R_{Gg2}	21-kΩ

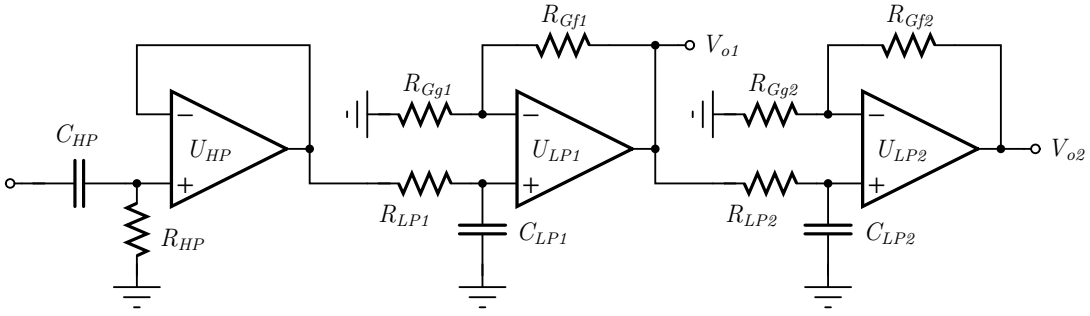


FIGURE 6.12. Schematic of the filter in the Heisenberg test board.

7. IMPLEMENTATION

7.1 Introduction

The present chapter describes the details of the Heisenberg test system implementation. Particularly, it details relevant information related to the chip layout, the test PCB layout, the HDL firmware, and the data analysis software used to testbench Heisenberg.

7.2 The Heisenberg chip layout

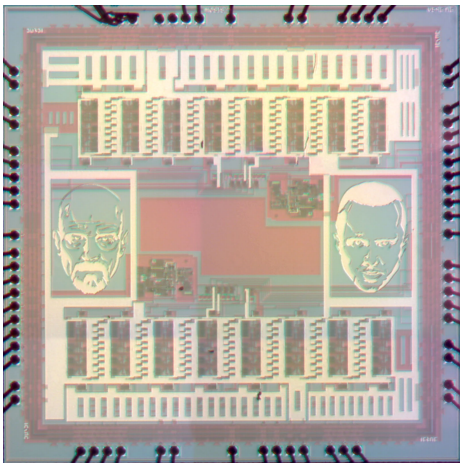


FIGURE 7.1. Heisenberg chip micrograph.

The Heisenberg chip was designed in a commercial 0.5-μm CMOS process. This technology has two polysilicon layers and three metal layers, and is meant for 5-V applications. The layout was designed using the Magic VLSI software (Magic VLSI (2018)). Figure 7.1 shows a die micrograph of Heisenberg.

Figure 7.2 shows the Heisenberg chip layout, of total dimensions $3\text{-mm} \times 3\text{-mm}$, including the pad frame. The effective dimensions for the circuit core are $2.55\text{-mm} \times 2.55\text{-mm}$. The chip has 70 connected pads, and uses a CQFP-100 package[1].

[1]CQFP-100: 100-pin Ceramic Quad Flat Package, a type of surface-mounted integrated-circuit packaging.

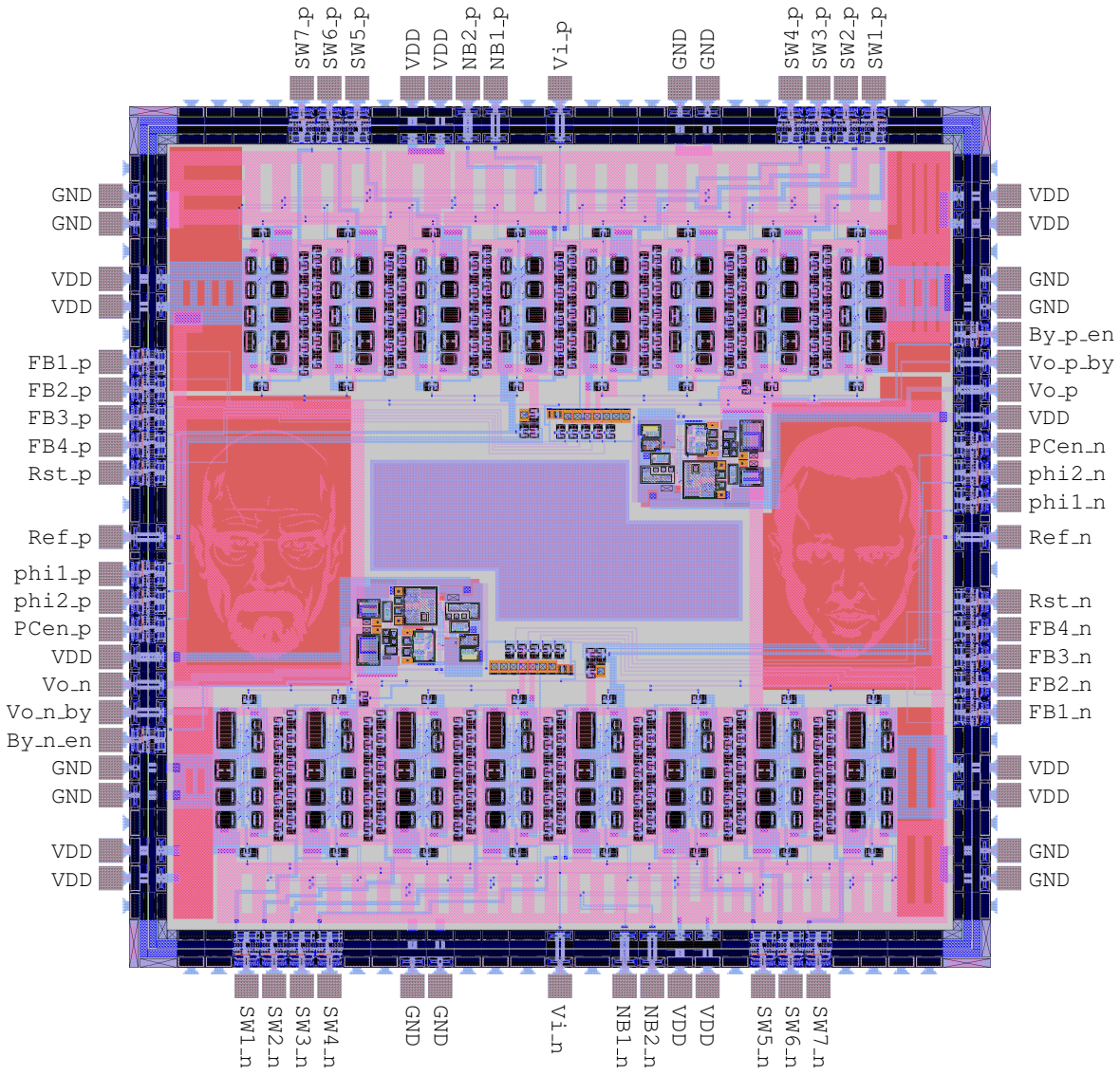


FIGURE 7.2. The Heisenberg chip layout. Die dimensions: 3-mm $\times$ 3-mm.

7.2.1 Floorplan

There are two channels in the Heisenberg chip, differing most importantly in the type of input device used. Figure 7.3 shows the chip floorplan for a single channel: the left side of the figure contains the configurable amplifier, consisting of individual amplifier slices and switch banks, whereas the right side contains the output buffer, the feedback network and the pre-charger circuit.

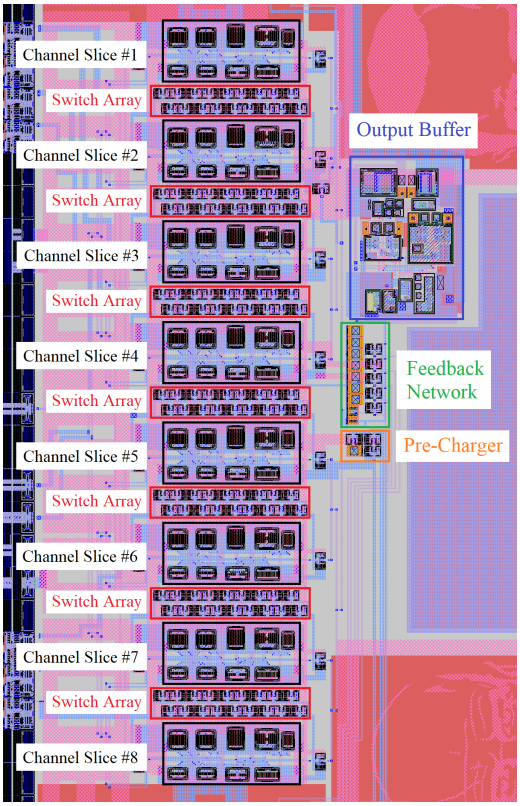
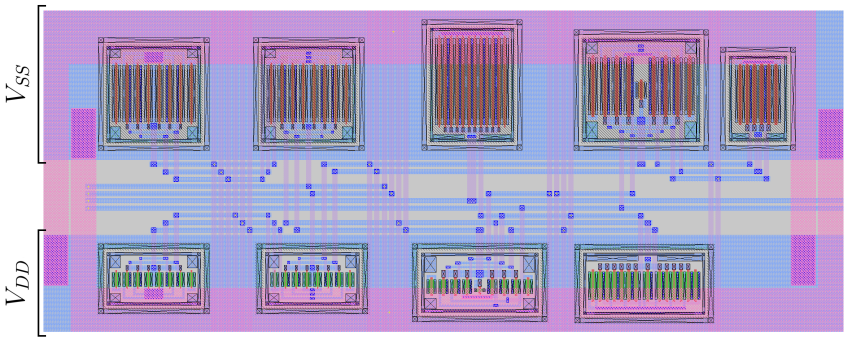


FIGURE 7.3. Heisenberg chip floorplan for a single PMOS based channel.

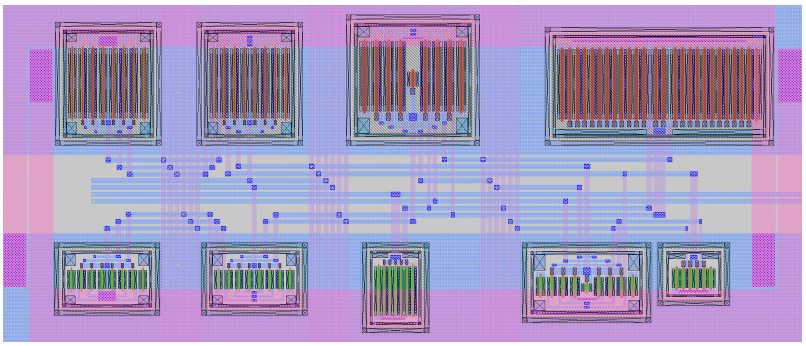
7.2.2 Amplifier slice and parallel connection

The amplifier slice layout was designed to include both the folded-cascode amplifier and the bias circuit, to match both amplifier and bias devices locally. The reference current for the amplifier cells is generated externally using a potentiometer. Figure 7.4 shows the layout for both the PMOS and NMOS-based amplifier slices. The dimensions of the amplifier slices are $402.6\text{-}\mu\text{m} \times 161.1\text{-}\mu\text{m}$ for the PMOS-based slice, and $421.5\text{-}\mu\text{m} \times 177.3\text{-}\mu\text{m}$ for the NMOS-based slice.

In both designs, all matched transistors are matched locally using interdigitated devices, and ended in dummies to reduce the effect of boundary conditions. Additionally, all sets of matched transistors are surrounded locally by guard rings connected to either V_{SS} or V_{DD}, for NMOS or PMOS devices respectively, to both provide signal shielding and to strongly tie the local substrate to the proper operating voltage.



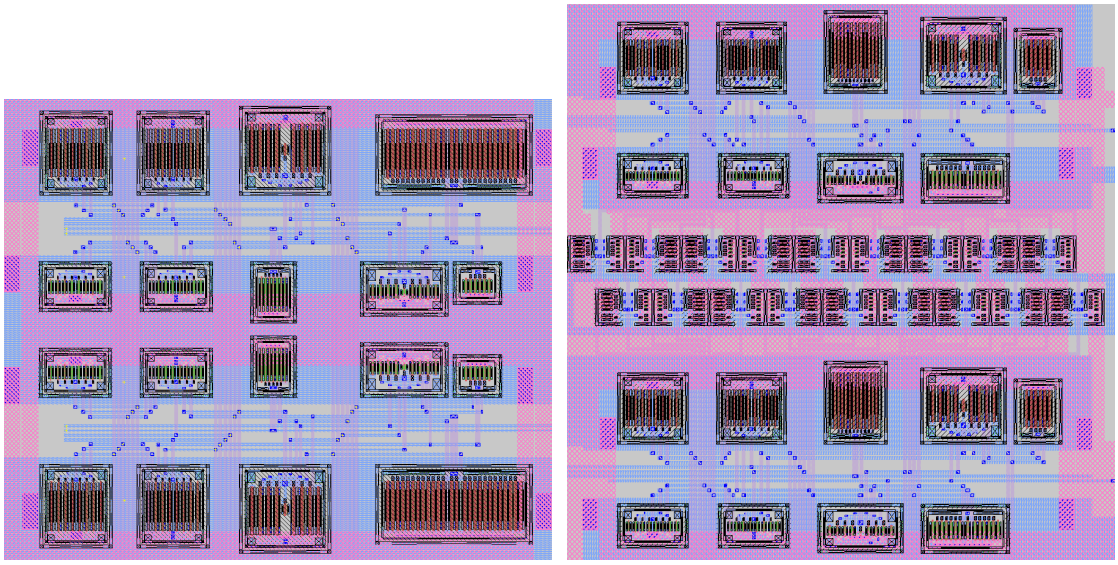
(A) PMOS-based amplifier slice.



(B) NMOS-based amplifier slice.

FIGURE 7.4. Amplifier slice layout for both PMOS and NMOS input device designs.

The layout of both types of amplifier slices share the same design philosophy. Inspired in digital cell layout design, the slice has clearly separated voltage rails in opposing sides, as shown in Figure 7.4(A). This allows for abutting the cells vertically, by mirroring the orientation of the cells. This is shown graphically in Figure 7.5(A) as an example, although this approach was not used in Heisenberg. All nodes have vertical tracks going through the amplifier slice to allow for an easy connection in a vertical stack, which can also be seen in Figure 7.5(A). This layout geometry also allows to potentially connect slices horizontally, and as such grow the equivalent circuit in both dimensions. One downside of this parallel connection approach is that the long vertical tracks increase the capacitance to ground on all nodes.



(A) Direct abutting by vertical mirroring.

(B) Configurable slice array using a switch bank.

FIGURE 7.5. Different approaches to slice stacking.

Figure 7.5(B) shows the approach used in the Heisenberg chip for the parallel connection of the different slices. A switch bank, with one CMOS switch corresponding to each node, is used as the interface between the slices. This allows to connect and disconnect the adjacent slices, and adjust the performance of the equivalent circuit accordingly. Since the corresponding voltage rails are physically separated between slices, no cell mirroring is necessary.

7.2.3 Feedback network, pre-charger and output buffer

Both the feedback network and the pre-charger circuit in Heisenberg consist only in CMOS switches and PIP[2] capacitors. Figure 7.6 shows the layout for both circuits. For the pre-charger, a single 1-pF capacitor was used as the reference. For the feedback network in the other hand, seven 1-pF and three 333-fF capacitors were used to configure the equivalent feedback capacitor.

[2]PIP: Polysilicon-Insulator-Polysilicon.

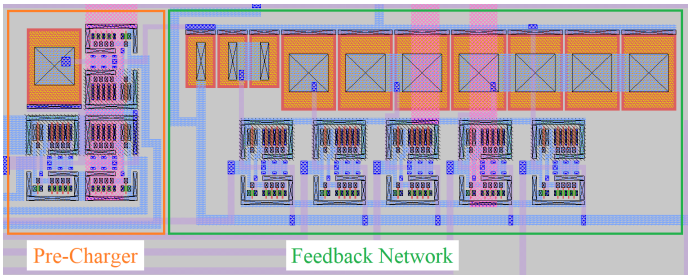


FIGURE 7.6. Feedback network and pre-charger circuit layout.

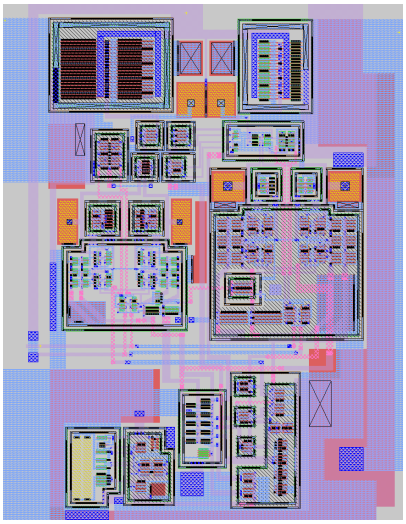


FIGURE 7.7. Operational amplifier cell layout.

For the output buffer, a pre-existing operational amplifier design was used, connected in a buffer configuration[3]. The operational amplifier cell layout is shown in Figure 7.7.

7.3 Heisenberg Test PCB

A custom PCB[4] was designed for the Heisenberg chip. The PCB layout design consists of four layers, in the following order: a top signal layer; an internal power layer; an internal ground plane; and a bottom signal layer. Figure 7.8 shows the test PCB layout, of dimensions 124.9-mm $\times$ 64.4-mm.

[3]Thanks to Diego Ávila and Hernán Campillo for lending me their OPAMP design.

[4]PCB: Printed Circuit Board.

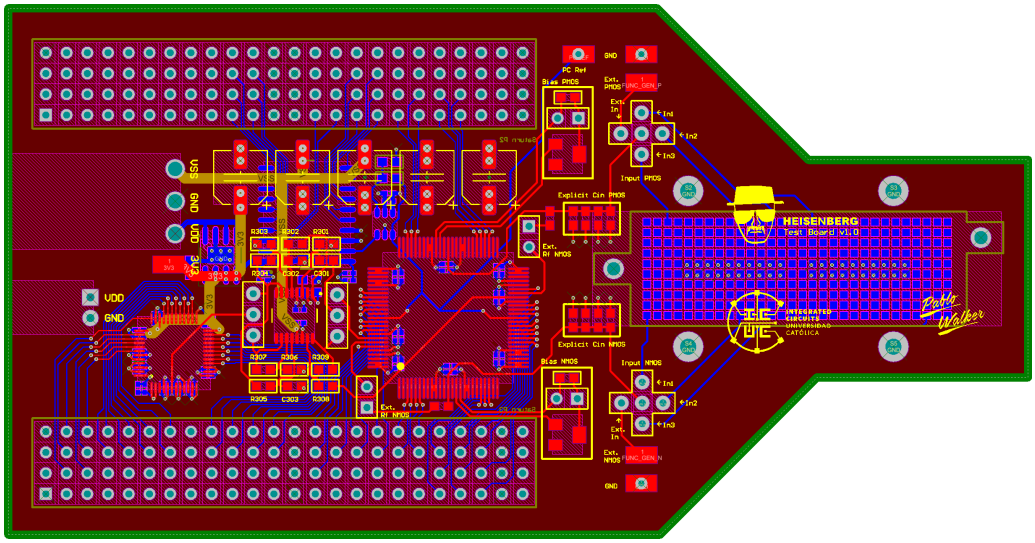
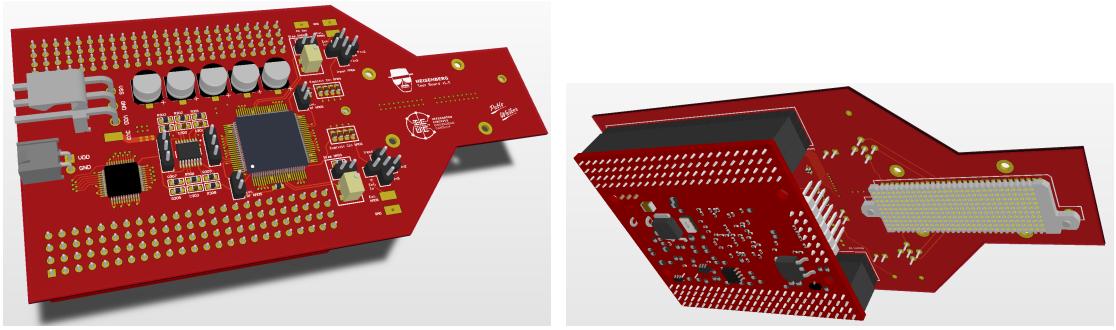


FIGURE 7.8. Heisenberg test board. Dimensions: 124.9-mm $\times$ 64.4-mm.

The Heisenberg test board is comprised by the following on-board devices: a low-dropout (LDO) linear voltage regulator (PN:[5] ADP1706)(Analog-Devices (2016)); bypass capacitors; the Heisenberg chip; bias potentiometers for the CSA channels; a pulse shaping filter including passive components and operational amplifiers (PN: LT6232)(Linear-Technology (2003)); a 16-bit 2MSPS ADC (PN: ADS8411)(Texas-Instruments (2004)) for sampling the filter output; and a 12-bit DAC (PN: DAC7621)(Texas-Instruments (1999)) that serves as the pre-charger reference. The board uses a dense multiple-pin connector (PN: GFZ-30-01-G-10-AD)(Samtec (2015)) for external inputs, as seen in the right-hand side of the board in Figure 7.8, with the option to alternatively simply use pin headers. Additionally, the board has two 96-pin header arrays, as seen in the left-hand side in Figure 7.8, to interface with an FPGA development board (PN: Numato Saturn)(Numato (2016)). A 3D model of the board can be seen in Figure 7.9(A).

Figure 7.10 shows the FPGA development board used for the tests. The board uses a Xilinx Spartan-6 FPGA (PN: XC6SLX16)(Xilinx (2011)), has 150 user I/Os, and uses USB for both FPGA programming and serial communication. The FPGA handles all the control signals for the Heisenberg chip, the reference DAC and the ADC, while also

[5]PN: Part Number



(A) Top view. (B) Bottom view.

FIGURE 7.9. 3D model of the board stack between the Heisenberg test PCB and the FPGA board.

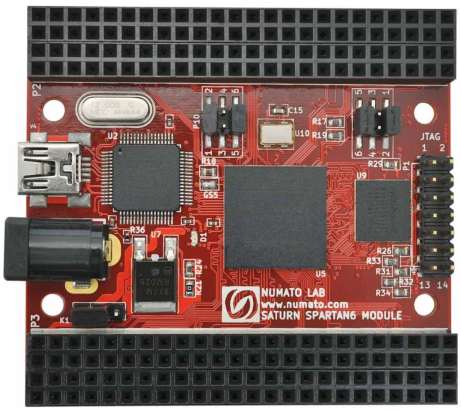


FIGURE 7.10. Numato Saturn - Spartan-6 FPGA Development Board. Dimensions: 62.8mm $\times$ 57.5mm.

reading the ADC output and dumping the data into a personal computer through USB. The FPGA board is stacked directly underneath the Heisenberg Test Board, as shown in Figure 7.9(B).

7.4 Firmware and data sampling

The FPGA firmware was written in Verilog and compiled using Xilinx ISE WebPack v14.7. The FPGA interfaces with four different devices: the Heisenberg chip, the ADC, the DAC, and a PC. Table 7.1 shows all the signal I/Os handled by the FPGA.

TABLE 7.1. FPGA digital I/Os.

Name	Device	Type	Description
UART_tx	PC	Output	UART transmit signal.
UART_rx	PC	Input	UART receive signal.
FB_n[4 − 1]	Heisenberg	Output	NMOS channel: Feedback network configuration.
SW_n[7 − 1]	Heisenberg	Output	NMOS channel: Slice connection configuration.
PCen_n	Heisenberg	Output	NMOS channel: Pre-charger enable.
By_en_n	Heisenberg	Output	NMOS channel: Output buffer bypass enable.
Rst_n	Heisenberg	Output	NMOS channel: CSA feedback reset.
phi_n[2 − 1]	Heisenberg	Output	NMOS channel: Pre-charger charge/discharge control.
FB_p[4 − 1]	Heisenberg	Output	PMOS channel: Feedback network configuration.
SW_p[7 − 1]	Heisenberg	Output	PMOS channel: Slice connection configuration.
PCen_p	Heisenberg	Output	PMOS channel: Pre-charger enable.
By_en_p	Heisenberg	Output	PMOS channel: Output buffer bypass enable.
Rst_p	Heisenberg	Output	PMOS channel: CSA feedback reset.
phi_p[2 − 1]	Heisenberg	Output	PMOS channel: Pre-charger charge/discharge control.
AD[15 − 0]	ADC	Input	ADC output.
AD_BUSY	ADC	Input	Status output. High when conversion in progress.
$\overline{\text{AD\_RST}}$	ADC	Output	Conversion abort. Sets ouput to zero.
AD_BYTE	ADC	Output	Byte select. Read in folded 8-bit mode or 16-bit parallel mode.
$\overline{\text{AD\_CONVST}}$	ADC	Output	Convert start.
$\overline{\text{AD\_RD}}$	ADC	Input	Read-synchronization pulse.
$\overline{\text{AD\_CS}}$	ADC	Output	Chip select.
DA[11 − 0]	DAC	Output	DAC input.
$\overline{\text{DA\_CS}}$	DAC	Output	Chip select.
DA_R/$\overline{\text{W}}$	DAC	Output	Read and write control.

The function of the FPGA in relation to the operation of Heisenberg is twofold: to set the static control signals to configure the feedback network and the number of connected slices, and to control the timing sequence for the pre-charger and feedback reset signals.

The aforementioned timing sequence consists in the following steps: the CSA feedback reset is engaged, the pre-charger is charged, the CSA reset is disengaged, the pre-charger is discharged, and finally the waveform is sampled. The ammount of charge deposited in the pre-charger capacitor is dependent on the reference voltage, which is set by the DAC, and controlled by the FPGA. The sampling of the waveform is done using the ADC, the timing of which is controlled by the FPGA to properly sample at the peaking time. Finally, the CSA feedback reset is done after the sample has been read by the FPGA. Figure 7.11 shows the most relevant signals in the sampling sequence. The timescales are

TABLE 7.2. Timing parameters.

Parameter	Duration [μs]
t_{w1}	3
t_{w2}	1
t_{w3}	250
t_{w4}	1
t_{d1}	1
t_{d2}	260
τ_P	20

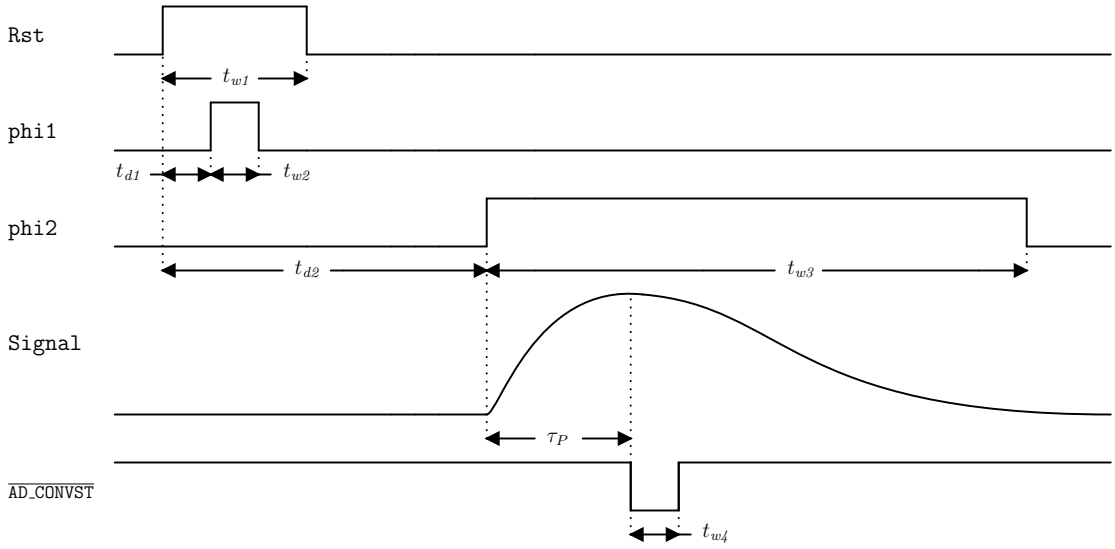


FIGURE 7.11. Timing for the operation of the pre-charger circuit and data sampling. The timescale is distorted to better display the signal sequence.

distorted to better display the signal sequence. The timing parameters used in the tests are displayed in Table 7.2.

The communication between the FPGA and the PC is done via USB. The FPGA development board has a USB-to-serial converter chip (PN: FT2232H)(FTDI (2016)), so the Verilog code is written to support the UART serial protocol. All the static control signals have a corresponding register in the Verilog code, which can be set externally by the PC. As for the timing signals, a periodic synchronous sequence can be triggered by setting other registers.

From the PC side, the control of the FPGA is done using MATLAB scripts. Mainly two scripts are used for gathering data: one configuration script and one sampling script. The configuration script is used to set the value of the feedback capacitor, the number of connected slices, the pre-charger enable and the DAC reference voltage. Additionally, another register can be set to trigger a periodic synchronous sequence on the FPGA to allow the user to observe waveforms in an oscilloscope. The sampling script, which is meant to be run after the configuration script, is used to configure the number of desired samples and to set the ADC registers on the FPGA to trigger data sampling. The PC serial buffer size is always set to a larger value than the number of desired samples, so that no samples are lost.

8. TEST RESULTS

8.1 Introduction

The present chapter provides an in-depth look at the most important results extracted from the measurements of the Heisenberg chip, specifically the configurable NMOS-input charge-sensitive amplifier (CSA).[1] Results related to the general functionality and the noise performance of the amplifier as a function of the number of parallel-connected slices are presented, and are contrasted with the expected performance of the CSA based on circuit analysis and simulations.

8.2 Noise PSD and bandwidth scaling

Before delving into the results of the Heisenberg chip, some preliminary comments related to the operation of the CSA are necessary in order to understand how noise was measured. As detailed in Section 2.2.2, when identical copies of an amplifier are connected in parallel, the bandwidth of the resulting circuit will be the same as that of the individual circuit copies only if both the equivalent capacitances and admittances of the dominant pole scale in the same proportion.

The above statement is not the case on a typical CSA, as the equivalent capacitance that sets the bandwidth of the amplifier is dominated by the very large external detector capacitance, as shown in Section 6.2.2. When CSA slices are connected in parallel, the effective transconductance of the resulting circuit increases proportionally, while the equivalent capacitance remains mostly unchanged. As a result, the bandwidth of the amplifier increases.

In Section 2.2.3 it was shown that the power spectral density (PSD) of noise decreases for an increasing number of parallel-connected slices. It follows that, if noise was measured directly at the CSA output, while the PSD of the noise would decrease, the

[1] Although two CSA channels were designed and implemented in Heisenberg, one for each MOS flavor, only the NMOS-input CSA was tested due to issues with the other channel.

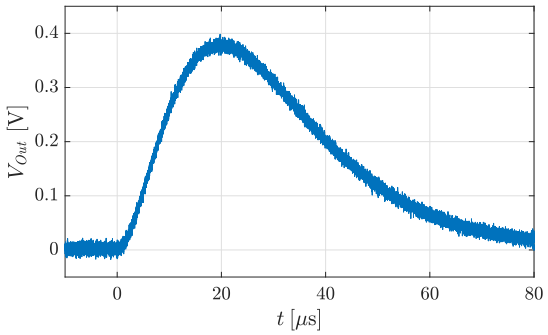


FIGURE 8.1. Example of a noisy waveform ($k = 2$, $C_F = 8$ pF).

bandwidth of the circuit would increase, resulting in no obvious improvement in the total integrated noise performance. This is not really a problem, however, as a typical CSA has a pulse-shaping filter connected in cascade, which limits the bandwidth of the circuit. Thus, in order to characterize the noise performance of the Heisenberg chip, noise was measured at the output of the $CR - 2RC$ pulse-shaping filter.

8.3 Input stimuli and noise measurements

It is a common practice to measure the noise of a circuit in the absence of input stimuli.[2] In the case of the Heisenberg test board, noise must be sampled at the filter output, which has no DC bias (*i.e.* it fluctuates between positive and negative voltages). The analog-to-digital converter (ADC) used to sample the noise has a unipolar reference, meaning that it cannot sample negative signals. This oversight meant that either the board had to be modified, or noise needed to be measured by inducing an output signal. For the sake of simplicity, the latter approach was selected, as the output of the filter is slow enough to not introduce additional problems.

The approach used to measure noise is better illustrated by Figure 8.1, which shows an example of a noisy waveform without a DC component. This is the exact type of waveform that was used to characterize the noise performance of the Heisenberg chip.

[2]The presence of input stimuli and output signal can even become a hindrance. For example, it leads to the necessity to sample at the exact same time for every output signal, to avoid superposing noise with amplitude variations, and can be influenced by the *jitter* of the sampling clock.

8.4 Pad leakage, feedback reset and baseline

While measuring the output waveform of the CSA it was observed that, on a large timescale, the output DC voltage, also known as the baseline, was discharging with a constant slope. This is an indication that there is a constant current source on one of the nodes of the feedback capacitor, and it was concluded that pad leakage on the input node is the most likely explanation.

In integrated circuits, pad leakage is an undesired effect caused by the leakage current of the protection diodes of a pad, which manifests as a sink or source constant current, depending on the dominant diode, on that particular node.

The value of the baseline is defined by the gate-to-source voltage (V_{GS}) of the input transistor of the CSA when the feedback reset is asserted, after which it starts discharging due to pad leakage. The presence of pad leakage means that, to assure that the output transistors are properly biased, the feedback reset needs to be asserted not long before signal is injected with the pre-charger.

It is possible to compute a simple expression for the behavior of the baseline to better understand which factors have an influence over it. Let us consider the circuit model presented in Figure 6.1 of Section 6.2.1, with a constant current source as the input to represent the pad leakage. From this circuit, it is possible to compute the following expression:

$$V_O(t) = V_{GS} - \left(\frac{I_P}{C_F}\right) \cdot \gamma_{ol} \cdot t \cdot u(t) \tag{8.1}$$

$$\gamma_{ol} = \frac{A_v C_F}{C_D + C_{gg} + (1 + A_v)C_F} \tag{8.2}$$

where I_P is the leakage current of the pads, $u(t)$ is the Heaviside step function, and the moment $t = 0$ is when the reset is released. This expression is not always applicable, since it was calculated through small-signal analysis, and thus it will no longer be valid when the output transistors are improperly biased.

The formula (8.2) shows that the baseline voltage starts at V_{GS}, and then discharges with a constant slope, which is a function of the magnitude of the pad leakage, the open-loop gain of the amplifier, and several circuit capacitances.

Let us assume that $\gamma_{ol} \approx 1$, which is true for an amplifier with a large open-loop gain. The slope of the baseline voltage with pad leakage is $-I_P/C_F$. Thus, if the same sampling time is used between different measurements, it is to be expected that the baseline will be lower for smaller values of C_F, *i.e.* the baseline discharges faster for a smaller feedback capacitance.

8.5 Test methodology

In order to test the NMOS-input CSA on the Heisenberg chip, two types of measurements were defined: average waveform measurements to assess general functionality, and statistical measurements of the output voltage to characterize noise performance. All measurements were repeated multiple times for the different configurations of two independent variables, namely for a varying number of parallel-connected slices from 1 to 8 (referred to as k), and for different values of the feedback capacitance (C_F).

The average waveform measurements were performed by injecting electrical charge into the CSA using the pre-charger circuit, the operation of which was described in Section 6.3. An oscilloscope was used to measure signal waveforms both at the CSA output and at the $CR - 2RC$ filter output. Several thousand events were averaged in order to filter out noise and obtain an accurate representation of the systematic response of the circuit.

The noise measurements were done by sampling the output of the $CR - 2RC$ filter with an ADC in the presence of input stimuli generated by the pre-charger circuit, in order to force the output signal within the limits of the ADC references.[3] Figure 8.1 shows an example of the kind of signals that were fed to the ADC. The signal was sampled at the filter peaking time ($\tau = 20\mu$s). Several tens of thousand samples were taken for each

[3]The noise measurements can be a time-consuming process. The use of a dedicated ADC, instead of an oscilloscope, allows for better low-level integration with custom FPGA firmware, which helps expedite the process.

different combination of k and C_F to obtain enough statistics to characterize the noise performance of the charge-sensitive amplifier.

As signal propagates through the amplifier, amplitude changes can have an effect on the instantaneous g_m/I_D of the transistors and consequently on their noise spectra. This effect might be negligible if signal amplitude is small. Nonetheless, it is desirable to maintain roughly the same amplitude between the different measurements for a fair comparison of the noise performance. And since the feedback capacitance has a direct effect on signal amplitude, it was necessary to adjust the amount of deposited charge when testing different values of C_F to obtain roughly the same amplitude.

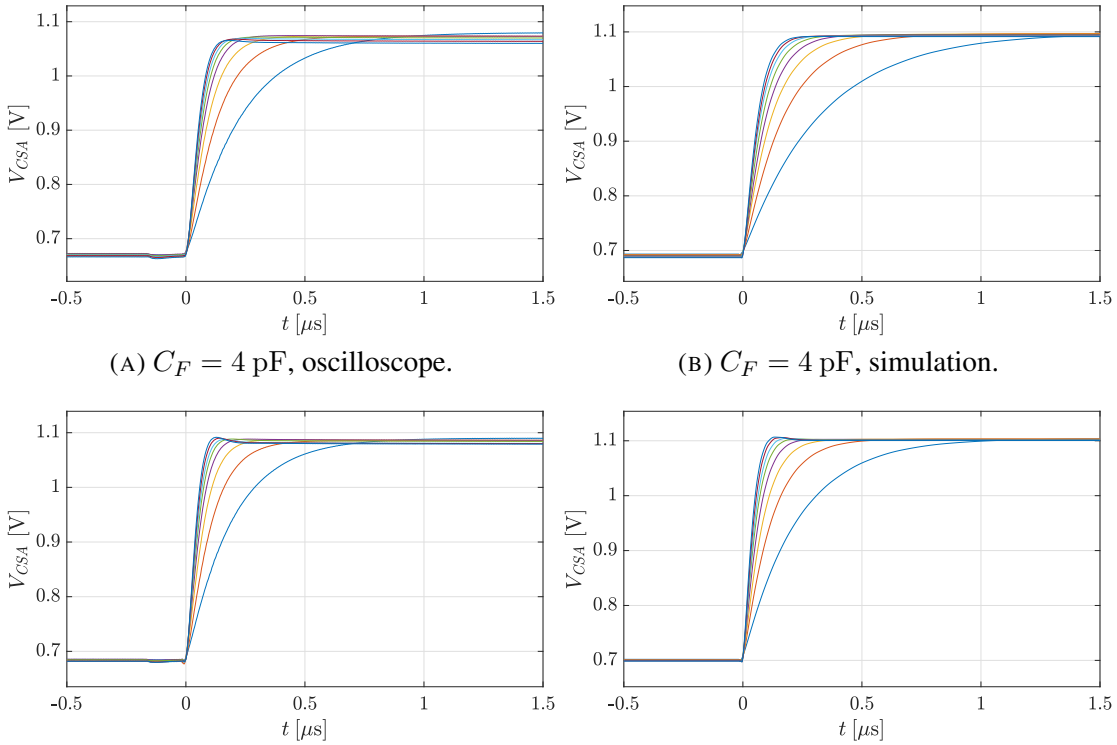
8.6 CSA step response

8.6.1 Setup

An internal voltage buffer connects the configurable CSA output to one of the analog pads of the Heisenberg chip, which was measured using an oscilloscope to analyze the average waveform of the circuit. The measured waveforms were obtained by averaging $8{,}192$ identical events to remove noise, whereas the simulated waveforms were obtained from post-layout simulations.[4]

The amount of electrical charge injected into the CSA by the pre-charger is given by $Q_{PC} = C_{PC}V_{ref}$, where C_{PC} (which has a value of $1\ \mathrm{pF}$) is the pre-charger capacitance, and V_{ref} is the reference voltage set by a DAC on the Heisenberg Test Board. For the results presented in the following section, the reference voltage values were set at $1.6\,\mathrm{V}$ and $3.2\,\mathrm{V}$, for feedback capacitance values of $4\ \mathrm{pF}$ and $8\ \mathrm{pF}$, respectively. The resulting amount of deposited charge produces an expected step amplitude of $400\ \mathrm{mV}$ in both scenarios, considering nominal component values.

[4]Post-layout simulations are simulations performed on a netlist extracted from the circuit layout, to account for parasitic effects that might be present on the physical layout of the circuit and not on the nominal design.



(A) $C_F = 4$ pF, oscilloscope.

(B) $C_F = 4$ pF, simulation.

(C) $C_F = 8$ pF, oscilloscope.

(D) $C_F = 8$ pF, simulation.

FIGURE 8.2. Charge-sensitive amplifier output for different number of parallel-connected slices. Both measured waveforms and simulation waveforms are presented, for different values of C_F. The measured waveforms were obtained using an oscilloscope, and averaging 8192 identical events for each value of k to remove noise. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$.

8.6.2 Results

Figure 8.2 shows the output of the NMOS-input charge-sensitive amplifier for both measurements and simulations. There are three key aspects of the measured step waveforms that are worth analyzing: baseline, amplitude and bandwidth. A brief summary of the expected behavior of these three aspects for the response of a typical folded-cascode CSA is presented below:

- **Baseline**: The baseline or DC bias of the voltage step is defined by the gate-to-source voltage V_{GS} of the input device of the CSA when the reset is engaged.

- **Amplitude**: The expected step amplitude for an injected charge Q_I is given by $V_O = \gamma_{ol} \cdot Q_I / C_F$, where $\gamma_{ol} \approx 1$ for an amplifier with a large open-loop gain.
- **Bandwidth**: The dominant pole for large capacitance detectors is given by

$$p = -\frac{G_{meff} \cdot C_F}{(C_L + C_F) \cdot C_D},$$

  which means that, for fixed capacitance values, the bandwidth of the amplifier increases for increasing values of the effective transconductance G_{meff}.

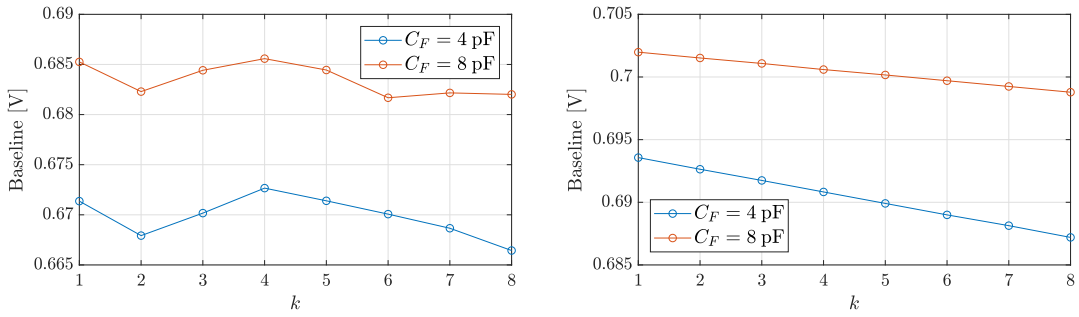
A first-order analysis of the plots shown in Figure 8.2 reveals that the circuit operates as expected. The baseline is similar between the different plots, the amplitude is roughly 400 mV, and both of them do not change drastically for different numbers of parallel-connected slices. Conversely, the bandwidth of the amplifier notably increases as more slices are connected in parallel, since the effective transconducatance of the amplifier increases as well.

A more detailed analysis reveals that there are some small but noticeable differences in the behavior of the baseline, amplitude and bandwidth when comparing the measured results to both the expected behavior and simulation results, which are explored in the following sections.

8.6.3 The baseline

The baseline is one of the least consequential aspects of the operation of a charge-sensitive amplifier, and it is only relevant in that it limits the output swing. As such, the results related to the baseline do not warrant extensive analysis. Nonetheless, a closer inspection of the baseline does allude to the presence of noticeable device mismatch, which will become relevant in the analysis of other results.

Figure 8.3 shows a close-up view of the baseline values of the plots presented in Figure 8.2. Discrepancies in absolute baseline values between simulation and measurements are of little significance, as there are many reasons why these discrepancies are expected to occur. These might include process variations, device mismatch, bias differences, imperfect simulation models, among others.



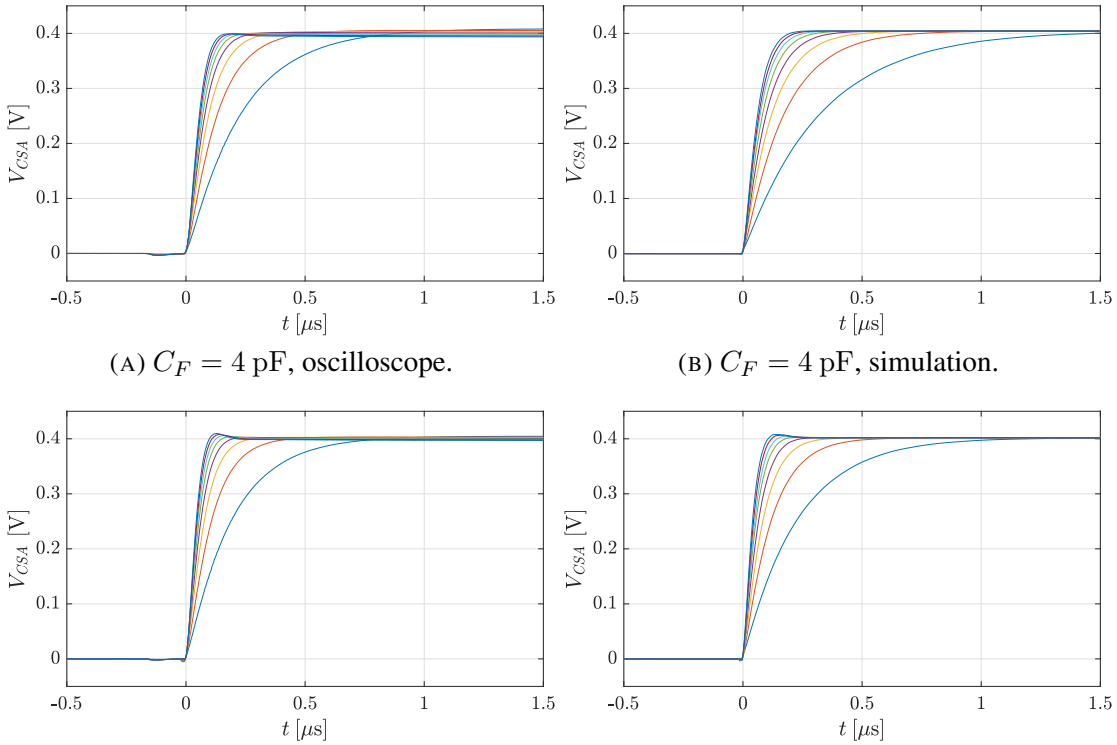
(A) Oscilloscope. (B) Simulation.

FIGURE 8.3. Close-up of the baseline values just before the CSA step response.

Furthermore, the relative baseline differences within each plot in Figure 8.3 as a function of the feedback capacitance are also of little significance to the present analysis. As it was alluded to in Section 8.4, because of the presence of pad leakage both in measurements and simulations, baseline value differences are to be expected for different values of C_F, since the baseline discharges faster for smaller values of the feedback capacitance.

The relative baseline differences within each plot in Figure 8.3 as a function of the number of parallel-connected slices are harder to explain. On the model shown in Section 8.4, the only term affected by k in (8.1) and (8.2) is the gate capacitance C_{gg}, which scales proportionally as more slices are connected in parallel. This model is not sufficient to explain the behavior of the baseline as observed in either measurement or simulation. It is likely that a more complex model is necessary to explain these differences.

Nonetheless, some relevant observations can still be made. Particularly, in Figure 8.3(B), it can be observed that there is a monotonic decrease in the baseline as a function of the number of parallel-connected slices, while in Figure 8.3(A) the behavior is non-monotonic. It is likely that the behavior observed in the measured results is a combination of the negative slope observed in the simulations and device mismatch on the input transistor of the amplifier. As slices are connected in parallel, the gate-to-source voltage (V_{GS}) of the equivalent input device changes due to size and gradient-related mismatch, which then sets the baseline value when the feedback reset is asserted.



(A) $C_F = 4$ pF, oscilloscope.

(B) $C_F = 4$ pF, simulation.

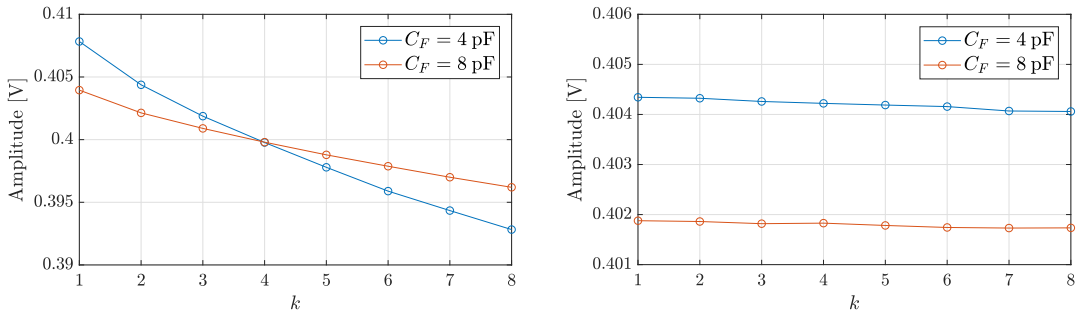
(C) $C_F = 8$ pF, oscilloscope.

(D) $C_F = 8$ pF, simulation.

FIGURE 8.4. Charge-sensitive amplifier output for different number of parallel-connected slices, without baseline. Both measured waveforms and simulation waveforms are presented, for different values of C_F. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$.

8.6.4 Step Amplitude

The step amplitude, unlike the baseline, is a fundamental aspect of the operation of a charge-sensitive amplifier. It is directly proportional to the amount of deposited charge, and consequently, to the amount of energy of an event in a particle physics experiment. Therefore, it is important to understand the mechanisms that might have an effect on signal amplitude, in order to calibrate during pre-testing or correct for them during data post-processing.



(A) Oscilloscope. (B) Simulation.

FIGURE 8.5. Close-up of the amplitude values for the CSA step response. These values are limited to the AC amplitude, and do not include the baseline. Furthermore, these values were sampled when the waveform reached steady-state.

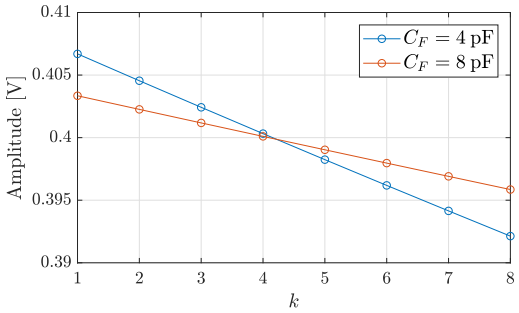


FIGURE 8.6. Plot for $V(k) = Q_I/(C_F + k \cdot C_{FP})$, where Q_I and C_{FP} were obtained through linear regression of the measured amplitude results. The estimated values are $Q_I = 1.635$ pC and 21.34 fF for the 4 pF curve, and $Q_I = 3.235$ pC and 21.69 fF for the 8 pF curve.

Figure 8.4 shows the same plots presented in Figure 8.2 but without the baseline, *i.e.* only the AC signal. Additionally, Figure 8.5 shows a close-up view of the step amplitude, which was measured after the signal reached the steady-state.

Similarly to the analysis of the previous section, absolute amplitude differences between simulation and measurements are not significant to the present analysis, as these are mostly influenced by external factors, such as the DAC voltage. Moreover, voltage differences within each plot with respect to the feedback capacitance are also not important, as the amount of injected charge between the two curves is different, as it was explained in

Section 8.6.1. Instead, a special emphasis will be placed in relative amplitude differences within each plot with respect to the number of parallel-connected slices.

Before delving into an analysis of the results, it is worth reiterating the expression that describes the amplitude behavior of the charge-sensitive amplifier. From Section 6.2.1, considering a generic implementation of a CSA, and a current impulse (*i.e.* a step of charge) as the input stimulus, it is possible to derive a simple expression for the output voltage of the amplifier:

$$v_O(t) = \frac{Q_I}{C_F} \cdot \gamma_{ol} \cdot u(t) \tag{8.3}$$

$$\gamma_{ol} = \frac{A_v C_F}{C_D + C_{gg,T}(k) + (1 + A_v)C_F} \tag{8.4}$$

where γ_{ol} represents the effect of the finite open-loop gain of the amplifier, which produces an static error on the output. The total gate capacitance is expressed as $C_{gg,T}(k) = k \cdot C_{gg}$, to emphasize its dependence on the number of parallel-connected slices. This expression is unrealistic in order to analyze the time response of the amplifier, but it is sufficient to understand the behavior of the amplitude.

It can be observed from Figure 8.5(B) that there is a very small amplitude decrease as more slices are connected in parallel. There is roughly a $280\ \mu\mathrm{V}$ amplitude difference between $k = 1$ and $k = 8$ for $C_F = 4\ \mathrm{pF}$, which is the most significant variation shown in the plot. The static error cannot explain this amplitude difference, since it has only a very weak dependency to the number of parallel-connected slices in the form of the total gate capacitance. Nonetheless, the amplitude variations are small enough to not pose a problem during the design process, and do not warrant further analysis.

In contrast, there is a much more significant amplitude difference observable in Figure 8.5(A) as a function of the number of parallel-connected slices. The amplitude difference between $k = 1$ and $k = 8$ for $C_F = 4\ \mathrm{pF}$ is roughly $15\ \mathrm{mV}$, the largest observable difference in the plot. This can be explained through the presence of a parasitic component to the feedback capacitance, which increases in size as more slices are connected in parallel.

Let us consider (8.3), and for the sake of simplicity, let us assume that the parasitic contribution of each individual slice is the same, even though the layout of the connection of each slice to the feedback network is different. With this parasitic component, (8.3) can be rewritten as

$$v_O(t) = \frac{Q_I}{C_F + C_{FP,T}(k)} \cdot \gamma_{ol} \cdot u(t) \tag{8.5}$$

where $C_{FP,T}(k) = C_{FP} \cdot k$, and C_{FP} is the parasitic contribution of each additional slice connected. This parasitic component also appears inside the static error factor γ_{ol} in parallel with C_F, but the effect is comparatively minuscule.

Only the first factor in (8.5) is relevant to understand the behavior of the amplitude, as it is dominant over the static error, which can be assumed to be $\gamma_{ol} \approx 1$ for the sake of simplicity. It can be observed that the reciprocal of this first term is linear, and can be written as

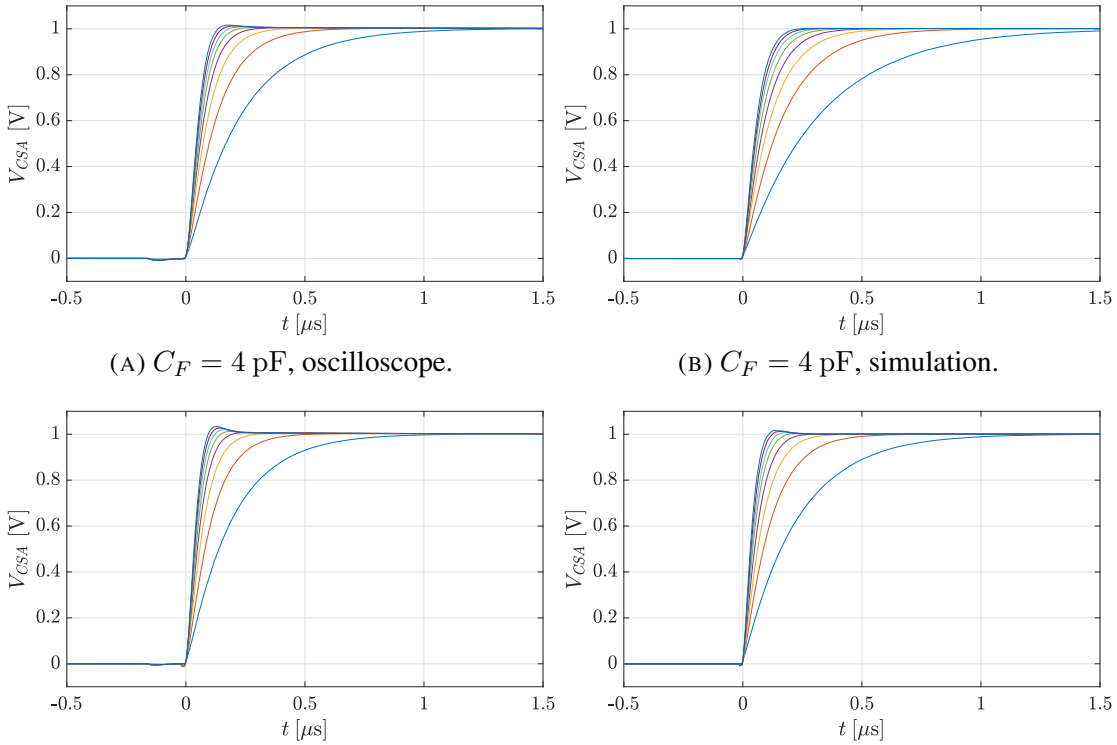
$$\text{Amplitude}^{-1} = \frac{C_F}{Q_I} + \frac{C_{FP}}{Q_I} \cdot k$$

From this expression, and through simple linear regression from the data shown in Figure 8.5(A), a best-fit value for Q_I and C_{FP} can be estimated for both curves. Figure 8.6 shows a plot for the amplitude term from (8.5) for these estimators. The estimated values of C_{FP} are 21.34 fF and 21.69 fF, for the $C_F = 4$ pF and the $C_F = 8$ pF curves, respectively. Given the similarities between Figures 8.5(A) and 8.6, it becomes clear that the amplitude decrease can be mostly attributed to a roughly 21.5 fF parasitic component to the feedback capacitance, added with each additional parallel-connected slice.

In principle, these parasitic effects should also be observable in the simulation results. However, due to the small size of each individual contribution to the feedback capacitance, these fall under the size cut-off for device extraction by the post-layout netlist generator, which is set to 100 fF.

8.6.5 Bandwidth

The importance of the bandwidth of a charge-sensitive amplifier is primarily in its interaction with the bandpass pulse-shaping filter, which limits the bandwidth of the system



(A) $C_F = 4$ pF, oscilloscope.

(B) $C_F = 4$ pF, simulation.

(C) $C_F = 8$ pF, oscilloscope.

(D) $C_F = 8$ pF, simulation.

FIGURE 8.7. Charge-sensitive amplifier output for different number of parallel-connected slices, without baseline and normalized to unit amplitude. Both measured waveforms and simulation waveforms are presented, for different values of C_F. In all cases, there is a monotonic increase in signal bandwidth from the rightmost plot corresponding to $k = 1$, to the leftmost plot corresponding to $k = 8$.

to minimize noise. As long as the amplifier is significantly faster than the filter, the exact bandwidth of the amplifier is irrelevant.

Figure 8.7 shows the same plots presented in Figure 8.2, but without baseline, and their final value normalized to unit amplitude. These plots focus exclusively in the shape of the waveform, and their relative differences with respect to the number of parallel-connected slices.

As described in Section 6.2.2, a typical CSA can be approximated to be a single-pole circuit, and the location of the pole is given by:

$$p \approx -\frac{G_{meff} \cdot C_F}{C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg}} \approx -\frac{G_{meff} \cdot C_F}{(C_L + C_F) \cdot C_{D+gg}} \tag{8.6}$$

The bandwidth of the amplifier is determined in large part by the large external detector capacitance C_D, which does not scale as more slices are connected in parallel. On the Heisenberg chip, the value of the load capacitance C_L is determined by the parasitic capacitances of the output transistors of the CSA and the input transistors of the cascading buffer. Thus, C_L can be written as:

$$C_L = C_{Lcsa} + C_{Lbuf} \tag{8.7}$$

Let us consider the simplified form of (8.6): if k identical amplifier slices are connected in parallel, the effective transconductance G_{meff} scales proportionally to the number of slices ($k \cdot G_{meff}$); the parasitic contributions to the load capacitance C_L which are affected by the CSA also scale proportionally ($C_L = k \cdot C_{Lcsa} + C_{Lbuf}$); while C_F remains unchanged; and C_{D+gg} remains mostly unchanged. The proportion in which the position of the pole moves is not obvious, as it is dependent on the value of C_L, but it is limited to a maximum of a k-fold increase in bandwidth given by the scaling of G_{meff}.

In practice, as it can be seen in Figure 8.7, the output of the Heisenberg CSA appears to behave as a second-order circuit, as there is a small amount of overshoot for some of the curves, both in simulation and measurements, particularly for larger values of the feedback capacitance. The presence of overshoot indicates that the circuit is behaving as an underdamped second-order system, meaning that the two poles of the circuit are complex conjugates, and as such, are not separable. It also appears that the response of the amplifier goes from being overdamped, or at least underdamped with a damping factor (ζ) very close to unity, to being notably underdamped as more slices are connected in parallel.

This can be explained by the presence of a secondary non-dominant pole in the circuit. From the transistor design parameters for the NMOS-input CSA presented in Section

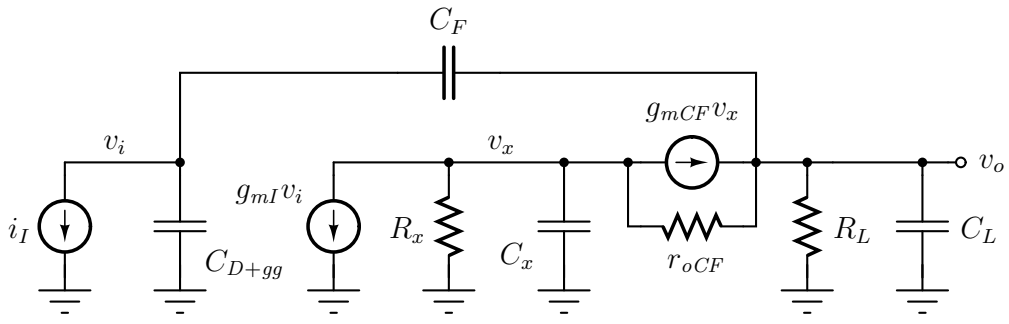


FIGURE 8.8. Schematic of the NMOS-input folded-cascode amplifier for small-signal analysis, with capacitive feedback. An explicit capacitor to account for the non-dominant pole was added to the node v_x to analyze the second-order response of the circuit.

6.2.6, it can be observed that the folding transistor M_F has a very large width W. From this, it is reasonable to assume that the comparatively large shunt parasitic capacitance on the drain of this particular device, plus the parasitic contributions of the other transistors on this node (M_I and M_{CF}), introduces the secondary non-dominant pole. This was later confirmed from the simulation of a single slice, by adding an explicit capacitor on the drain of M_F, which produces overshoot on the output depending on the value, which does not happen on any other node.

To understand the condition that the circuit must meet to go from being overdamped to underdamped as more slices are connected in parallel, let us consider the circuit shown in Figure 8.8, where

$$R_x = r_{oI} \| r_{oF} \tag{8.8}$$

$$C_x = C_{gdI} + C_{dbI} + C_{gdF} + C_{dbF} + C_{gsCF} + C_{sbCF} \tag{8.9}$$

$$R_L = g_{mCL} \cdot r_{oCL} \cdot r_{oL} + r_{oCL} + r_{oL} \approx g_{mCL} \cdot r_{oCL} \cdot r_{oL} \tag{8.10}$$

This circuit is a small-signal schematic of the NMOS-input folded cascode amplifier used in the Heisenberg chip, but with the addition of an explicit capacitance C_x on the drain of the folding transistor M_F. The term C_x in (8.9) includes the parasitic capacitances from gate-to-drain (C_{gd}) and drain-to-body (C_{db}) for transistors M_I and M_F, and the parasitic capacitances from gate-to-source (C_{gs}) and source-to-body (C_{sb}) for the transistor M_{CF}.

From Figure 8.8 it is possible to compute a simplified expression for the damping factor (see Annex A) by considering large intrinsic gain values for transistors M_I, M_F and M_{CF}, to be

$$\zeta^2 = \frac{1}{4} \cdot \frac{g_{mCF}}{C_x} \cdot \frac{C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg}}{g_{mI} \cdot C_F} \tag{8.11}$$

It can be observed that the expression for the dominant-pole approximation of the amplifier (8.6) appears in the damping factor (since $G_{meff} \approx g_{mI}$), which can be rewritten as a function of p

$$\zeta^2 = \frac{\frac{1}{4} \cdot \frac{g_{mCF}}{C_x}}{|p|} \tag{8.12}$$

From (8.12) it can be observed that the numerator does not change as additional slices are connected in parallel, since the capacitance and the transconductance scale in the same proportion. In contrast, as it was described in an earlier paragraph, the denominator ($|p|$) gets increasingly larger as more slices are connected in parallel.

A circuit is said to be overdamped when ($\zeta > 1$), and underdamped when ($0 < \zeta < 1$). Thus, the CSA will be underdamped if the following condition holds true

$$4 \cdot \frac{g_{mI}}{g_{mCF}} \cdot C_x > \frac{C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg}}{C_F} \tag{8.13}$$

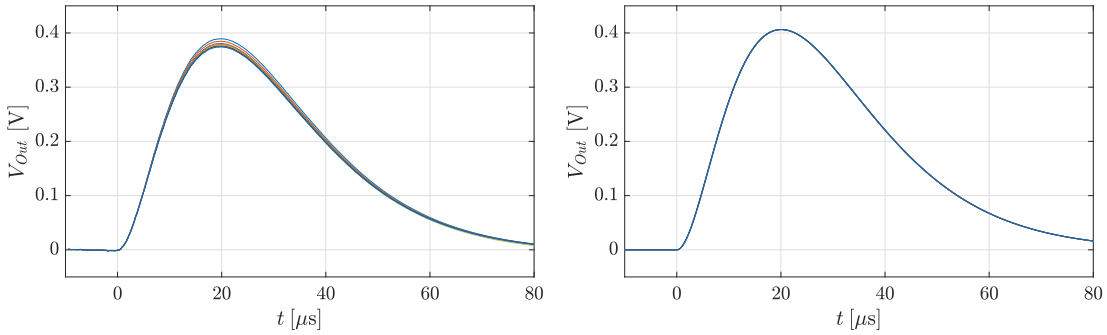
The transconductance ratio in the left term of (8.13) for this particular CSA can be calculated to be roughly 13.5. Assuming that $C_{D+gg} \gg C_L$ and $C_{D+gg} \gg C_F$, (8.13) can be rewritten as

$$C_x > \frac{C_D}{54} \cdot \left(1 + \frac{C_L}{C_F}\right) \tag{8.14}$$

Using g_m/I_D curves it was calculated that C_x is roughly 950 fF for a single slice. Thus, for 8 parallel-connected slices, $C_{x,Tot} = 7.6\,\text{pF}$, which is very close to the critically damped condition ($C_D/54 = (390\,\text{pF})/54 = 7.2\,\text{pF}$).

8.7 Filter output

Figure 8.9 shows the output of the $CR - 2RC$ filter on the Heisenberg test board for a feedback capacitance of 4 pF. The filter is connected in cascade to the output of the



(A) Oscilloscope. (B) Simulation.

FIGURE 8.9. $CR - 2RC$ filter output for different number of parallel-connected slices. Both measured waveforms and simulation waveforms are presented for $C_F = 4$ pF.

NMOS-input charge-sensitive amplifier, and thus these waveforms are the filtered version of the ones presented in Figures 8.4(A) and 8.4(B).

The active filter was designed to have a unity gain from input to output, and thus there is some gain on each active stage to compensate for filter attenuation (see Section 6.4). The amplitude of the measured waveforms is slightly lower than that of the CSA output, which shows that the active filter on the Heisenbeg test board has a gain slightly smaller than unity, possibly due to component tolerance.

8.8 Noise measurements

As described in Chapter 3, the amplifier is one of two circuit blocks that introduce electronic noise to the output of a particle physics front-end, the other one being the detector itself. As such, minimizing the amplifier noise is of paramount importance for accurate energy measurements of physics events. One of the main objectives of applying the slice-based design methodology to the design of a CSA was to observe the noise behavior as an increasing number of amplifier slices are connected in parallel. The results of these measurements are presented in this section.

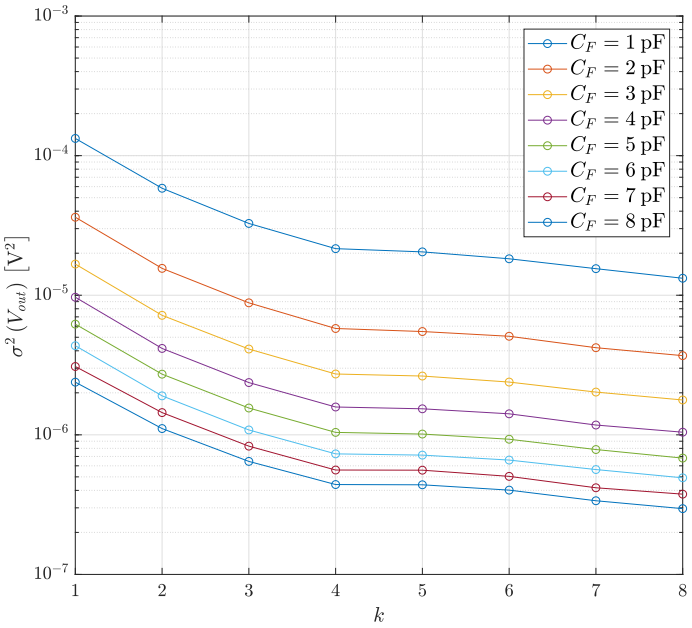


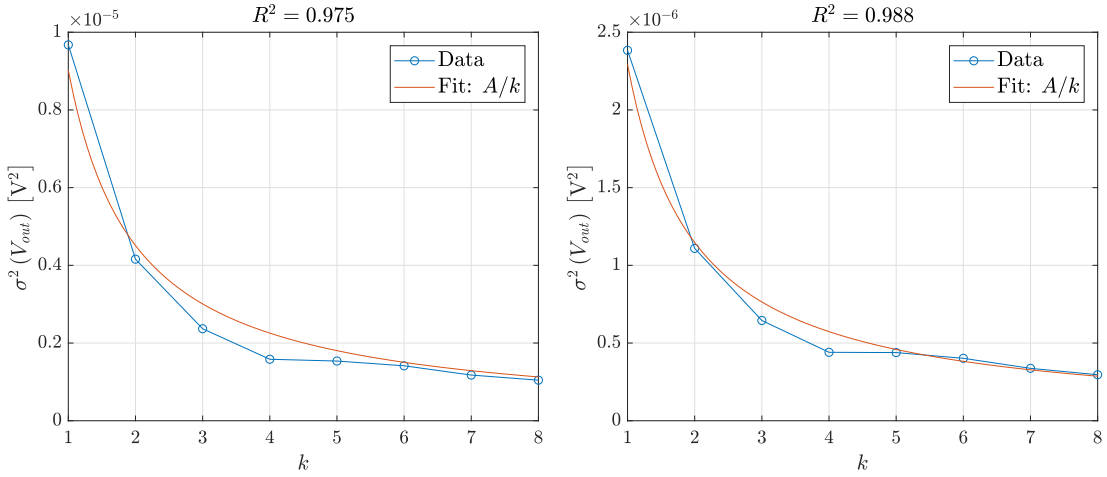
FIGURE 8.10. Noise measurements for all the combinations of k and C_F used in the testing of the Heisenberg chip. The plot is presented with a logarithmic y axis to be able to display all curves properly.

8.8.1 Setup

The filter output shown in Figure 8.9(A) was sampled at the peaking time by the ADC of the Heisenberg test board to get enough statistical data to characterize the noise performance of the charge-sensitive amplifier. An example of one of the noisy signals used for this purpose is shown in Figure 8.1.

For each discrete value of k ranging from 1 to 8, and for each value of C_F ranging from $1\,\mathrm{pF}$ to $8\,\mathrm{pF}$ (with increments of $1\,\mathrm{pF}$), a total of $75{,}000$ voltage samples were taken by the ADC and stored on a PC for subsequent analysis.

For the sake of simplicity, the sampled voltage will be referred to as V_{out} in the present section. The noise of the circuit can be computed by calculating the variance of the collected voltage samples $(\sigma^2(V_{out}))$, which will yield an accurate measurement with enough statistical data.



(A) $C_F = 4$ pF (B) $C_F = 8$ pF

FIGURE 8.11. Noise measurements for an increasing number of parallel-connected slices, and two values of the feedback capacitance C_F. A nonlinear least-squares fit is also included using the model for expected behavior ($\propto 1/k$).

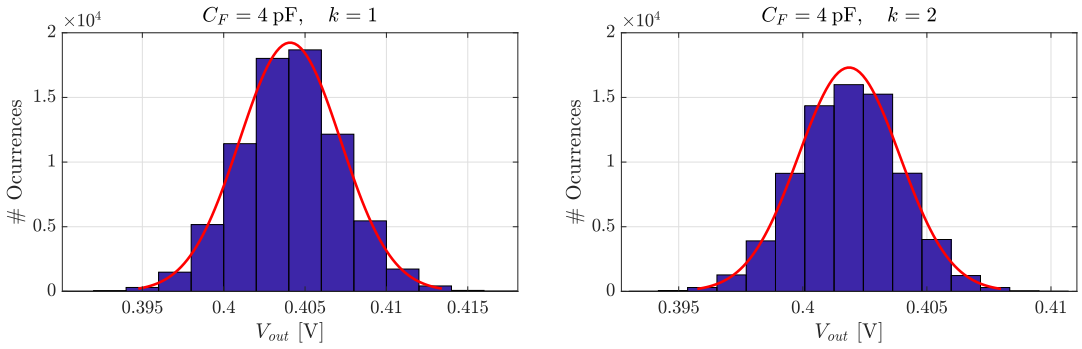
8.8.2 Results

Figure 8.10 shows a semi-logarithmic plot for the results of all the noise measurements done on the Heisenberg chip used for testing. Each marked point on each one of the curves corresponds to the variance of $75{,}000$ voltage samples.

Figure 8.11 shows a closer view with linear axes at the noise curves for feedback capacitance values of 4 pF and 8 pF. These particular plots also include a fitted curve ($\propto 1/k$) for the expected behavior of the noise.

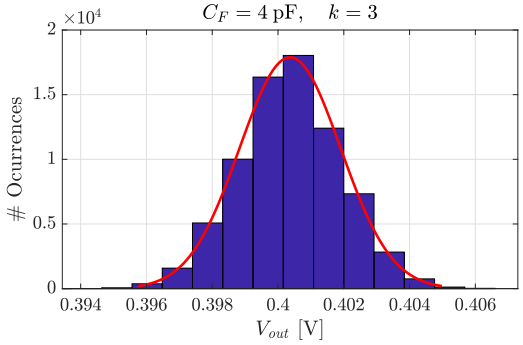
Figure 8.12 shows the histograms of the measured voltage for the different numbers of parallel-connected slices, considering a feedback capacitance of 4 pF. The variance of each of these histograms represents the noise for that particular combination of k and C_F.

A quick inspection of the plots reveals that noise is indeed reduced as more slices are connected in parallel, as predicted. A closer view of the magnitude of the obtained results, and of the scaling behavior of the noise curves is presented in the following sections.

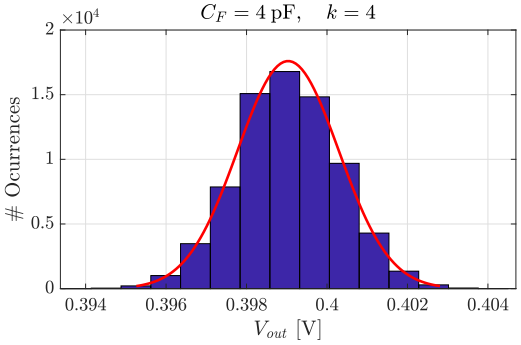


(A) $k = 1$

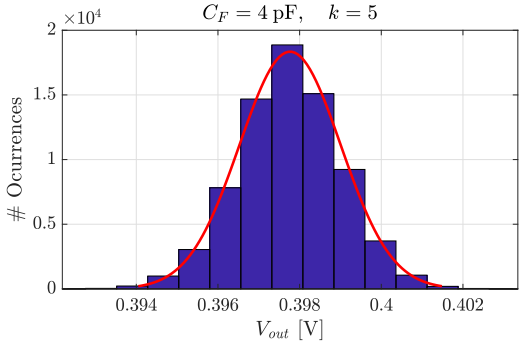
(B) $k = 2$



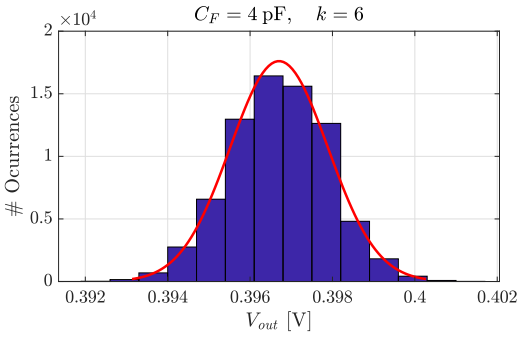
(C) $k = 3$



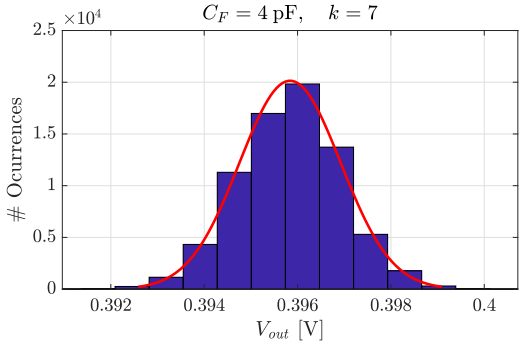
(D) $k = 4$



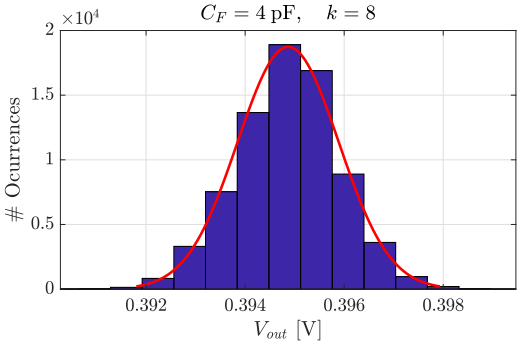
(E) $k = 5$



(F) $k = 6$



(G) $k = 7$



(H) $k = 8$

FIGURE 8.12. Histograms for the filter output voltage, being sampled at the peaking time, for different numbers of parallel-connected slices (k) using $C_F = 4$ pF. The variance of each histogram represents the noise for that particular configuration of variables.

8.8.3 Noise magnitude and thermal noise

The mathematical framework for noise analysis of particle physics front-ends in general, and the charge-sensitive amplifier in particular, was thoroughly analyzed in Chapter 3. From the equivalent noise charge (ENC) expression (3.23), it is possible to write a formula for the CSA output noise voltage in terms of the normalized noise coefficients, as follows

$$\overline{V_{O,noise}^2} = \frac{q^2}{C_F^2}\left[\tau_P N_{Pn}\cdot\overline{I_D^2} + (C_{D+gg}+C_F)^2\left(\frac{N_{Wn}}{\tau_P}\cdot\overline{V_{A,W}^2} + N_{Fn}\cdot K_F\right)\right] \quad (8.15)$$

where the flicker noise exponent was assumed to be unity ($A_F = 1$), and N_{Pn}, N_{Wn} and N_{Fn} are the normalized noise coefficients for parallel, thermal and flicker noise, respectively.

Since the Heisenberg chip was tested without a detector, and instead the detector capacitance was emulated with an explicit capacitor (C_D) on the Heisenberg test board, it could be assumed that there is no shot noise ($\overline{I_D^2}$) in the system. However, there should be a small amount of shot noise due to the pad leakage current, with a power spectral density (PSD) of $\overline{I_D^2}(f)/\Delta f = 2q|I_P|$ [A^2/Hz], where I_P is the DC leakage current. Regardless, after analyzing the numerical results, it was concluded that the shot noise introduced by the pads should be several orders of magnitude smaller than the amplifier noise, and thus it will be ignored for the rest of the analysis.

In Section 6.2.5, it was mentioned that the input-referred CSA noise is mostly affected by two devices: the input transistor M_I and the folding transistor M_F. From the analysis presented in that section, it follows that the input-referred amplifier noise can be written as

$$\overline{V_A^2} = \overline{V_{NI}^2} + \left(\frac{g_{mF}}{g_{mI}}\right)^2\overline{V_{NF}^2}, \quad (8.16)$$

where $\overline{V_{NI}^2}$ and $\overline{V_{NF}^2}$ are the gate-referred noise contributions of transistors M_I and M_F, respectively. From this formulation, it is straightforward to see that the thermal and flicker

noise components in (8.15) can be written in the following form

$$\overline{V_{A,W}^2} = \overline{V_{NI,W}^2} + \left(\frac{g_{mF}}{g_{mI}}\right)^2 \overline{V_{NF,W}^2} \tag{8.17}$$

$$K_F = K_{FI} + \left(\frac{g_{mF}}{g_{mI}}\right)^2 K_{FF} \tag{8.18}$$

Let us consider (8.17), which can be written in terms of the normalized thermal noise PSD (see Section 3.4.2) as follows:

$$\overline{V_{A,W}^2} = \frac{\overline{\hat{V}_{NI,W}^2}}{I_{DI}} + \left(\frac{g_{mF}}{g_{mI}}\right)^2 \cdot \frac{\overline{\hat{V}_{NF,W}^2}}{I_{DF}} \tag{8.19}$$

where I_{DI} and I_{DF} are the drain currents for transistors M_I and M_F, respectively. If we multiply on both sides by I_{DI}, it is possible to write a normalized expression for the input-referred amplifier white noise $\overline{\hat{V}_{A,W}^2} = \overline{V_{A,W}^2} \cdot I_{DI}$, as follows

$$\begin{aligned}\overline{\hat{V}_{A,W}^2} &= \overline{\hat{V}_{NI,W}^2} + \left(\frac{I_{DI}}{I_{DF}}\right)\left(\frac{g_{mF}}{g_{mI}}\right)^2 \cdot \overline{\hat{V}_{NF,W}^2} \\ &= \overline{\hat{V}_{NI,W}^2} + \left(\frac{I_{DF}}{I_{DI}}\right)\left[\frac{(g_m/I_D)_{M_F}}{(g_m/I_D)_{M_I}}\right]^2 \cdot \overline{\hat{V}_{NF,W}^2}\end{aligned} \tag{8.20}$$

The same is also valid for the normalized flicker coefficient $\hat{K}_F = K_F \cdot I_{DI}$

$$\hat{K}_F = \hat{K}_{FI} + \left(\frac{I_{DF}}{I_{DI}}\right)\left[\frac{(g_m/I_D)_{M_F}}{(g_m/I_D)_{M_I}}\right]^2 \cdot \hat{K}_{FF} \tag{8.21}$$

The output noise voltage (8.15) can be written in terms of the normalized thermal noise PSD $\overline{\hat{V}_{A,W}^2}$ and the normalized flicker coefficient $\hat{K}_F$, as follows

$$\overline{V_{O,noise}^2} = \frac{q^2}{k \cdot I_{DI}}\left(\frac{C_{D+gg} + C_F}{C_F}\right)^2 \left[\frac{N_{Wn}}{\tau_P} \cdot \overline{\hat{V}_{A,W}^2} + N_{Fn} \cdot \hat{K}_F\right] \tag{8.22}$$

where I_{DI} is the drain current of the input transistor for a single amplifier slice. The normalized noise terms $\overline{\hat{V}_{A,W}^2}$ and $\hat{K}_F$ are solely functions of g_m/I_D and the drain current ratio I_{DF}/I_{DI}.[5]

[5]The ratio between the folding transistor current and the input transistor current is constant for a given slice, and does not change as more slices are connected in parallel.

TABLE 8.1. Parameters for noise calculations.

Parameter	Value
$q^2 N_{Wn}$	0.82
$q^2 N_{Fn}$	0.54
C_D	390 [pF]
C_{gg}	160 [fF]
C_F	8 [pF]
I_{DI}	250 [μA]
I_{DF}	275 [μA]
$(g_m/I_D)_{M_I}$	16.4 [mS/mA]
$(g_m/I_D)_{M_F}$	11.5 [mS/mA]
$\overline{\hat{V}^2_{A,W}}$	1.586×10^{-21} [VJ]
τ_P	20 [μs]

Let us consider the white noise component of (8.22) to try to compute a value for the noise of a single slice, considering $C_F = 8$ pF. The values of the normalized noise coefficients are shown in Chapter 3, in Table 3.1, where the values for a RU-3 filter (*i.e.* the $CR - 2RC$ filter) can be retrieved. Table 8.1 summarizes these and other important circuit parameters necessary to calculate the noise. The resulting value for the thermal noise contribution can be computed to be

$$\begin{aligned} \overline{V^2_{O,noise}} &= \frac{1}{I_D} \cdot \left(\frac{C_{D+gg} + C_F}{C_F} \right)^2 \cdot \frac{q^2 N_{Wn}}{\tau_P} \cdot \overline{\hat{V}^2_{A,W}} \\ &= 6.44 \times 10^{-10}\ \mathrm{V}^2 \end{aligned} \tag{8.23}$$

The RMS value of the calculated noise is $V_{O,RMS} = 22.4\ \mu$V. If the CSA output noise were dominated by thermal noise, then (8.23) should be close to the first point in Figure 8.11(B). However, from the plot it can be seen that the measured noise for $k = 1$ is $\sigma^2(V_{out})_{k=1} = 2.4 \times 10^{-6}\ \mathrm{V}^2$, or alternatively $\sigma(V_{out})_{k=1} = 1.5$ mV, which is orders of magnitude higher than the calculated value.

8.8.4 Flicker noise

Since thermal noise does not seem to be the dominant noise process on the output of the CSA, it follows that flicker ($1/f$) noise could be the culprit. From Figure 8.10 it can

be observed that the measured noise clearly scales with the number of parallel-connected slices, and from (8.22) it can be seen that both white and flicker noise scale with k.

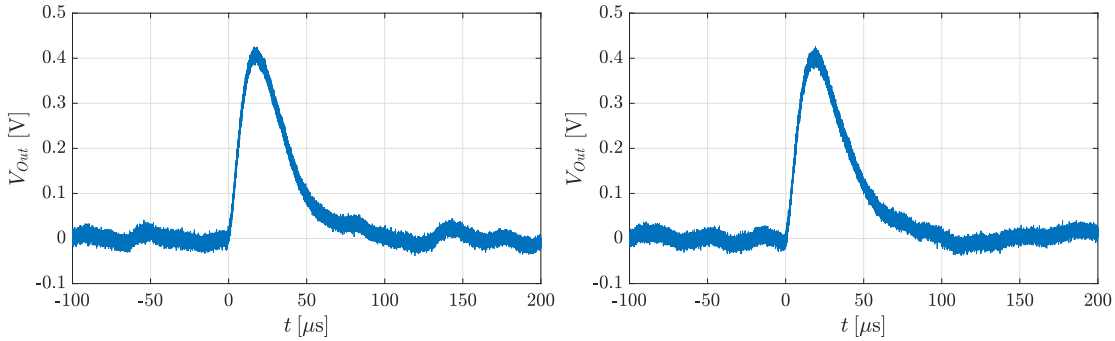
It is tempting to attempt to calculate the magnitude of the flicker noise from (8.22) to confirm that flicker noise is the dominant noise process. Unfortunately, the available SPICE models for the selected 0.5-$\mu\mathrm{m}$ CMOS technology do not include a value for the flicker coefficient, and thus it is not possible to calculate the flicker noise from simulation data, or to compute accurate g_m/I_D curves for $\hat{K}_F$.

Additionally, the circuit model used to derive (8.22) assumed a continuous reset element (see Figure 3.2), while in practice, a switch was used. Without delving into the topic of weighting functions (Radeka (1968), Goulding (1972)), which is an extension of the analysis done in Chapter 3 but valid also for time-varying systems, it can be said that the flicker noise component in (8.22) provides an upper bound for the measurable noise.

Nonetheless, several factors point to the possibility of flicker noise being the dominant noise process. First, an NMOS transistor was used as the input device, which typically has a much larger flicker noise coefficient (by several orders of magnitude) than their PMOS counterparts. Second, a relatively large time window of $280\ \mu\mathrm{s}$ was used between the instant that the reset was released and the time when the signal was sampled, giving the $1/f$ noise enough time to rise.

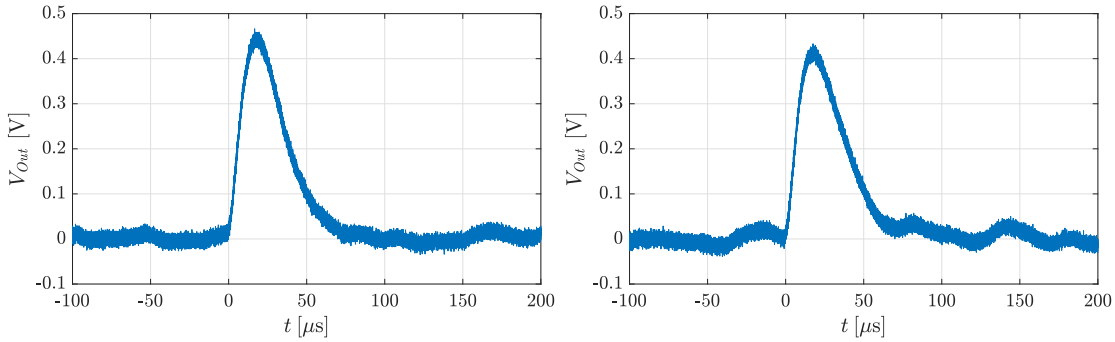
Finally, it is possible to observe from singular noisy waveforms that the noise does not behave as purely white noise (with a constant envelope). Figures 8.13(A)-(D) show individual waveforms for $C_F = 1\ \mathrm{pF}$ and $k = 1$, the noisiest combination of variables measured, and it is possible to observe from these plots random low frequency oscillations, which clearly have an effect at the moment of sampling. These curves can be contrasted with the ones presented in Figures 8.13(E)-(F), for $C_F = 1\ \mathrm{pF}$ and $k = 8$, where these random low frequency oscillations are still present, but are less noticeable.

From the above observations, and given that the measured noise scales with the number of parallel-connected slices, it is likely that flicker noise is the dominant noise process in the measured waveforms.



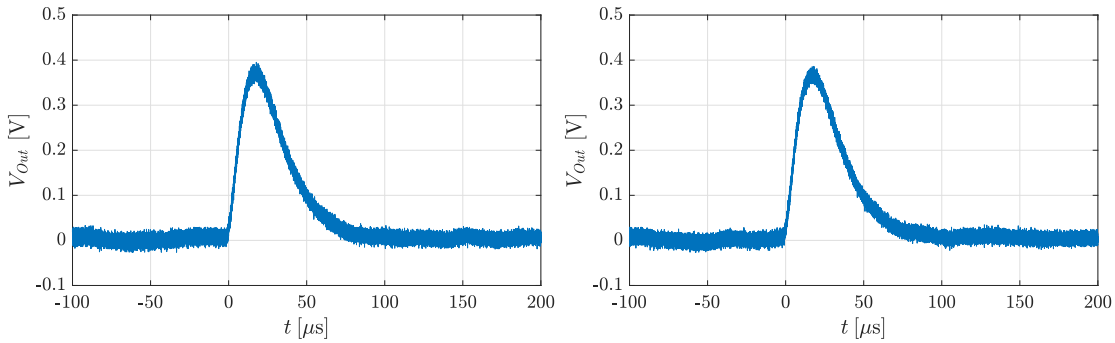
(A) $k = 1$.

(B) $k = 1$.



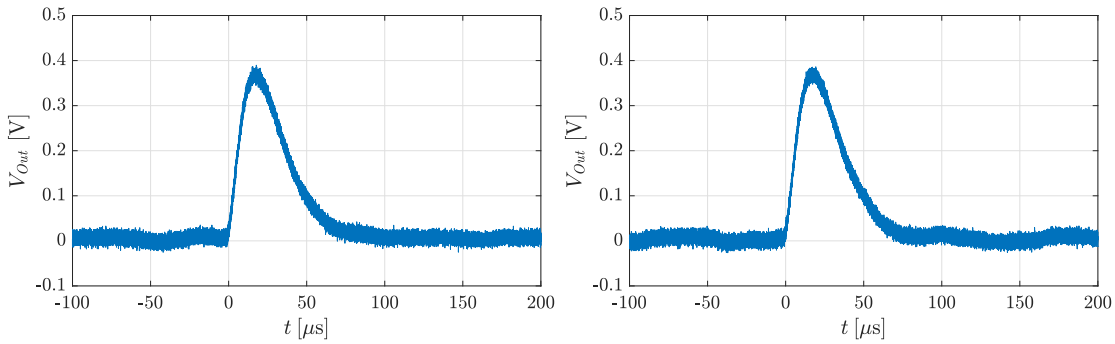
(C) $k = 1$.

(D) $k = 1$.



(E) $k = 8$.

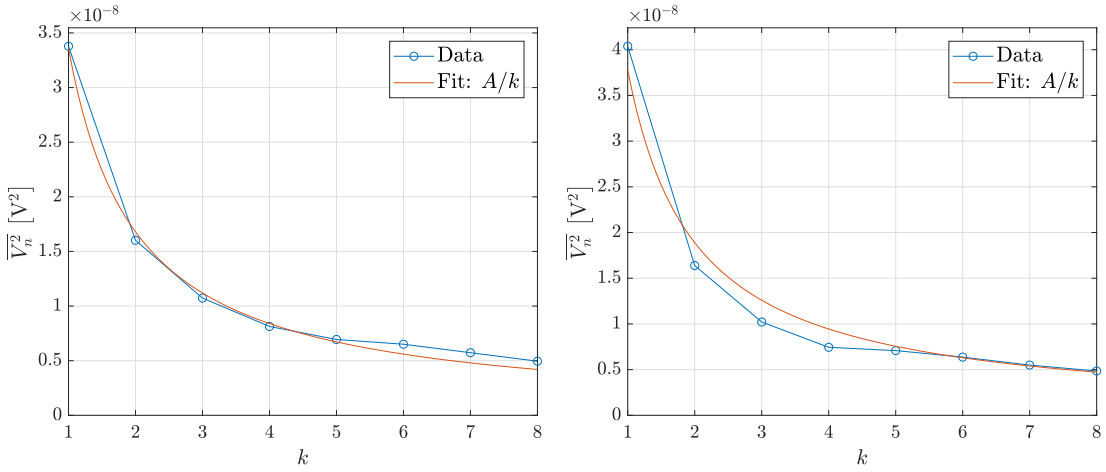
(F) $k = 8$.



(G) $k = 8$.

(H) $k = 8$.

FIGURE 8.13. Individual waveforms for the filter output for $C_F = 1\,\mathrm{pF}$, which is the smallest value available for the feedback capacitance on the Heisenberg chip. Two different values of k ($k = 1$ and $k = 8$) are presented.



(A) Iteration # 1064 (B) Iteration # 1826

FIGURE 8.14. Handpicked iterations of the Monte Carlo simulation of the CSA ($C_F = 8$ pF).

8.8.5 Noise scaling

From Figure 8.10 it can be observed that the scaling of the noise as a function of k appears to be insensitive to the value of the feedback capacitance C_F, and all the curves seem to be offset by a constant value. This is clear from (8.22), since the capacitive term appears as an independent factor with a very weak dependency with k ($C_{D+gg}(k) \approx C_D$), and thus as an additive constant in the semi-logarithmic plot.

Given that all curves scale as functions of k in almost identical fashion, it is not relevant which curve is used to analyze the scaling of the noise. Let us consider the two curves presented in Figure 8.11, which also include a fitted curve for the expected behavior. It can be observed that the tendency of the noise scaling follows closely with the expected behavior ($\propto 1/k$), but it is not a perfect fit.

The deviation of the curves with respect to the expected behavior can be explained by gradient-related and size-related mismatch. The interaction of device mismatch with the proposed slice-based design methodology was explored in detail in Chapter 4. In particular, a Monte Carlo simulation of the CSA (see Section 4.7.3) with plausible values for the mismatch variances was performed to confirm whether it was possible to replicate similar results to the ones obtained with the Heisenberg chip.

Each one of the iterations of the Monte Carlo simulation of Section 4.7.3 corresponds to a single realization of the Heisenberg chip under the influence of device mismatch. The integrated noise curves as a function of the number of parallel-connected slices were plotted for all 2000 realizations and analyzed through visual inspection. In most cases, the mismatch has very little impact on the behavior of the noise and the scaling. However, in some outlier cases, there were more obvious deviations with respect to the expected noise scaling. A couple of examples are presented in Figure 8.14, which were selected specifically because of their similar shape to the measured curves.

The above explanation is not conclusive, but only a possible explanation, and the most likely one given the available information. To further test this hypothesis, a larger number of chips manufactured in different wafers (so that the process gradients are randomized and uncorrelated) would be necessary. Unfortunately, only a very small number of chips were available for testing, all of which most likely shared a wafer, and were manufactured in close physical proximity on the wafer.

8.9 Summary of design flaws

Three design flaws became apparent during the testing of the Heisenberg chip and test board, and it would be desirable to correct these if another version of the Heisenberg chip was ever manufactured. Fortunately, none of these flaws compromised the results in a meaningful way, and in some cases even provided additional insight into the operation of the circuit.

The first flaw was the selection of a unipolar ADC, which cannot measure negative signals. An ADC with bipolar references or used in a differential configuration could have overcome this limitation. However, the necessity of inducing a signal to measure noise did not affect the results, but instead introduced additional practical considerations that could have been avoided.

The second flaw of the design was the use of a very large folding transistor M_F which introduced a second, non-dominant pole to the analysis. In practice, it is desirable to keep

the analysis of the amplifier simple and predictable, and it is complicated by the addition of the second pole. Nonetheless, the use of a relatively slow pulse-shaping filter meant that the bandwidth of the amplifier became effectively irrelevant to the measured results.

The third flaw was the of combination factors that lead to flicker noise being the dominant noise process. In most fast front-end circuits, it is typically white noise the dominant noise process. Plus, flicker noise is less intuitive to analyze and requires additional process parameters provided by the manufacturer to characterize accurately. This could have been avoided by using a smaller peaking time τ_P and a shorter time window from when the reset was released to when the signal was sampled. There is a limit to how much both of these can be reduced in the current design, given by the bandwidth of the CSA. The apparent dominance of the flicker noise did not affect the main hypothesis of this thesis, that is, that amplifier noise can be easily and effectively reduced by connecting additional amplifier slices in parallel.

The large cell pitch for each amplifier slice could also be considered as a design flaw, since it makes the circuit more susceptible to gradient-related mismatch. However, it lead to a more detailed analysis of the effects of mismatch on the proposed design methodology, which is critical for the further development of the slice-based design technique.

9. CONCLUSION

9.1 Summary

This thesis serves as a practical exploration of a particular idea, the slice-based design methodology, inspired by the g_m/I_D design technique. The slice-based design methodology is an innovative approach to analog design through the use of pre-designed circuit cells, which can be connected in parallel to scale important performance metrics. Since it is not possible to analyze the applicability and practicality of the proposed design methodology to any arbitrary circuit topology, it was decided to limit the scope of this exploration to a particular target application: particle physics instrumentation. Within this context, it was decided to focus on the design of low-noise charge-sensitive amplifiers (CSA), with a particular emphasis on the scaling of the noise performance.

In order to evaluate practical design considerations and to measure the real-world performance of a CSA designed with the proposed design technique, a custom application-specific integrated circuit (ASIC) was designed, fabricated and tested. The ASIC prominently includes a configurable CSA, comprised of several amplifier slices which can be connected in parallel, from a single slice to a total of eight. The integrated noise of the CSA can be measured on the output of the pulse-shaping filter to characterize the scaling of the noise performance.

One of the questions addressed by this thesis is whether it is possible to design a CSA with an optimum g_m/I_D that minimizes noise independently of the problem specifications, for a given technology. It was concluded that this is not possible, as the optimum g_m/I_D is not detachable from the system the peaking time. There are, however, some practical considerations that can be made for a white-noise dominated system to design a single slice that is widely applicable in terms of performance scalability, die area and noise performance, but it is highly technology-dependent.

The second, more general question addressed by this thesis is what are the practical considerations when this approach is implemented on a custom ASIC. Does it work,

how well does it work, and what are the caveats. Through the practical exploration of this problem, it was concluded that it indeed works in practice: noise and other important performance metrics can be scaled. It works mostly as intended, but there might be some considerations that are highly problem-specific. Additionally, gradient-related device mismatch can become a relevant issue depending on cell pitch and the number of parallel-connected slices.

Among the problem-specific considerations that might arise are the dependency of circuit performance metrics to boundary conditions, *i.e.* what is connected on the input and output. Let us consider the CSA as an example. What is connected to the input is critical to circuit operation. The amount of charge deposited by the detector influences the size of the feedback capacitance, and the size of the parasitic capacitance of the detector influences the speed of the CSA and the noise of the system as a whole. What is connected to the output of also important, as the pulse-shaping filter limits the bandwidth of the amplifier. If the bandwidth of the CSA was not limited externally, as additional slices are connected in parallel, the bandwidth of the amplifier would increase while the power-spectral densities (PSD) of the white and flicker noise would decrease, resulting in no obvious gain in terms of integrated noise performance. Thus, a thorough understanding of the target application is necessary to use the proposed design methodology effectively.

9.2 Future work

The are several unexplored topics related to the slice-based design methodology. Two of them were alluded to in Chapter 2, related to non-ideal effects of connecting parallel circuits through long wires. First, there are some parasitic components implied in the stackable layout due to the metal traces that connect the parallel cells. These might have an effect on node impedances, and might become more relevant for very fast applications. Second, mismatch might cause voltage differences between nominally identical nodes, which would translate to current flow through the wires. A better understanding of these effects and their importance to different target applications is desirable.

In Chapter 4 it was observed that the proposed parameter mismatch model does not match perfectly with the results of the Monte Carlo simulation of the CSA. To further understand the interaction of device mismatch with the slice-based design methodology, it is desirable to derive a generalized model for parameter mismatch on an arbitrarily complex multi-transistor circuit.

The optimal layout geometry is also unclear. The cell and interconnection geometry proposed in 2 and implemented in Chapter 7 was conceived with a focus in ease of use. There might be some physical way to distribute and interconnect the cells at the layout level that reduces mismatch effects and allows for optimal area utilization.

Finally, and more generally, other target applications need to be analyzed to assess the practicality and applicability of using the proposed design technique.

REFERENCES

Alvarez, E., & Abusleme, A. (2012, April). Noise power normalisation: extension of g_m/i_D technique for noise analysis. *Electronics Letters*, *48*(8), 430-432. doi: 10.1049/el.2011.3730

Alvarez, E., Avila, D., Campillo, H., Dragone, A., & Abusleme, A. (2012, Oct). Noise in charge amplifiers - a g_m/i_D approach. *IEEE Transactions on Nuclear Science*, *59*(5), 2457-2462. doi: 10.1109/TNS.2012.2208270

Analog-Devices. (1999). *Ltspice by analog devices.* `https://analog.com/`. (Accessed: 2018-09-18)

Analog-Devices. (2016). *1a, low dropout, cmos linear regulator* (No. ADP1706ARDZ-3.3-R7). (Datasheet - Rev. A)

Aspell, P., Barney, D., Bloch, P., Jarron, P., Lofstedt, B., Reynaud, S., & Tabbers, P. (2001). Delta: a charge sensitive front-end amplifier with switched gain for low-noise, large dynamic range silicon detector readout. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *461*(1), 449 - 455. (8th Pisa Meeting on Advanced Detectors) doi: https://doi.org/10.1016/S0168-9002(00)01270-5

ATLAS Collaboration. (2008). The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, *3*, S08003. doi: 10.1088/1748-0221/3/08/S08003

ATLAS Collaboration. (2012). Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, *716*(1), 1 - 29. doi: https://doi.org/10.1016/j.physletb.2012.08.020

Baker, R. J. (2010). *Cmos circuit design, layout, and simulation* (3rd ed.). Wiley-IEEE Press.

Barros, M., Guilherme, J., & Horta, N. (2010). *Analog circuits and systems optimization based on evolutionary computation techniques* (1st ed.). Springer.

Binkley, D. M. (2007, June). Tradeoffs and optimization in analog cmos design. In *2007 14th international conference on mixed design of integrated circuits and systems* (p. 47-60). doi: 10.1109/MIXDES.2007.4286119

Cadence. (2006). *Virtuoso layout suite by cadence design systems inc.* `https://www.cadence.com/`. (Accessed: 2018-09-18)

CMS Collaboration. (2012). Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, *716*(1), 30 - 61. doi: https://doi.org/10.1016/j.physletb.2012.08.021

Croon, J. A., Rosmeulen, M., Decoutere, S., Sansen, W., & Maes, H. E. (2002, Aug). An easy-to-use mismatch model for the mos transistor. *IEEE Journal of Solid-State Circuits*, *37*(8), 1056-1064. doi: 10.1109/JSSC.2002.800953

De Geronimo, G., & O'Connor, P. (2005, Dec). Mosfet optimization in deep submicron technology for charge amplifiers. *IEEE Transactions on Nuclear Science*, *52*(6), 3223-3232. doi: 10.1109/TNS.2005.862938

Drennan, P. G., & McAndrew, C. C. (2003, Mar). Understanding mosfet mismatch for analog design. *IEEE Journal of Solid-State Circuits*, *38*(3), 450-456. doi: 10.1109/JSSC.2002.808305

Evans, L., & Bryant, P. (2008, aug). LHC machine. *Journal of Instrumentation*, *3*(08), S08001–S08001. doi: 10.1088/1748-0221/3/08/s08001

Flandre, D., Viviani, A., Eggermont, J. P., Gentinne, B., & Jespers, P. G. A. (1997, Jul). Improved synthesis of gain-boosted regulated-cascode cmos stages using symbolic analysis and gm/id methodology. *IEEE Journal of Solid-State Circuits*, *32*(7), 1006-1012. doi: 10.1109/4.597291

Foty, D., Bucher, M., & Binkley, D. (2002, Sep.). Re-interpreting the mos transistor via the inversion coefficient and the continuum of g/sub ms//i/sub d/. In *9th international conference on electronics, circuits and systems* (Vol. 3, p. 1179-1182 vol.3). doi: 10.1109/ICECS.2002.1046463

FTDI. (2016). *Dual high speed usb to multipurpose uart/fifo ic* (No. FT2232H). (Datasheet - Rev 2.5)

Gielen, G. G. E. (2007, Jan). Design tool solutions for mixed-signal/rf circuit design in cmos nanometer technologies. In *2007 asia and south pacific design automation conference* (p. 432-437). doi: 10.1109/ASPDAC.2007.358024

Gielen, G. G. E., & Rutenbar, R. A. (2000, Dec). Computer-aided design of analog and mixed-signal integrated circuits. *Proceedings of the IEEE*, *88*(12), 1825-1854. doi: 10.1109/5.899053

Goulding, F. (1972). Pulse-shaping in low-noise nuclear amplifiers: A physical approach to noise analysis. *Nuclear Instruments and Methods*, *100*(3), 493-504. Retrieved from `https://www.sciencedirect.com/science/article/pii/0029554X72908282` doi: https://doi.org/10.1016/0029-554X(72)90828-2

Gray, P., Hurst, P., Lewis, S., & Meyer, R. (2001). *Analysis and design of analog integrated circuits*. Wiley.

Griffiths, D. J. (2008). *Introduction to elementary particles; 2nd rev. version*. New York, NY: Wiley.

Grupen, C., & Shwartz, B. (2008). *Particle detectors*. Cambridge University Press.

Harris, I., Sutherland, I., Sproull, R., Sproull, B., Harris, D., & Harris, P. (1999). *Logical effort: Designing fast cmos circuits*. Elsevier Science.

Kinget, P. R. (2005). Device mismatch and tradeoffs in the design of analog circuits. *IEEE Journal of Solid-State Circuits*, *40*(6), 1212-1224. doi: 10.1109/JSSC.2005.848021

Linares-Barranco, B., & Serrano-Gotarredona, T. (2007, Aug). On an efficient cad implementation of the distance term in pelgrom's mismatch model. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *26*(8), 1534-1538. doi: 10.1109/TCAD.2007.893546

Linear-Technology. (2003). *215mhz, rail-to-rail output, 1.1nv/sqrthz, 3.5ma op amp family* (No. LT6232). (Datasheet - Rev. C)

Liu, W., Jin, X., Chen, J., Jeng, M.-C., Liu, Z., Cheng, Y., ... Hu, C. (1998). *Bsim 3v3.2 mosfet model users' manual* (Tech. Rep. No. UCB/ERL M98/51). EECS Department, University of California, Berkeley.

Magic VLSI. (2018). *Magic vlsi layout design software.* `http://opencircuitdesign.com/magic/magic_docs.html`.

Martins, R., Lourenço, N., & Horta, N. (2012). *Generating analog ic layouts with laygen ii*. Springer. doi: 10.1007/978-3-642-33146-6

Mentor. (2006). *Calibre by mentor graphics.* `https://mentor.com/`. (Accessed: 2018-09-18)

Middlebrook, R. (1991). Low-entropy expressions: the key to design-oriented analysis. In *Proceedings frontiers in education twenty-first annual conference. engineering education in a new world order* (p. 399-403). doi: 10.1109/FIE.1991.187513

Muller, R., & Kamins, T. (2002). *Device electronics for integrated circuits* (3rd ed.). Wiley.

Nagashima, Y. (2013). *Elementary particle physics: Foundations of the standard model*. Wiley.

Numato. (2016, Feb). *Saturn - spartan 6 fpga development board with ddr sdram.* https://numato.com/docs/saturn-spartan-6-fpga-development-board-with-ddr-sdram/. (Accessed: 2017-12-10)

O'Connor, P., & De Geronimo, G. (1999). Prospects for charge sensitive amplifiers in scaled cmos. In *1999 ieee nuclear science symposium. conference record. 1999 nuclear science symposium and medical imaging conference (cat. no.99ch37019)* (Vol. 1, p. 88-93 vol.1). doi: 10.1109/NSSMIC.1999.842454

Pelgrom, M. J. M., Duinmaijer, A. C. J., & Welbers, A. P. G. (1989, Oct). Matching properties of mos transistors. *IEEE Journal of Solid-State Circuits*, *24*(5), 1433-1439. doi: 10.1109/JSSC.1989.572629

Radeka, V. (1968). Optimum signal-processing for pulse-amplitude spectrometry in the presence of high-rate effects and noise. *IEEE Transactions on Nuclear Science*, *15*(3), 455-470. doi: 10.1109/TNS.1968.4324970

Radeka, V. (1984). Semiconductor position-sensitive detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *226*(1), 209 - 218. Retrieved from http://www.sciencedirect.com/science/article/pii/0168900284901931 doi: https://doi.org/10.1016/0168-9002(84)90193-1

Razavi, B. (2001). *Design of analog cmos integrated circuits* (first ed.). McGraw-Hill.

Samtec. (2015, Jul). *Ultra low profile one piece array* (No. GFZ-30-01-G-10-AD). (Datasheet - Rev. 17)

Sansen, W. M. C., & Chang, Z. Y. (1990, Nov). Limits of low noise performance of detector readout front ends in cmos technology. *IEEE Transactions on Circuits and Systems*, *37*(11), 1375-1382. doi: 10.1109/31.62412

Serrano-Gotarredona, T., & Linares-Barranco, B. (2003, Sept). Cmos transistor mismatch model valid from weak to strong inversion. In *Esscirc 2004 - 29th european solid-state circuits conference (ieee cat. no.03ex705)* (p. 627-630). doi: 10.1109/ESSCIRC .2003.1257213

Silveira, F., Flandre, D., & Jespers, P. G. A. (1996, Sep). A gm/id based methodology for the design of cmos analog circuits and its application to the synthesis of a silicon-on-insulator micropower ota. *IEEE Journal of Solid-State Circuits*, *31*(9), 1314-1319. doi: 10.1109/4.535416

Snoeys, W., Anelli, G., Campbell, M., Cantatore, E., Faccio, F., Heijne, E. H. M., ... Wyllie, K. (2000, Dec). Integrated circuits for particle physics experiments. *IEEE Journal of Solid-State Circuits*, *35*(12), 2018-2030. doi: 10.1109/4.890318

Spieler, H. (2005). *Semiconductor detector systems*. Oxford University Press.

Texas-Instruments. (1999). *12-bit, parallel input digital-to-analog converter* (No. DAC7621). (Datasheet)

Texas-Instruments. (2004). *16-bit, 2 msps, unipolar input, micro power sampling analog-to-digital converter with parallel interface and reference* (No. ADS8411). (Datasheet)

Vittoz, E. A. (1994). *Design of analog-digital vlsi circuits for telecommunications and signal processing* (J. E. Franca & Y. Tsividis, Eds.). Prentice-Hall, Inc.

Xilinx. (2011). *Spartan-6 family overview* (No. XC6SLX16). (Datasheet)

ANNEXES

A. SECOND-ORDER TRANSFER FUNCTION OF THE CSA

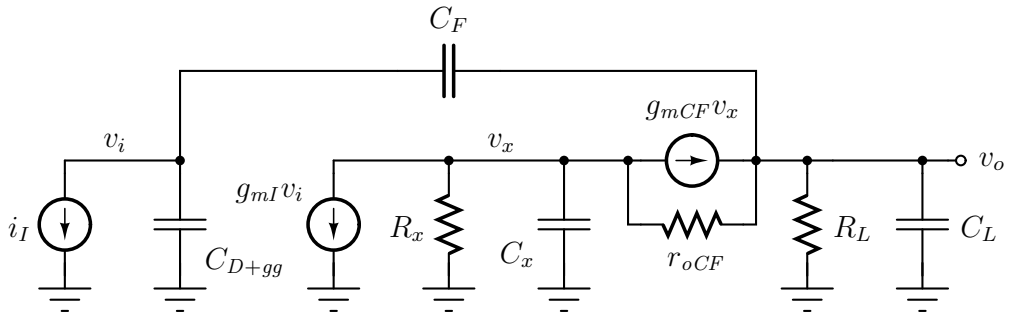


FIGURE A.1. Schematic of the NMOS-input folded-cascode amplifier for small-signal analysis, with capacitive feedback. An explicit capacitor to account for the non-dominant pole was added to the v_x node to analyze the second-order response of the circuit.

Let us consider the circuit presented in Figure A.1, which is a small-signal schematic of the NMOS-input folded cascode amplifier used in the Heisenberg chip in feedback configuration, but with the addition of an explicit capacitance C_x on the drain of the folding transistor M_F to account for the secondary non-dominant pole.

Through straightforward circuit analysis, and with reasonable circuit assumptions such as large intrinsic gain for the input device and the cascodes, it is possible to derive an expression for the transfer function of the circuit, as follows

$$\frac{v_O(s)}{i_I(s)} = \frac{-1}{s} \cdot \frac{(C_F \cdot C_x)s^2 + (g_{mCF} \cdot C_F)s - g_{mI} \cdot g_{mCF}}{a \cdot s^2 + b \cdot s + c} \tag{A.1}$$

where

$$a = C_x \cdot (C_F \cdot C_{D+gg} + C_F \cdot C_L + C_{D+gg} \cdot C_L) \tag{A.2}$$

$$b = g_{mCF} \cdot (C_F \cdot C_{D+gg} + C_F \cdot C_L + C_{D+gg} \cdot C_L) \tag{A.3}$$

$$c = g_{mCF} \cdot g_{mI} \cdot C_F \tag{A.4}$$

Since the denominator is a second degree polynomial, the roots can be trivially obtained

$$s = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{A.5}$$

Or, alternatively

$$s = \frac{-b}{2a} \pm \sqrt{\frac{c}{a}} \cdot \sqrt{\frac{b^2}{4ac} - 1} \tag{A.6}$$

A second-order circuit is typically characterized by its natural resonant frequency (ω_0) and its damping factor (ζ), as follows

$$s = -\zeta\omega_0 \pm \omega_0\sqrt{\zeta^2 - 1} \tag{A.7}$$

From (A.6) and (A.7), it can be observed that

$$\omega_0 = \sqrt{\frac{c}{a}}, \quad \zeta = \frac{b}{2\sqrt{ac}} \tag{A.8}$$

Finally, the damping factor for the second-order circuit presented in Figure A.1 can be calculated to be

$$\zeta^2 = \frac{1}{4} \cdot \frac{g_{mCF}}{C_x} \cdot \frac{C_L \cdot C_F + C_L \cdot C_{D+gg} + C_F \cdot C_{D+gg}}{g_{mI} \cdot C_F} \tag{A.9}$$