SYNTHESIS AND MONOLITHIC INTEGRATION OF ANALOGUE SIGNAL PROCESSING NETWORKS

by

Piotr Jan Osuch

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Electronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering

Faculty of Engineering, Built Environment and Information Technology

UNIVERSITY OF PRETORIA

December 2017

SUMMARY

SYNTHESIS AND MONOLITHIC INTEGRATION OF ANALOGUE SIGNAL PROCESSING NETWORKS

by

Piotr Jan Osuch

Supervisor: Dr Tinus Stander

Department: Electrical, Electronic and Computer Engineering

University: University of Pretoria

Degree: Philosophiae Doctor (Electronic Engineering)

Keywords: All-pass networks, analog integrated circuits, analog processing circuits, analogue signal processing, BiCMOS integrated circuits, CMOS integrated circuits, current-mode circuits, design optimization, group-delay engineering, millimeter wave integrated circuits, ultra-wideband technology.

Data traffic of future 5G telecommunication systems is projected to increase 10 000-fold compared to current rates. 5G fronthaul links are therefore expected to operate in the mm-wave spectrum with some preliminary International Telecommunication Union specifications set for the 71-76 and 81-86 GHz bands. Processing 5 GHz as a single contiguous band in real-time, using existing digital signal processing (DSP) systems, is exceedingly challenging. A similar challenge exists in radio astronomy, with the Square Kilometer Array project expecting data throughput rates of 15 Tbits/s at its completion. Speed improvements on existing state-of-the-art DSPs of 2-3 orders of magnitude are therefore required to meet future demands.

One possible mitigating approach to processing wideband data in real-time is to replace some DSP blocks with analog signal processing (ASP) equivalents, since analogue devices outperform their digital counterparts in terms of cost, power consumption and the maximum attainable bandwidth. The fundamental building block of any ASP is an all-pass network of

prescribed response, which can always be synthesized by cascaded first- and second-order all-pass sections (with two cascaded first-order sections being a special case of the latter). The monolithic integration of all-pass networks in commercial CMOS and BiCMOS technology nodes is a key consideration for commercial adaptation of ASPs, since it supports mass production at reduced costs and operating power requirements, making the ASP approach feasible. However, this integration has presented a number of yet unsolved challenges.

Firstly, the state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks. In this work an analytically-based solution to the synthesis problem is presented that produces an all-pass network with a response approximating the required group delay to within an arbitrary minimax error. This method is shown to work for any physical realization of second-order all-pass elements, is guaranteed to converge to a global optimum solution without any choice of seed values as an input, and allows synthesis of pre-defined networks described either analytically or numerically.

Secondly, second-order all-pass networks are currently primarily implemented in off-chip planar media, which is unsuited for high volume production. Component sensitivity, process tolerances and on-chip parasitics often make proposed on-chip designs impractical. Consequently, to date, no measured results of a dispersive on-chip second-order all-pass network suitable for ASP applications (delay Q-value (Q_D) larger than 1) have been presented in either CMOS or BiCMOS technology nodes. In this work, the first ever on-chip CMOS second-order all-pass network is proposed with a measured Q_D-value larger than 1. Measurements indicate a post-tuning bandwidth of 280 MHz, peak-to-nominal delay variation of 10 ns, Q_D-value of 1.15 and magnitude variation of 3.1 dB. An active on-chip mm-wave second-order all-pass network is further demonstrated in a 130 nm SiGe BiCMOS technology node with a bandwidth of 40 GHz, peak-to-nominal delay of 62 ps, Q_D-value of 3.6 and a magnitude ripple of 1.4 dB. This is the first time that measurement results of a mm-wave bandwidth second-order all-pass network have been reported.

This work therefore presents the first step to monolithically integrating ASP solutions to conventional DSP problems, thereby enabling ultra-wideband signal processing on-chip in commercial technology nodes.

ACKNOWLEDGEMENT

To God for His love.

To my parents, Elżbieta and Krzysztof Osuch, for their upbringing and instilling in me from a young age the deep desire to question and to never stop searching. To my mother for everything, for giving herself, the list is infinite. To my father for teaching me the art of logical thought, for setting in motion, in me, an unquenchable thirst for Truth. To my brother, Stanisław Osuch, for always being there, no matter what. For his thought inspiring comments. For those nights spent under the Drakensberg sky in conversation. To Fr. Bogdan Wilkaniec, for his upbringing, for being a second father, for shaping me, for showing me how to live, for being my mentor and best friend. For shaping me into the person I am today. To all my friends, who have shaped me over the years - to all the many conversations under the African sun with a glass of red wine.

To my study advisor, Dr Tinus Stander, who has lead me through my postgraduate career – for all your support, comments and thought-provoking questions. I have been privileged to have had you as my study leader. Thank you for giving me freedom in my investigation and guiding me when necessary. That takes courage, thank you.

For

ANNA,

my love for you shall live forever

LIST OF ABBREVIATIONS

ADC	Analog-to-Digital-Converter
ams AG	Austriamicrosystems AG
ASP	Analog Signal Processor / Processing
BEOL	Back-End-of-Line
BER	Bit Error Rate
BiCMOS	Bipolar Complementary Metal-Oxide Semiconductor
BPF	Bandpass Filter
BSIM	Berkeley Short-Channel IGFET model
CCI	First Generation Current Conveyor
CCII	Second Generation Current Conveyor
CMOS	Complementary Metal-Oxide Semiconductor
DAC	Digital to Analog Converter
DCMA	Dispersion Code Multiple Access
DDS	Dispersion Delay Structure
DRC	Design Rule Check
DSP	Digital Signal Processor / Processing
DUT	Device Under Test
ESD	Electrostatic Discharge
FFT	Fast Fourier Transform
GSG	Ground-Signal-Ground
HBT	Heterojunction Bipolar Transistor
ITU	International Telecommunications Union
LVS	Layout Versus Schematic
MATLAB	Matrix Laboratory
MC	Monte Carlo

MIM	Metal-Insulator-Metal
MIMO	Multiple Input Multiple Output
MOS	Metal-Oxide Semiconductor
NDA	Non-Disclosure Agreement
NPN	Negative-Positive-Negative
OFDM	Orthogonal Frequency Division Multiplexing
PCB	Printed Circuit Board
PDK	Process Design Kit
PNP	Positive-Negative-Positive
PPM	Pulse Position Modulation
PSU	Power Supply Unit
RTFT	Real-Time Fourier Transform
SKA	Square Kilometer Array
SNR	Signal-to-Noise Ratio
SOLT	Short-Open-Load-Thru
SOTA	State-of-the-Art
TEM	Transverse Electromagnetic
THD	Total-Harmonic-Distortion

TABLE OF CONTENTS

CHAPTER 1   INTRODUCTION

1.1   BACKGROUND TO THE RESEARCH

1.1.1   Context of the problem

Data traffic of future telecommunication systems is projected to increase 10 000-fold in the next 20 years [1]. Various approaches have been proposed to meet the increase in throughput and can be grouped into three categories, namely: improving spectral efficiency, increasing spectrum availability and the massive densification of small cells. Spectral efficiency can be improved by techniques such as multiple input multiple output (MIMO), carrier aggregation, coordinated multipoint, heterogeneous networks, authorized shared access, as well as various interference cancellation techniques. This approach will, at best, only provide a 1000-fold speed improvement [1]. Increasing spectrum availability by moving to higher mm-wave frequencies (30 – 300 GHz) is, therefore, inevitable. In turn, the migration to mm-wave frequencies, practically, also implies the massive densification of the network through the deployment of small cells, due to the high atmospheric attenuation of mm-waves compared to microwaves. Next generation 5G fronthaul links will therefore be comprised of a mesh network of small-cells operating in the mm-wave spectrum with some preliminary International Telecommunication Union (ITU) specifications set for the V-band (57-66 GHz) and E-band (70-90 GHz) where contiguous bands of up to 5 GHz are available [1]. At these modulation bandwidths, DSP-based design solutions for telecommunication systems are exceedingly challenging. Moreover, the digitization process would require sampling speeds of at least 10 GSPS, currently only achievable at the cost of sacrificing resolution and dynamic range [2] or through interleaving, which increases cost and power consumption. A similar challenge exists in radio astronomy, with the Square Kilometer Array (SKA) project expecting data throughput rates of 15 Tbits/s at its completion [3]–[7].

Since these rates are higher than those that can be analyzed or stored affordably with current practices, this is termed a "big data" problem [8]. Speed improvements on existing state-of-the-art DSPs of 2-3 orders of magnitude are therefore required to meet future demands. Improving spectral efficiency by means of more sophisticated data processing algorithms, has only partially alleviated processing requirements, justifying the need for higher bandwidth hardware. Since modern DSP bandwidths are limited by the capabilities of CMOS technology nodes [9], increasing processing speeds is, in principle, no longer a viable solution to meet the big data problem.

In consequence, replacing traditional DSP tasks with ASP equivalents has seen renewed interest in the literature [5], [10], as analogue devices outperform their digital counterparts in terms of cost, power consumption, and the maximum attainable bandwidth [10]. Various applications where APSs could replace or ease current DSP tasks have been proposed, such as real-time spectrum analysis, frequency sniffing, beam steering, pulse position modulation (PPM), distortion-encoding multiple-access (DCMA), interference cancelation, time-stretching and compression (for alleviating the burden on analogue-to-digital-converters (ADCs) and enhancing the signal-to-noise ratio (SNR)) as well as signal de-dispersion [11]–[24]. The fundamental building block of any ASP is a *dispersive* all-pass network of prescribed response [11], [12]. For example, all-pass networks with non-constant linear group delays are used for real-time spectrum analysis [11], [13]–[17], M-ary PPM [18], time-stretching and compression systems [19], [20], stepped group delays for frequency sniffing with potential applications in cognitive radio [21], general polynomial delay functions for arbitrary frequency beam scanning in antenna arrays [22], [23] and Chebyshev delays for DCMA communication channels [12]. Quasi-arbitrary group delay functions are further used in signal de-dispersion [24]. Although constant delay functions can be employed in IIR-type analog processing systems, such as beamforming [25], this work focuses only on ASPs employing dispersive delay functions. IIR-type ASP solutions experience similar drawbacks as purely digital implementations and do not benefit from the advantages of purely analog implementations. Moreover, IIR implementations can be seen as quasi-digital, quasi-analog systems. Therefore, in this work ASP will refer only to purely analog implementations [11], which, as will be shown, can only be realized with dispersive delay networks.

1.1.2 Research gap

Any delay function can be synthesized by cascaded first- and second-order all-pass networks [26], [27], with two cascaded first-order networks being a special case of the latter. Although various off-chip realizations of all-pass networks have been proposed in the peer-review literature [12], [13], [24], [28]–[32], their monolithic integration in commercial CMOS and BiCMOS technology nodes has presented a number of yet unsolved challenges. Implementing all-pass networks on-chip in a commercial technology node allows mass production at reduced costs and operating power requirements which is key to successfully replacing certain DSP tasks with ASP equivalents. However, several research gaps must be addressed before the synthesis and monolithic integration of all-pass networks is possible.

Firstly, the state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks [10], [27], [31]–[39]. Synthesizing a minimum-order network is particularly important for on-chip implementations where reducing costs and operating power requirements are key considerations, as justified earlier. Existing analytical approaches are sensitive to both the desired group delay function and the choice of initial solutions, often failing to converge for practical cases [27], [31], [32], [40], [41]. Numerical synthesis techniques also suffer from convergence issues and it is impossible to guarantee that the obtained solution is optimal (minimum-order). Furthermore, numerical approaches are often computationally intensive such as those based on differential evolution or genetic algorithms.

Secondly, second-order all-pass networks, the building blocks of ASPs, have traditionally been limited to off-chip media which are unsuited for volume production. Moreover, the high losses of on-chip distributed line elements make these approaches unfeasible on-chip [40]. In recognition of this, numerous active implementations of all-pass networks have been proposed [29], [40], [42]–[57], suitable for monolithic integration. However, component sensitivity, process tolerances and on-chip parasitics often make the proposed designs impractical. Consequently, to date, no measured results of an on-chip second-order all-pass

network, suitable for ASP applications (delay Q-value (Q_D) larger than 1), have been presented in either CMOS or BiCMOS technology nodes.

1.2   RESEARCH OBJECTIVE AND QUESTIONS

This thesis investigates, firstly, the synthesis of quasi-arbitrary group delay functions with a prescribed delay response (corresponding to certain ASP applications) using a minimum-order all-pass network, and secondly, the practical microelectronic implementation of the all-pass network on-chip, in the commercial CMOS and BiCMOS technology nodes. This will enable monolithic integration of ASPs on-chip, making the analogue approach a viable alternative to existing DSP techniques.

The primary research question addressed in this thesis can therefore be stated as follows.

How can a quasi-arbitrary group delay function be synthesized with a minimum-order network and implemented on-chip in commercial CMOS and BiCMOS technology nodes?

The following derived (secondary) research questions arise.

1. What class of delay functions can be approximated by all-pass networks? What is the minimum-order approximation? What is the error of approximation? How can this error be minimized?

2. How can existing all-pass networks be improved to obtain larger delay Q-value (Q_D)? Is an inductorless solution practically viable?

3. How do process tolerances and parasitic elements affect the design of the all-pass network? What can be done to reduce the network's sensitivity to these non-ideal effects? If sufficient sensitivity reduction is impossible, can post-production tunability be used to compensate for these effects?

4. What can be done to increase the bandwidth of the all-pass network? Is a mm-wave bandwidth solution possible?

1.3   METHODOLOGY

First, a literature review is conducted and shortcomings in the state-of-the-art identified. The literature review is presented in three parts: delay function network synthesis, physical realization of all-pass networks and ASP applications.

Next, a group delay function network synthesis theory is presented. The synthesis theory produces a minimum-order all-pass network approximating the desired delay function to within an arbitrary minimax error. This method is shown to work for any physical realization of second-order all-pass elements, is guaranteed to converge to a global optimum solution without any choice of seed values as an input, and allows synthesis of pre-defined networks described both analytically and numerically.

Then, CMOS and BiCMOS implementations of second-order all-pass networks, the building blocks of any higher-order all-pass network, are proposed and demonstrated with measured results, achieving delay Q-values above 1 for the first time in published literature. The effects of process tolerances and parasitic effects are investigated in implementing a practical system. Methods of increasing the bandwidth, a key requirement of all-pass networks with application in ASPs, are investigated. The feasibility of implementing a mm-wave bandwidth all-pass network is also further investigated.

Lastly, the research questions are evaluated and future work proposed.

1.4   OVERVIEW OF STUDY

This thesis is organized as follows.

1. Chapter 1: *Introduction*

   This Chapter presents the context of the research problem, research gaps in the peer-reviewed literature and the motivation for this research. Research objectives are specified and the research questions asked. Finally, the contribution of this work to the field of Electronic Engineering is summarized and research outputs listed.

2. Chapter 2: *Literature review*

In this Chapter, the existing body of knowledge is presented and critically analyzed in the context of the research questions. First, existing analytical and numerical all-pass network synthesis techniques are summarized and their various shortcomings examined. Next, various passive and active realizations of all-pass networks, both on-chip and off-chip, are examined in the literature. A literature review of various building blocks of all-pass networks and associated design challenges, such as multi-loop circuit stability analysis, is also conducted.

3. Chapter 3: *Research methodology*

The research methodology followed in this work to answer the research questions is presented in this Chapter. The design flow, techniques and corresponding software packages used are discussed. The fabrication process of the submitted designs is discussed and the measurement setup and procedure presented.

4. Chapter 4: *Mathematical synthesis*

The state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks. In this Chapter, an analytical solution to this problem is presented that produces an all-pass network with a response approximating the required group delay to within an arbitrary minimax error. This method is shown to work for any physical realization of second-order all-pass elements, is guaranteed to converge to a global optimum solution without any choice of seed values as an input, and allows synthesis of pre-defined networks described either analytically or numerically. An application of the proposed method is also demonstrated by reducing the delay variation of a practical system by any desired amount, and compared to state-of-the-art methods in comparison examples.

5. Chapter 5: *On-chip implementation of all-pass networks*

In this Chapter, a novel active on-chip CMOS second-order all-pass network is proposed. A post-production tuning mechanism that uses a genetic local optimizer to control bias voltages to tune the design is used to account for process tolerances and parasitics. The design is prototyped in a 0.35 µm CMOS process and measurements indicate a post-tuning bandwidth of 280 MHz, peak-to-nominal delay variation of 10 ns, magnitude variation of 3.1 dB and a Q_D-factor of 1.15. This is the first time that measurement results have been reported for an active on-chip second-order all-pass network with $Q_D > 1$.

The proposed all-pass network makes use of a novel second-generation current conveyor (CCII), itself a versatile analogue signal processing element that could see application in future wideband telecommunication and data processing solutions. Despite this versatility, CCIIs have seen limited commercial adoption due to, among other reasons, high sensitivity to process tolerances, layout parasitics, and possible instability. This Chapter therefore presents, a practical CCII design methodology that incorporates accurate foundry device models, process corners, layout parasitics, and explicit multi-loop feedback stability analysis into the design flow. A post-production tunable feedback compensation network is further included to enable post-production stability and peaking control. The method is experimentally verified by producing a high precision (1.15% transfer error), high-bandwidth (> 500 MHz) CCII+ in 0.35µm CMOS with port X input resistance (R_X) < 5 Ω, along with an astable counter-example designed by state-of-the-art methods.

Finally, the bandwidth extension of all-pass networks is examined by an active on-chip mm-wave second-order all-pass network in a 130 nm SiGe BiCMOS technology node with a measured bandwidth of 40 GHz, peak-to-nominal delay of 62 ps, magnitude ripple of 1.4 dB and Q_D-factor of 3.6, which is the highest in reported literature. A layout-focused design methodology incorporating layout parasitics is followed.

6. Chapter 6: *Conclusion*

    This Chapter concludes the thesis and provides a critical evaluation of the research questions. Suggestions for improvements and future work are made.

1.5   DELIMITATIONS TO THE SCOPE OF THE RESEARCH

The scope of this work is limited to developing a synthesis theory for minimum-order all-pass networks and proposing an on-chip implementation of second-order all-pass networks in CMOS and BiCMOS technology nodes. Implementations of specific ASPs using the proposed synthesis theory and second-order all-pass building blocks is beyond the scope of this work. ASP applications are reviewed in Chapter 2 in order to establish context and requirements, however no ASPs are implemented in this work.

1.6   CONTRIBUTION TO THE FIELD

1. The state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks. In this work, a new numerical synthesis procedure of minimum-order cascaded all-pass networks having a quasi-arbitrary group delay response is presented, making the following contributions to the state-of-the-art.

    a. The proposed method does not require an initial value set. Convergence is therefore not dependent on an appropriate selection.

    b. The resulting group delay function is an approximation of the required response to within any arbitrarily specified maximum delay variation across the passband, extending on methods in literature [27], [28], [31]–[33], [36], [58] where the network's order is chosen and not, explicitly, the resulting maximum delay variation.

c. The method is implementation abstracted in that any theoretical, circuit, or parametrized numerical model description of a second-order all-pass section can be applied in the algorithm.

d. Due to the underlying analytical nature of the approach, rapid convergence is achieved; typically, within 20 iterations for practical cases (as will be shown in the examples). This is an order of magnitude fewer iterations required by more general state-of-the-art optimization methods (genetic algorithm and simulated annealing), as is shown in comparative examples.

2. Second-order all-pass delay networks – the building blocks of ASPs – are usually implemented in off-chip planar media, which is unsuited for volume production. In this work, a novel active on-chip CMOS second-order all-pass network is proposed, making the following contributions to the state-of-the-art.

a. An inductorless second-order all-pass network is synthesized which can be tuned post-production to concurrently account for non-unity voltage and current following CCII non-idealities (A_v and A_i), as well as a non-zero port X input resistance (R_X), which is a unique contribution to the state-of-the-art.

b. Varactors and NMOS resistors are used to provide post-production tunability to account for process tolerances and parasitic elements. A post-production *automated* tuning method is further proposed, whereby measured data from a VNA is used in a real-time genetic local optimizer on a PC (which controls various bias voltages using a digital-to-analogue (DAC) card) to tune the physical all-pass network. It is shown that this step is crucial to realizing a practical system, leading to the first ever measured results of an active on-chip CMOS second-order all-pass network with a Q_D-value larger than 1.

c. Measurements indicate a post-tuning bandwidth of 280 MHz, peak-to-nominal delay variation of 10 ns, a Q_D-value of 1.15 and magnitude variation of 3.1 dB. This is the first time that measured results with a Q_D-value larger than 1 have been reported in literature.

3. CCIIs are versatile analogue signal processing elements that could see application in future wideband telecommunication and data processing solutions. Despite this versatility, CCIIs have seen limited commercial adoption due to, among other reasons, high sensitivity to process tolerances, layout parasitics, and possible instability. For the first time, a CCII design methodology is presented that incorporates accurate foundry device models, process corners, layout parasitics, and explicit multi-loop feedback stability analysis into the design flow. The following contributions are made to the state-of-the-art.

    a. A post-production tunable RC compensation network is used to reduce gain peaking and to tune the phase margin, to ensure stability despite process variations. An optimization based synthesis technique is presented and implemented in 0.35 µm CMOS, incorporating multi-loop feedback analysis.

    b. A multi-loop cut-based numerical feedback analysis technique is applied combining single-loop [59] and multi-loop [60], [61] feedback analysis and applied in a practical CCII microelectronic design which is a novel contribution to the state-of-the-art.

4. An active on-chip mm-wave second-order all-pass network is demonstrated in a 130 nm SiGe BiCMOS technology node. Measured results of a mm-wave bandwidth all-pass network with a Q_D-value larger than 1 are presented for the first time. The following improvements are made to the state-of-the-art.

    a. A bandwidth of 40 GHz with a peak-to-nominal delay of 62 ps and a magnitude ripple of 1.4 dB is achieved, improving significantly on the state-of-the-art. A layout-focused design methodology incorporating layout parasitics, distributed effects and accurate device models is followed. This is the first time that measurement results of a mm-wave second-order all-pass network have been reported. Furthermore, a Q_D-value of 3.6 is measured, which is the highest in reported literature.

1.7   RESEARCH OUTPUTS

The following peer-reviewed journal articles related to this work have been published.

1. P. J. Osuch and T. Stander, "A geometric approach to group delay network synthesis," *Radioengineering*, vol. 25, no. 2, pp. 351–364, 2016.

2. P. J. Osuch and T. Stander, "A millimeter-wave second-order all-pass delay network in BiCMOS," *IEEE Microwave and Wireless Components Letters (MWCL)*, 2018 (accepted).

The following peer-reviewed journal articles related to this work have been submitted for publication.

1. P. J. Osuch and T. Stander, "A high-Q second-order all-pass delay network in CMOS," *IET Circuits, Devices & Systems*, 2018 (submitted).

2. P. J. Osuch and T. Stander, "Wideband high-precision CCII+ with stability and peaking control," *Elsevier Microelectronics Journal*, 2018 (submitted).

The following peer-reviewed conference proceedings related to this work have been accepted for publication.

1. P. J. Osuch and T. Stander, "An on-chip post-production tunable group delay equaliser," in *International Conference on Actual Problems of Electron Devices Engineering* (APEDE) 2014, vol. 1, pp. 177–184, 2014.

All journals are listed by Thomson Reuters Web of Knowledge (formerly ISI).

1.8   CONCLUSION

Replacing traditional DSP tasks with ASP equivalents has seen renewed interest in the literature. The fundamental building block of any ASP is a delay network of prescribed response. The monolithic integration of all-pass networks in commercial CMOS and BiCMOS technology nodes has presented a number of yet unsolved challenges. Firstly, the state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks. Secondly, second-order all-pass networks, the building blocks of ASPs, have traditionally been limited to off-chip media which are unsuited for volume production. This work will focus on addressing both of these issues thus laying the foundations for the future monolithic integration of high-speed ASP systems on-chip in commercial CMOS and BiCMOS technology nodes.

CHAPTER 2   LITERATURE REVIEW

2.1   INTRODUCTION

A literature review of the research topic is presented in this Chapter. First, the second-order all-pass network (a fundamental building block of any ASP) is defined and characterized. Next, analytical and numerical synthesis techniques of higher-order delay functions using second-order all-pass networks are reviewed and limitations identified. The literature on physical implementations of second-order all-pass networks is then reviewed. The review is divided into the passive and active realization of all-pass networks in soft-substrate and on-chip. The CCII element is identified as a promising building block of active all-pass networks on-chip, justifying a literature review of the CCII, subsequently leading to the review of single- and multi-loop stability analysis of electronic circuits. Finally, the research is contextualized with a review of state-of-of-the-art (SOTA) methods for analog signal processing with all-pass networks such as real-time spectrum analysis, frequency sniffing, arbitrary frequency beam scanning, dispersion code multiple access (DCMA) and M-ary pulse position modulation (PPM). This is presented at the end of the Chapter after the introduction of all-pass networks and their building blocks as knowledge of delay networks is necessary to understand their system application in ASPs.

2.2   DELAY FUNCTION NETWORK SYNTHESIS

The fundamental building block of any analogue signal processor is a *dispersive* all-pass network [28], [31], [32] with a frequency-time delay function, or group delay, of prescribed response. Ideal all-pass networks transmit all frequency components of an input signal to the output without changing their magnitude in any way, whilst simultaneously changing the phase response. An all-pass network therefore delays each Fourier component of the input

signal by a certain amount, depending on its group delay function. Example passive circuit representations of first- and second-order all-pass sections are shown in Figure 2.1.
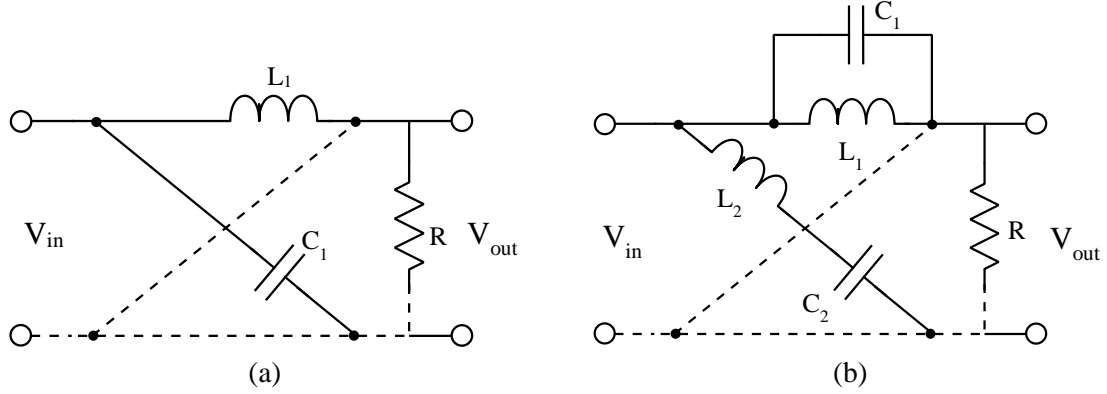
Figure 2.1. Circuit representation of (a) first- and (b) second-order all-pass networks [36].

The transfer function of a second-order all-pass network may be written as:

$$T(s) = \frac{V_{out}}{V_{in}} = K \cdot \frac{s_n^2 - s_n / Q_D + 1}{s_n^2 + s_n / Q_D + 1},$$

(2.1)

where $s_n = s/\omega_0$, $\omega_0 = 2\pi f_0$ is the center frequency of the second-order all-pass delay function, and $\Delta\tau$ is the corresponding peak-to-nominal group delay. The quantity Q_D in (2.1) is referred to as the Q_D-value, defined as [62]:

$$Q_D = \frac{\Delta\tau \cdot \omega_0}{4}$$

(2.2)

and should not be confused with the transmission Q-factor [63]. The Q_D-factor determines whether the poles and zeros of the transfer function in (2.1) are real or complex. If $Q_D > 1/2$ then the poles and zeros contain a non-zero imaginary component. The pole and zero locations of the transfer function in (2.1) are shown on the Argand plane of Figure 2.2, where σ_{px} is the real component and ω_{px} the imaginary component.

The group delay of a linear system is defined as:

$$\tau_T = -\frac{\partial \angle T(s)}{\partial \omega}\bigg|_{s=j\omega},$$

(2.3)

where $\angle T(s)\big|_{s=j\omega}$ is the system's phase response. Substituting (2.1) into (2.3) results in the well-known second-order all-pass delay response as shown in Figure 2.2.
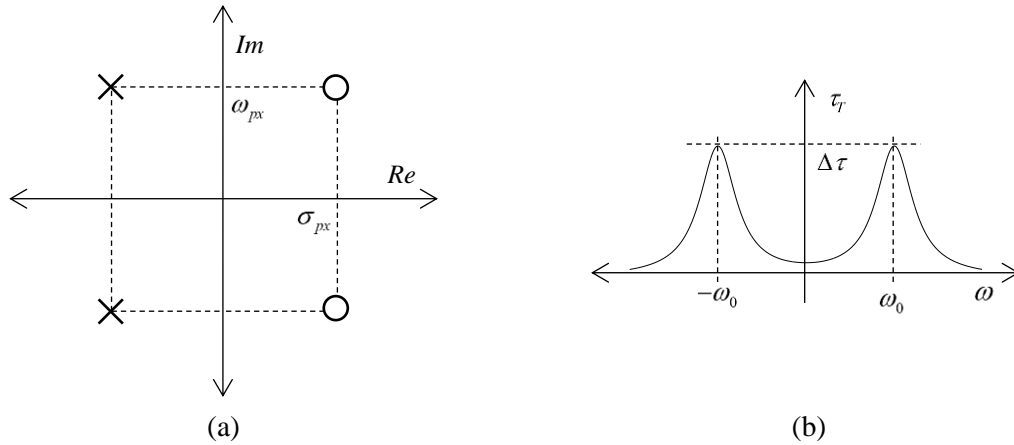

(a)   (b)

Figure 2.2. (a) Pole and zero locations of a second-order all-pass network and (b) the resulting group delay response.

Two cascaded first-order all-pass networks are, therefore, a special case of the second-order all-pass section (in the group delay sense) where the poles and zeros in Figure 2.2 merge at the real-axis as the imaginary component disappears. It is important to note, however, that cascading two first-order networks does not result in a generalized second-order all-pass section, as can be seen by considering Figure 2.2, since $\omega_0 = 0$.

Any all-pass network can be synthesized by cascading first- and second-order all-pass sections, to obtain a quasi-arbitrary group delay function, approximating the required response to within a constant [27], [36]. It is important to emphasize that the fundamental building block of ASPs is a *dispersive* all-pass network and not flat delay networks such as those used in true time delay networks [64], [65]. The delays of cascaded all-pass sections add linearly due to the additive property of exponent multiplication. This synthesized network may then be inserted into a system to perform an ASP function. This approach is compatible with several existing techniques [10], [33]–[35], [40], [41], [58], [62] for the synthesis of group delay networks. For example, a non-constant linear group delay can be

used for real-time Fourier transformation, a stepped group delay for distortionless frequency discrimination, and a Chebyshev delay for distortion-encoding multiple-access communication channels [11], easing the burden on the system DSP.

In practically implementing a first- or second-order all-pass network, a trade-off is made between the maximum achievable all-pass delay ($\Delta\tau$) and the insertion loss of the network at the resonant frequency, as is described later in Section 2.3. The Q_D-value is used in this work as a figure of merit to gauge the maximum achievable $\Delta\tau$ independent of ω_0. In literature $\Delta\tau$ and ω_0 are typically reported, in which case (2.2) is used in this work to calculate the more meaningful parameter Q_D. A Q_D-value larger than 1 is required for many ASP applications [11]. Real-time spectrum analysis [11], [13] requires a $Q_D > 1.57$. In frequency scanning antenna arrays, a Q_D of 2 has been shown to result in a mapping of 60°/GHz [23], with higher Q_D required for finer spatial resolution. In M-ary PPM, a Q_D of 3.14 is required to create a maximum delay of one pulse width [18]. The requirement for $Q_D > 1$ is further justified in detail in Section 2.6.

2.2.1   Analytical synthesis techniques

Initial methods for synthesizing continuous group delay functions (using analogue all-pass networks) were aimed at reducing the group delay variation of color television receivers [34]. These methods typically assume the form of design tables and graphs, as well as trial-and-error design approaches and are therefore limited to particular applications.

In an effort to address these limitations, analytical techniques for synthesizing the group delay of an electrical network have been developed [27], [31]–[33], [36], [37]. Of these, only one presents a rigorous theoretical treatment to finding minimum-order solutions to the synthesis problem [27]. This is done by approximating the required system's group delay characteristic (τ_s), to within a specified error, using the Tchebycheff polynomial series with an optimal number of all-pass sections:

$$\tau_s \approx -\sum_{m=0}^{n} C_{2m} T_{2m}\left(\frac{\omega}{\omega_B}\right) \forall\ \omega \in [-\omega_B : \omega_B], \text{ where}$$

$$T_{2m} = \cos(2\,m\,\phi) \text{ and} \tag{2.4}$$

$$\phi = \sin^{-1}\left(\frac{\omega}{\omega_B}\right),$$

where ω_B is the bandwidth over which the approximation is made and where $2n$ is the network order. Choosing a higher n leads to a smaller approximation error. It is shown in [27] that the coefficients C_{2m} satisfy the following relationship:

$$-\frac{\omega_B}{8}(C_{2m} + C_{2m} + 2) = \sum_{k=1}^{n}\left(\frac{1}{z_k^{2m+1}} + \frac{1}{(z_k^*)^{2m+1}}\right), \tag{2.5}$$

where z_k is a zero of the network satisfying the inequality $|z_k| > 1$. The equations in (2.4) and (2.5) lead to a non-liner system of $2n$ equations. In this method, convergence of the system of analytical equations is, however, sensitive to both the original group delay function and the choice of initial solutions, often failing to converge for practical cases [27]. This method is also limited to the lowpass case and provides poor convergence if the bandpass signal is treated as an extended lowpass spectrum [40], [41]. The approach presented in [31], [32] solves the approximation problem by generating a Hurwitz polynomial (with the desired phase response) at specified frequencies chosen *a priori*. However, no rigorous method of choosing these frequency points such that the resulting solution is minimum-order, is presented.

2.2.2   Numerical synthesis techniques

Due to these problems with state-of-the-art analytical synthesis methods, numerical approaches relying on optimization algorithms have been widely sought in the group-delay synthesis literature [10], [28], [36]–[38], [58], [66], [67] over the past two decades. The most popular approach [36] is based on approximating a desired group delay curve with a summation of second-order all-pass networks to within an arbitrary additive constant. The

theoretical approximation is expressed in terms of nonlinear constraints at specific points that are then perturbed to achieve quasi equi-ripple convergence. Convergence of this method is sensitive to the chosen initial value set, though this shortcoming is partially alleviated by a trial-and-error approach. This method also fails to account for deviations of practical realizations from theoretical models, which are of particular importance in the design of high-frequency systems [39]. A similar approach is presented in [37], with the exception that the imaginary components of the all-pass poles and zeros are assumed to be the frequency locations of the local maxima of the resulting delay curve. As with the method in [36], it is impossible to analytically determine the initial solution set which guarantees convergence. The approach presented in [66] is also similar to [36] albeit suitable for digital IIR implementations. An equi-ripple solution is sought by subdividing the optimization space into smaller local segments and developing decision rules to guide the optimizer to an equi-ripple solution. Rapid convergence with few iterations is obtained for higher-order cases. However, the algorithm performs poorly for lower-order equalization problems (typically the same number of iterations are required as for higher-order problems). Furthermore, the algorithm does not guarantee equi-ripple convergence. Other approaches using the differential evolution and genetic algorithms [10], [38], [58], [67] have also been proposed. Such approaches tend to be computationally intensive and are prone to converge to local minima, as opposed to a global optimum, as will be shown later in this work. Moreover, initial solutions are computed using a rule-of-thumb analytical approach based on experimental results [67] and not a rigorous theoretical treatment.

The preceding discussion highlights the limitations of existing quasi-arbitrary group delay function synthesis theories with all-pass networks. A numerical synthesis procedure is therefore required, which is based on an analytical approach and which does not suffer from the aforementioned limitations. Such an analytical approach should be able to guarantee an optimal solution in terms of network order as well as computational complexity, which is particularly important for on-chip applications where real-estate, cost and power consumption are of importance.

2.3   PHYSICAL REALISATIONS OF ALL-PASS NETWORKS

Various technologies have been reported in the literature that can be used to achieve dispersion effects, which can be used to physically implement the synthesized group delay responses discussed earlier [10]. These technologies can be grouped into those exploiting the dispersion of highly-dispersive media [11], [68], [69] and those using filter networks for group delay engineering. The network synthesis approach provides greater control over the resulting delay function and is better suited to electronic implementation. Network synthesized all-pass structures can further be subdivided into reflection-type and transmission-type networks, with the latter being better suited for integration into an ASP system due to its two-port integration [11]. Both passive and active transmission-type second-order all-pass networks have been proposed in the peer-reviewed literature [10], [12], [24], [28], [29], [40], [42]–[56], [62].

In practically implementing a first- or second-order all-pass network [10], [12], [24], [28], [29], [42]–[55], [62], a trade-off is made between the maximum achievable all-pass delay ($\Delta\tau$) and the insertion loss of the network at the resonant frequency. The Q_D-value is used as a figure of merit to gauge the maximum achievable $\Delta\tau$ independent of ω_0. A Q_D-value larger than 1 is required for many ASP applications [11] (as will be shown in Section 2.6) to provide sufficient frequency discrimination in time. However, this increases the insertion loss of the network at the resonant frequency and deteriorates the original assumption of all-pass transmission.

2.3.1   Passive realizations

Passive lumped element configurations are particularly vulnerable to the aforementioned problem of insertion loss at the resonant frequency due to low achievable practical Q-factors for lumped inductors. This is especially true for on-chip inductors with Q-factors typically less than 10 [30], [62]. Furthermore, to realize most practical $\Delta\tau$ values, unrealistically large component ratio spreads are required. This is particularly problematic on-chip because of stringent space limitations. Novel network transformations with simple design equations

leading to reduced component ratio spread have been presented in the recent literature [62] to address this problem.

One approach to improving on-chip inductor Q-factors has focused on emulating the inductor's behavior by means of active gyrator-C (Gm-C) realizations, requiring only capacitors and transistors. However, realizing floating Gm-C inductors (where neither end of the inductor is grounded) is challenging, since a parasitic RC network to ground is inevitably formed [70]. The second approach focuses on active Q-factor enhancement of passive inductors by means of negative resistors [71]. This leads to filters with high Q-factors, but inadequate bandwidths due to the negative resistors' strong frequency dependence [72]. To circumvent this distortion, a particularly attractive transformer-based Q-factor compensation technique has been proposed [73]. Various papers have demonstrated variations of this technique [73]–[77] with measured Q-factors as high as 380 at frequencies approaching 2 GHz at the cost of higher chip real-estate usage.

Another method of improving on-chip Q-factors is by using distributed element all-pass realizations. To this end, various microwave dispersion delay structures (DDSs) have been proposed as transmission line networks in soft substrate [10]–[12], [24], [28], [29], [31], [32], [40], [78]. Microwave C- and D-sections are popular approaches [10], [28], [31], [32], exhibiting large operational bandwidths and center frequencies in the microwave region. It has been shown that any arbitrary delay function supported by a transverse electromagnetic (TEM) commensurate microwave network can be realized with C-type and D-type all-pass sections to within an additive constant [35]. However, since quarter-wavelength transmission lines are used in these designs, an on-chip implementation is only realistic at mm-wave frequencies. Unfortunately, at these frequencies, on-chip losses become prohibitively large at resonance [40]. The microstrip-slotline stub C-section [24] has been shown to produce larger peak-to-peak (p-p) delay variations than C- and D-sections, for the same transmission loss at resonance. This improvement is however not sufficient to make an on-chip implementation practical. To circumvent this problem, active enhancement techniques for microwave DDSs have been proposed [12], [29]. The approach in [12] makes use of balanced loss-gain pairs to simultaneously amplify $\Delta\tau$ and equalize the transmission

loss. The design relies on ideal wideband amplifiers which are difficult to implement on-chip at microwave and mm-wave frequencies. On the other hand, in [29], $\Delta\tau$ is enhanced by passing the time signal several times through the DDS using a loop design. On each pass, the signal is amplified to compensate for transmission losses. This technique does not eliminate the transmission notch of the DDS element and, as a result, the notch is effectively enhanced on each signal pass through the loop. Both techniques could potentially be combined to overcome the above limitations and to implement actively enhanced on-chip all-pass networks with large $\Delta\tau$ values and acceptable transmission variations; however, this option has yet to be investigated in the literature.

2.3.2 Active realizations

In recognition of the disadvantages of passive realizations of all-pass networks, numerous active *inductorless* implementations of all-pass networks have been proposed [79]–[83]. In [79] an active inductorless approximation to a second-order all-pass network is proposed based on a CMOS inverter, and achieves a Q_D-value of 0.19 (extracted from the data presented). In [80] an LC network is implemented by using an active inductor, approximating a second-order delay response with an overall Q_D-value of 0.071. In [81]–[83] a special case of a second-order delay network is obtained with the resonant frequency at 0 Hz and therefore a Q_D-value of 0. To increase the achievable Q_D-value operational amplifier and CCII-based realizations have been proposed [29], [42]–[55] which can potentially achieve Q_D-values larger than 1. Op-amp based realizations are generally undesirable due to low bandwidths, sparking interest in current conveyor (CC) based implementations [43]. These devices exhibit higher bandwidth, greater linearity and lower power consumption than op-amps, making them better suited for implementation of second-order all-pass networks [29], [42]–[55]. The CC can be seen as the current-mode equivalent of an operational amplifier, overcoming the bandwidth limitations imposed by the gain-bandwidth product. Various circuits for first- and second-order all-pass delay elements using first- and second-generation current conveyors (CCI/CCII) have been proposed [42], [45], [47]–[55], [84]–[86] in literature. However, CCII non-idealities such as non-unity voltage mirroring (A_v) and current conveying (A_i), non-zero input resistance at port X (R_X), and finite values of output

resistance at ports Y and Z (R_Y and R_Z) make many of these designs impractical, even when using high-precision CCIIs [87]. This shortcoming is discussed in more detail in Section 2.4.2 but is apparently not addressed in literature, as many papers assume ideal CCII elements. In [43], [46] a synthesis approach based on partial fraction decomposition is presented and used to synthesize a second-order all-pass network with a minimal number of active devices which utilizes only one CCI/CCII element, thereby improving the achievable bandwidth. The sensitivity of the circuit to passive component variations is, however, still prohibitively large. Since on-chip component tolerances can be significant (the absolute values of well-capacitors vary around 10% and metal capacitors around roughly 25% [88] in CMOS processes), post-production tunability is necessitated to correct for these variations. Moreover, the design is sensitive to CCI/CCII non-idealities, particularly to a non-zero R_X (as is demonstrated in Section 5.2), requiring a high-precision CCII with a low R_X for practical applications.

In an effort to reduce the network's sensitivity, a tradeoff can be made between a minimum-order network (using fewer CCII elements) and a design with added design flexibility and potential tunability. In [47] a multifunction biquadratic filter is proposed which uses four CCIIs to implement a second-order all-pass network with low sensitivities to passive component tolerances. Grounded capacitors are further employed, which are more suitable for monolithic integration. The design has a low sensitivity to non-unity A_i which can be compensated for entirely by appropriate choice of RC values and only requires an $R_X < 50$ Ω, which is achievable using most CCIIs. However, non-unity A_v cannot be compensated for, and requires a voltage following precision within 6 % of unity for a magnitude variation lower than 0.5 dB, as will be shown in Section 5.2.2. In [51], [52], [54], [55] universal filters using only two CCIIs are proposed, which exhibit reduced sensitivities to both the passive and active components. In all cases a second-order all-pass network can be implemented. It can be shown that by appropriate selection of the RC components both a non-unity A_i and a non-zero R_X can be perfectly compensated for. However, a voltage following precision within 2.7 % of unity is required for an all-pass magnitude variation lower than 0.5 dB. Such a precision is only achievable using a high-precision CCII with feedback control, which complicates the design and reduces the achievable bandwidth [87]. In [48] a universal filter

is proposed using two CCIIs where an all-pass response can be realized by using an additional inverter. The design only uses two resistors and two capacitors and has a low sensitivity to RC variations. Similarly, in [50], a universal filter is proposed using two CCIIs which has a low component spread and low passive device sensitivities. The authors of the paper do not consider CCII non-idealities. In [48] and [50] the sensitivity to non-unity A_v and A_i is high and precisions within ~1% are required for a magnitude notch variation below 0.5 dB. It also requires an R_X lower than 10 Ω, achievable only with high precision CCIIs. This will be shown in more detail in Section 5.2.2.

The above discussion illustrates the importance, in the practical design of on-chip all-pass circuits, of considering and accounting for CCII non-idealities through appropriate selection of RC components. Furthermore, since both the CCII parameters and the RC components themselves change with process parameter variation, post-production tunability of the circuit is also necessary. Both of these considerations have, in the past, not been addressed in literature, which may explain the lack of measured results of a second-order all-pass network on-chip with a Q_D-value larger than 1.

Apart from the CCII-based realizations summarized above, a particularly attractive single transistor second-order all-pass network with inductor Q-factor enhancement has been proposed in [56] which alleviates many of the aforementioned shortcomings and enables much larger bandwidths. A bipolar transistor is used in a common emitter configuration, where the base-collector terminal is shunted by an LC resonant tank. Outside resonance the circuit behaves as a simple inverting amplifier (introducing a 180° phase shift) while at resonance the input signal is passed directly to the output terminal with a 0° phase shift. The transistor also compensates for series LC losses. However, to compensate for typical on-chip inductor losses, transistor transconductances of a few hundred mS are necessary. Achieving this transconductance with CMOS devices leads to impractical gate-width to gate-length ratios. For instance, in a 0.35 μm CMOS technology node, gate widths of a few millimeters are required. Furthermore, to achieve an optimal f_T, a bias current proportional to the gate width must be applied, resulting in an unrealistically large current. The design proposed in [56] is, therefore, only suitable for BiCMOS technology processes, though only a theoretical

design has been proposed with no physical on-chip implementation. At mm-wave frequencies, layout RCL parasitic components complicate the published design, which relies on zero length interconnects and ideal components. Significant additional design effort is, therefore, required to practically implement this single transistor all-pass network on-chip.

2.4   SECOND GENERATION CURRENT CONVEYORS

A promising approach to implementing second-order all-pass networks on-chip is based on the CCII as a building block of the circuit, as discussed in 2.3. In this Section, the SOTA methods for implementing CCIIs on-chip in both CMOS and BiCMOS technology nodes will be examined.

2.4.1   Introduction and operating principles

Since its introduction in the 1970's [89] the CCII has proven to be a versatile building block in analogue electronic design. Systems employing the CCII include active filters [90], impedance converters such as negative impedance converters (NICs), active inductance circuits [89], operational amplifiers, gyrators, mode converters, oscillators [89], [91], controlled voltage and current sources [89], analogue computation elements such as integrators and differentiators [92], current-mode DACs and variable-state filters [93]. Moreover, with the continuously increasing speed requirements of modern telecommunication and data processing systems and the bandwidth, power and cost advantages of analogue solutions over their digital counterparts [10]–[12], CCIIs could play an important role in future high-speed microelectronic design.

A CCII is a three-port network, as shown in Figure 2.3, defined by the hybrid matrix:

$$\begin{bmatrix} I_Y \\ V_X \\ I_Z \end{bmatrix} = \begin{bmatrix} Y_Y & 0 & 0 \\ A_v & Z_X & 0 \\ 0 & A_i & Y_Z \end{bmatrix} \begin{bmatrix} V_Y \\ I_X \\ V_Z \end{bmatrix} \tag{2.6}$$

where A_v and Ai represent the voltage and current transfer ratios respectively, Y_Y the admittance at port Y, Z_X the impedance at port X and Y_Z the admittance at port Z [91].
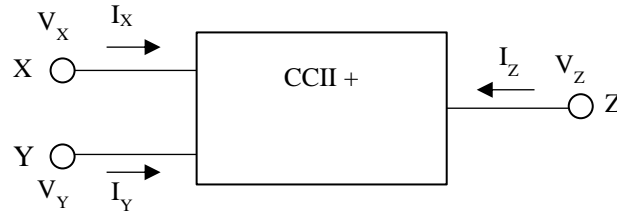

Figure 2.3. Second generation current conveyor.

In an ideal CCII the voltage at port X follows the voltage (in terms of magnitude and phase) at port Y (from a common reference) and the current flowing into port Z equals the current flowing into port X. Furthermore, ideally Y_Y, Y_Z and Z_X approach zero. A distinction between a CCII+ and CCII- is often made, where the former indicates that $A_i > 0$ and the latter that $A_i < 0$. It is desirable in many applications to minimize the deviation of A_v and A_i from unity (here referred to as the transfer error), minimize R_X and maximize R_Z and R_Y (the resistive components of Z_X, Y_Z and Y_Y respectively) over some desired bandwidth.

2.4.2 CMOS realizations

The first implementations of CCIIs used operational amplifiers as functional blocks due to a lack of high-quality PNP devices in bipolar technology nodes at the time [92]. Research efforts were subsequently focused on finding equivalent CMOS realizations [94], [95]. The CCII in [93] has formed the baseline of many subsequent designs, presenting an R_X of less than 50 Ω and a transfer error of below 1%. The CCII proposed in [94] improves on this transfer precision by employing a dual feedback mechanism as well as a novel high-swing cascode current mirror. The added complexity, however, reduces the achievable bandwidth to ~1 MHz in 0.6 µm CMOS. Furthermore, the efficacy of this approach is only demonstrated in simulation, and not in measurement. In [96], on the other hand, the CCII from [93] is simplified by incorporating the common drain feedback amplifier into the differential voltage follower stage. This has the effect of improving the bandwidth (20 MHz

in 1.2 µm CMOS) and reducing R_X to 0.3 Ω. An RC compensation network is also used, to reduce peaking.

In an effort to improve bandwidth, an implementation based on a source follower and cascode current mirror stage is proposed in [97] achieving a bandwidth of 700 MHz in a 1.2 µm CMOS node (for suitable load terminations). However, the resulting R_X is large (~50 Ω), illustrating the trade-off between bandwidth and R_X. In [98], a new high-precision (Av and A_i transfer errors of 30×10^{-6} and 10^{-6}, respectively) and ultra-low Rx (less than 0.1 Ω) CCII is proposed. The design, however, only achieves a bandwidth of 15 MHz in 0.5 um CMOS, demonstrating the trade-off that also exists between bandwidth and transfer precision. Other implementations of CCIIs not based on the aforementioned topologies have also been proposed, such as the translinear-loop based CCII in [99] with large bandwidths (2-3 GHz in 0.35 µm CMOS) and a low R_X below 20 Ω. This improvement comes at the price of a larger transfer error of 4%. Other recent CCII implementations based on a flipped voltage follower [100] and floating current source [101], [102] have also been proposed. In [100], low-power operation is achieved (less than 100 µW) and low total-harmonic-distortion (THD) (less than 0.21%) with a transfer error of less than 0.01% at the cost of bandwidth reduction (less than 60 MHz in 0.35 µm CMOS). In [101] a bandwidth of ~600 MHz (0.5 µm CMOS) with $R_X < 15$ Ω is presented, but only in simulation. P-type transistors are used, which limit the achievable bandwidth.

An analysis of the state-of-the-art in published literature, therefore, indicates that open-loop designs achieve higher bandwidths than designs with closed-loop feedback, but at the expense of poorer transfer precision and higher R_X values. On the other hand, closed-loop designs boast excellent precision and low R_X values but at the expense of lower bandwidths.

Besides having narrower operating bands, closed-loop CCIIs are also susceptible to instability (as will be shown in Section 5.3.4), especially at higher frequencies. Despite this risk, stability analysis in CCIIs has received minimal attention in the literature, as is further discussed in Section 2.5. Even though a general approach to the feedback analysis of low output impedance circuits (such as CCIIs) is presented in [103], often an analytical approach at high frequencies is too complex to attempt. Furthermore, analytical methods typically fail

to account for multiple loops which, if not explicitly present in the CCII schematic design (eg. the minimum of two loops in [94]), may exist through parasitic elements. Besides possible instability, feedback loops often also cause gain peaking at the upper frequency band-edge, with only the work in [96] addressing this problem through peaking reduction. The stability problem is often compounded by changing load impedance and environmental conditions during operation, as well as process variation. This would require post-production tunable phase margin and peaking control, which has not been addressed explicitly in the CCII literature.

Another significant shortcoming in the body of literature on CMOS CCIIs is limited published measurement data. In [93], [96], [94], [97]–[102], all results are based on simulations. Furthermore, few papers consider practicalities of device manufacturing such as non-ideal effects, process corners and random variation, device parasitics, and stability considerations. No paper describing a *physical implementation* (with measurement results) of a high bandwidth, high precision CMOS CCII has ever been published. This has possibly contributed to manufacturers' aversion to commercial adoption of CCIIs [92].

2.4.3   BiCMOS realizations

BiCMOS-based CCII circuits described in the literature can be subdivided into those based on translinear implementations [104]–[107] and those based on differential pair and current mirror implementations [91], [108]–[110]. Translinear designs generally exhibit lower R_X and higher R_Y values but at the cost of much lower bandwidths [91]. This is caused by the lower achievable peak f_T values of PNP bipolar transistors, which are integral to a translinear design, in contrast to the much higher values for NPN devices. For example, in the GlobalFoundaries 130 nm BiCMOS technology process the NPN transistors exhibit a peak f_T of 220 GHz, whereas the PNP transistors exhibit a peak f_T of only 18 GHz at optimal bias conditions. The translinear CCII presented in [111] forms the baseline of many subsequent designs. This design is improved in [104] achieving an $R_X = 1.3\ \Omega$, $R_Y = 70\ \text{k}\Omega$, $R_Z = 71\ \text{k}\Omega$ and a voltage and current following bandwidth of 1 GHz and 1.66 GHz respectively (in SPICE simulations for a 2 μm BiCMOS technology node with peak f_T values of 4 and 2 GHz,

respectively). Furthermore, the design achieves a voltage following error of 0.14% and a current following error of 2.3%. To improve A_i the design in [105] implements a Wilson current mirror achieving a current following error of 1.5% but at the cost of lower bandwidths. Moreover, R_X is degraded to 31.6 Ω. Similarly in [107] a cascode current mirror is used to improve voltage mirroring and current conveying precision, with similar results. Furthermore, the effect of a changing load is visible in A_i, indicating the importance of optimizing a design for a specific load impedance.

In an effort to increase bandwidths, designs based on the differential pair and current mirror have been proposed. The CCII in [108] achieves a voltage mirroring bandwidth of 4.5 GHz and a current mirroring bandwidth of 2 GHz in a 0.8 µm BiCMOS process. The reported R_X = 80 Ω, R_Y = 28 kΩ and R_Z = 28 kΩ illustrate the associated tradeoff in performance. Furthermore, the Av and A_i transfer errors in this design depend heavily on the load. This dependence can be reduced by implementing feedback control as shown in [91]. Once again, the design is prototyped in 0.8 µm BiCMOS this time achieving a voltage and current transfer bandwidth of 3.2 and 2.2 GHz, respectively, while improving R_X to 2.2 Ω. The values for R_Y and R_Z are 17 kΩ and 18 kΩ respectively, which is slightly lower than those reported in [108]. Lastly, in an effort to design a high-bandwidth CCII the solution in [109] proposes a simple two-transistor NPN current mirror biased by CMOS devices, achieving a voltage following bandwidth of 22 GHz and a current mirroring bandwidth of 35 GHz, in a simulation of a 0.25 µm BiCMOS process. However, the resulting R_X is 420 Ω clearly illustrating the tradeoff between obtaining a low R_X and a high bandwidth.

2.5   STABILITY ANALYSIS TECHNIQUES

In Section 2.4 it was shown that closed-loop feedback can be used to reduce CCII non-idealities, thereby making them acceptable for the purposes of implementing the second-order all-pass designs described in Section 2.3.2. It turns out however that closed-loop CCIIs are susceptible to instability (as will be shown in Section 5.3.4). An analytical stability analysis is often impractical as, firstly, foundries supply numerical models of the active and passive components and secondly, the analytical equations become too complicated to solve

by hand at high frequencies. In this Section, the SOTA techniques in single and multiple feedback loop analysis are presented and shortcomings identified.

2.5.1   Single-loop stability

Single-loop feedback theory is based on Bode's definition of the return ratio

$$T = -\frac{v_f}{v_e} = -\frac{i_f}{i_e} \,,$$ (2.7)

where v_e and i_e represent the injected and v_f and i_f the returned signals, as defined in Figure 2.4 [59], [112], [113].
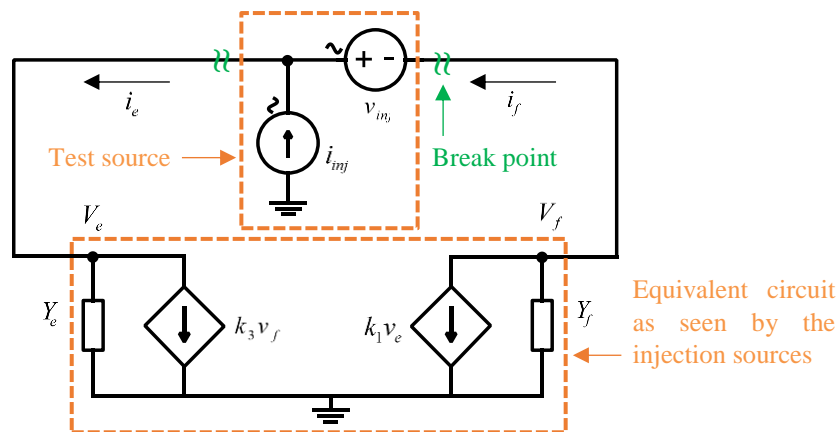

Figure 2.4. Single-loop feedback analysis using the double-injection technique.

Bode's original theory requires replacing an existing dependent source (that models the active device – typically current or voltage gain) with an independent test source, which is not always possible (especially if black-box models are used). This limitation is overcome by Middlebrook's and Tian's subsequent extensions to single-loop feedback theory [59]. Using Middlebrook's approach, the feedback loop can be "cut" at any point that breaks all possible feedback loops, and a test source inserted in the loop at the break, as shown in Figure 2.4. The test source consists of both a current and voltage source in order to maintain the original DC impedances at both break point terminals. Tian's approach further accounts

for reverse and forward loop transmission effects as modeled by the two dependent sources, k_1 and k_3 [59].

The return ratio can be derived from (2.7) and Figure 2.4 using Kirchhoff's laws, resulting in the equation [59]:

$$T = \frac{2(AD - BC) - A + D}{2(BC - AD) + A - D + 1},$$
(2.8)

where

$$A = \frac{-k_1 - Y_f}{k_1 + k_3 + Y_e + Y_f}, \quad B = \frac{Y_e Y_f - k_1 k_3}{k_1 + k_3 + Y_e + Y_f},$$

$$C = \frac{1}{k_1 + k_3 + Y_e + Y_f}, \quad D = \frac{k_3 + Y_f}{k_1 + k_3 + Y_e + Y_f}.$$
(2.9)

It is important to note that an intrinsic assumption has been made in the foregoing discussion – that the effective network seen by the test voltage and current sources can be modelled by the equivalent circuit shown in Figure 2.4. This assumption only holds if a break-point can be found which "breaks" all loops; more specifically, if a controlled source can be found which when nulled, renders the entire network passive. In the design of many microelectronic circuits, such as CCIIs, this assumption does not hold.

2.5.2   Multi-loop stability

The first extension of single-loop to multi-loop feedback theory was made in [114] where the concept of the return-difference matrix was introduced. Subsequent extensions of this theory [115] included the concept of the null return-difference matrix. However, these multi-loop analysis methods are based on analytical network techniques, which, as justified earlier, are often inadequate in the design of microelectronic circuits [115]–[117], where numerical techniques are required.

The approach presented in [60], [61], on the other hand, is sufficiently abstracted from any specific circuit implementation to support both analytical and numerical applications of the technique. The work in [60] presents a method of obtaining the denominator of the transfer function of a linear system represented as a signal flow graph. No restrictions as to the number or size of the loops is imposed. This approach is also compatible with numerical methods. It is shown that the denominator of the system's transfer function (which can be used to determine the system's stability by using a Nyquist plot for instance) is the sum of gain products of all possible combinations of non-touching loops (denoted by Δ). This statement is summarized by the equation:

$$\Delta = 1 - \sum_m P_{m1} + \sum_m P_{m2} - \sum_m P_{m3} + ...,$$ (2.10)

where P_{mr} is the gain product of the mth possible combination of r non-touching loops where $r > 0$. For example, considering the flow graph of Figure 2.5, four feedback loops can be identified:

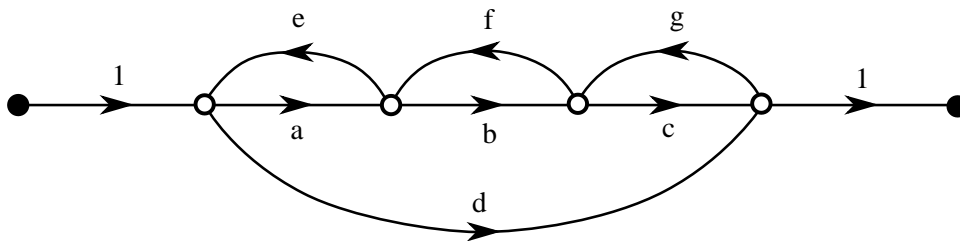$$\begin{aligned} T_1 &= ae & T_3 &= cg \\ T_2 &= bf & T_4 &= dgfe. \end{aligned}$$ (2.11)

Figure 2.5. A flow graph with four feedback loops [60].

The denominator of the transfer function can therefore be determined as:

$$1 - (T_1 + T_2 + T_3) + (T_1 T_3)$$ (2.12)

It is further shown in [60] that (2.10) can always be re-written in the factorized form:

$$\Delta = \prod_n 1 - T_n{'}, \qquad\qquad\qquad (2.13)$$

where $T_n{'}$ is the loop gain of the nth loop with all lower-numbered loops ($T_1 \dots T_{n-1}$) split. The result of (2.13) allows the application of Tian's approach to cut-based single-loop analysis in multi-loop feedback analysis, as will be shown in this work in Section 5.3.2.

2.6   ANALOGUE SIGNAL PROCESSING WITH ALL-PASS NETWORKS

In an effort to place this work in context, various potential ASP applications are summarized below. In each case the fundamental building block is an all-pass network of prescribed response, specific to the ASP application. The requirement that a Q_D-value larger than 1 is required for the second-order all-pass building blocks is justified in this Section. In literature $\Delta\tau$ and ω_0 are typically reported as opposed to the more universal parameter Q_D. Where this is the case (2.2) is used to obtain Q_D.

2.6.1   Real-time spectrum analysis

The Fourier transform and inverse Fourier transform have found application in many areas of science and engineering as two fundamental signal processing operations [11], [13]. Existing systems use DSP techniques to perform the fast-Fourier transform (FFT) in applications such as orthogonal frequency division multiplexing (OFDM) [11] and automotive RADAR, where the FFT is used to compute various physical parameters such as the car's speed and the distance from other objects [118]. Recently, several papers [11], [13], [15]–[17] have proposed to borrow techniques used in optics to perform a real-time Fourier transform (RTFT), using all-pass networks as shown in Figure 2.6 [13].

In the circuit of Figure 2.6 the input signal ($x(t)$) is multiplied by a linearly chirped signal $c(t) = \exp(jt^2/2C_1)$, where C_1 is the chirp parameter, producing signal $m(t)$ as shown in Figure 2.6. This signal is then injected into the all-pass network with an impulse response of $h(t) = \exp(jt^2/2C_2)$. The output of the all-pass network ($y(t)$) can be found as [13]:

$$y(t) = m(t) * h(t) = \exp\left(\frac{jt^2}{2C_2}\right) \int_{-\infty}^{+\infty} x(\tau) \exp\left(j\left(\frac{1}{2C_1} + \frac{1}{2C_2}\right)\tau^2\right) \exp\left(-j\frac{t\tau}{C_2}\right) d\tau \ . \ (2.14)$$
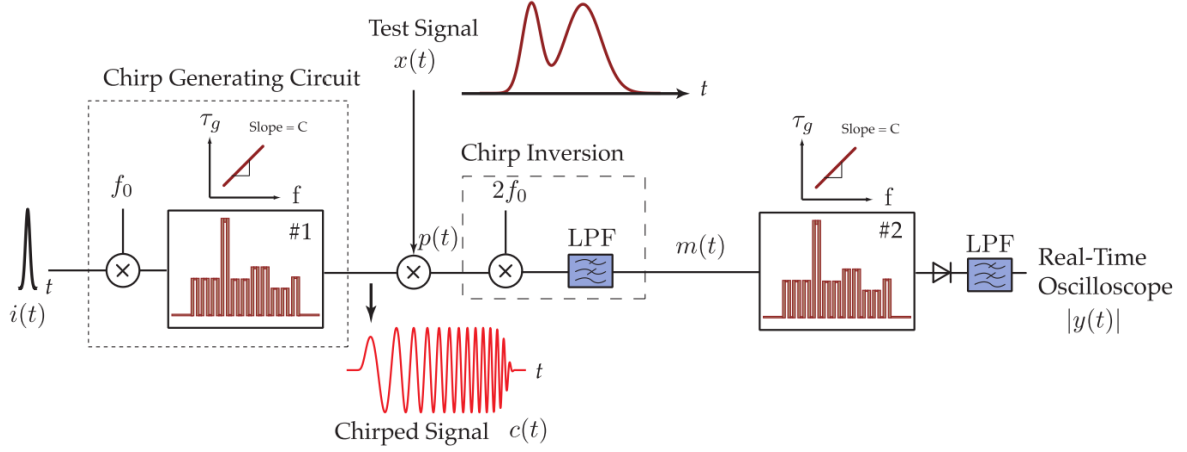

Figure 2.6. RTFT system [13], © 2010 IEEE.

If the condition $C_1 = C_2 = C$ is imposed, then it can be shown that:

$$|y(t)| = \int_{-\infty}^{+\infty} x(\tau) \exp(-j\omega\tau) \ d\tau , \qquad (2.15)$$

where $\omega = t/C$ [13]. Equation (2.15) therefore indicates that the magnitude of the signal at the output of the all-pass network is proportional to the Fourier transform of $x(t)$ with the frequency domain mapped onto the time axis. Increasing C increases the frequency discrimination in time. Prototypes of the system in Figure 2.6 have been reported in soft-substrate microstrip technology [11]. To ensure that the chirp signal contains frequency components which correspond to the all-pass network's passband, using (2.2) and the fact that the chirp signal must be sufficiently wide to cover at least one period of the lowest input frequency, it can be shown that $Q_D > \pi/2.(1 + BW/f_1)$, where BW represents the bandwidth of the all-pass network and f_1 is the lower frequency cutoff point. Therefore, a Q_D-value larger than at least $\pi/2$ is required [11]. Often all-pass sections with Q_D-values larger than 10 are used in soft-substrate designs. However, no on-chip implementations exist due to a lack of second-order all-pass networks with Q_D-values larger than 1 on-chip.

Using a similar approach, an inverse Fourier transform can be performed, where the output magnitude spectrum is identical to the input time signal [13].

2.6.2   Frequency sniffing for cognitive radio

Real-time spectrum sniffing is another promising application of all-pass dispersive networks with potential application in cognitive radio and automotive RADAR [21]. The principle of operation of real-time frequency sniffing is illustrated in Figure 2.7.

An input signal, with unknown frequency components, is passed through an all-pass network with a stepped delay response. Each band, centered at different frequencies corresponding to the locations of the group delay steps, is delayed by varying amounts. A simple Schmidt trigger can then be used to determine the availability of a given frequency band, provided that the introduced delay is sufficiently large to prevent overlapping. All-pass networks with Q_D-values larger than π are typically required to meet this requirement, as can be deduced in [21]. Existing solutions to this problem use digital FFT solutions which exhibit lower bandwidths and often require fast digital processors which, in many applications, are prohibitively expensive and power intensive. For mm-wave applications where contiguous bandwidths approaching 10 GHz are expected, digital techniques are no longer adequate for a comprehensive frequency analysis [21].
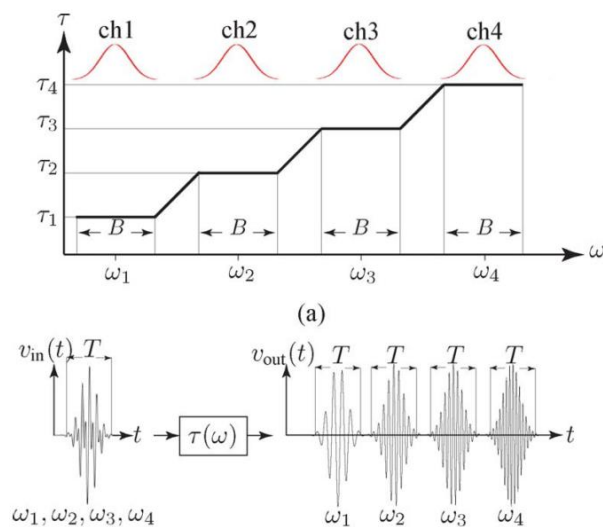

Figure 2.7.   Principle of operation of a real-time spectrum sniffer [21], © 2012 IEEE.

2.6.3   Arbitrary frequency beam scanning in antenna arrays

Another application of all-pass networks is in frequency scanning antenna arrays which are used in RADAR, imaging and spectrogram analysis [22], [23]. In the past, frequency independent (constant) delays have been used in antenna arrays to steer the beam, thus producing a certain time vs. spatial-domain radiation pattern. The concept of frequency scanning, achievable by using group delay networks with a non-constant delay function, allows frequency to beam angle mapping thus adding an additional orthogonal dimension, in effect improving achievable speeds and efficiency. Therefore, the spatial antenna radiation pattern of such a system is not only a function of time but also of frequency. A prototype of such a system has been demonstrated in soft-substrate using a microstrip all-pass network [23], achieving a linear frequency to beam angle mapping of 60°/GHz over a bandwidth of 1 GHz using all-pass sections with Q_D-values larger than 1 (in this case $Q_D = 2$ is extracted). A quadratic group delay function was required for this purpose. No on-chip implementations of such a system has ever been demonstrated, due to a lack of on-chip second-order all-pass networks with Q_D-value larger than 1.

2.6.4   Dispersion code multiple access (DCMA)

One way of further improving the throughput of future telecommunication systems is by introducing additional orthogonal multiple access schemes. One such candidate is dispersion code multiple access (DCMA) which is an analogue technique and is therefore suitable for next generation high-bandwidth systems [119]. The principle of operation of this scheme is illustrated in Figure 2.8 for a three-user communication system.

In this example, only users with opposite group delay receiver profiles can communicate successfully. User A expands a broadband signal which is then re-compressed by user B. User C, on the other hand, further expands the transmitted signal and therefore cannot reconstruct the original message [119]. It is shown in [119] that by using odd Chebyshev polynomial group delay functions, a theoretical bit-error rate (BER) of 0 is achievable.

However, a Q_D-value much larger than 1 is required (> 20). No measured results have yet been presented in literature in soft-substrate or on-chip.
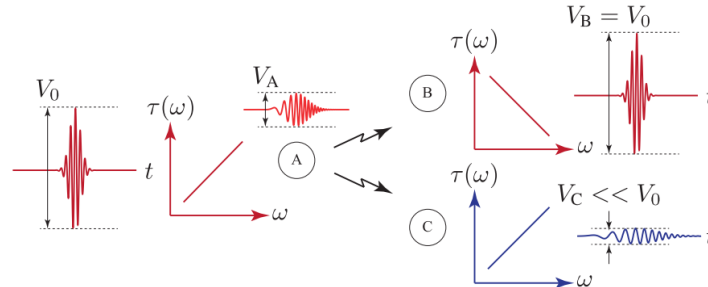

Figure 2.8. Principle of operation of DCMA [119], © 2015 IEEE.

2.6.5 M-ary pulse position modulation (PPM)

Another approach to achieving ultra-wideband telecommunication is by introducing M-ary pulse position modulation (PPM) [18]. A group delay function like that of Figure 2.7 can be used to delay pulses by varying amounts depending on the mixer frequency. A prototype is demonstrated in soft-substrate in [18]. For a pulse delay range comparable to the input pulse width a Q_D-value larger than 1 is required ($Q_D \geq \pi$ for a modulation delay of at least one pulse width). No on-chip implementations of such a system have ever been demonstrated, due to a lack of on-chip second-order all-pass networks with a sufficiently high Q_D-value.

2.7 CONCLUSION

The fundamental building block of any analogue signal processor is an all-pass network of prescribed response. Any all-pass network is shown to be synthesizable by cascading first- and second-order all-pass sections, with two cascaded first-order sections being a special case of the later (in the group delay sense). The state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements are reviewed and found to lack a theoretical synthesis procedure that guarantees minimum-order networks. Furthermore, literature on second-order all-pass networks, the building blocks of ASPs, are generally found to be limited to off-chip implementations, due to (among other reasons) high on-chip

losses. In response to this, numerous active inductorless implementations of all-pass networks are found in literature, but, no measured results of a second-order all-pass network on-chip with a Q_D-value larger than 1 (as required for ASP applications and illustrated in Section 2.6) have been presented. To address these research gaps, CCII-based all-pass realizations have been identified as a promising approach. However, it is found that CCII non-idealities, process tolerances and parasitics make many such designs proposed in literature impractical. Moreover, CCIIs have seen limited commercial adoption due to, among other reasons, high sensitivity to process tolerances, layout parasitics, and possible instability. In preparation to address this shortcoming the literature on single- and multi-loop stability analysis has also been reviewed and will be used in the novel all-pass network design of this work. Finally, a single transistor all-pass design has also been identified as a suitable approach to extend achievable bandwidths, as will be shown in a BiCMOS implementation later in this work.

CHAPTER 3   RESEARCH METHODOLOGY

3.1   INTRODUCTION

The research methods which are employed to answer the research questions are detailed in this Chapter. First a justification for the chosen paradigm is given and a research methodology is proposed. The methodology is detailed by further describing the simulation framework and software used in this work as well as the manufacturing processes employed for prototyping. Finally, the measurement setup and procedure are described.

3.2   JUSTIFICATION FOR THE PARADIGM AND METHODOLOGY

Analogue signal processing is a promising alternative to DSP techniques in future high-speed telecommunication and data-processing solutions, as analogue devices outperform their digital counterparts in terms of cost, power consumption, and the maximum attainable bandwidth. The fundamental building block of any ASP is an all-pass network of prescribed response. Implementing all-pass networks on-chip supports mass production at reduced costs and operating power requirements, which is a key requirement in future telecommunication and data-processing systems. Therefore, synthesizing a minimum-order network is particularly important for an on-chip implementation. Existing synthesis techniques suffer from convergence issues and no guarantee can be made that the obtained solution is optimal (minimum-order). Furthermore, to date, no measured results of a physical second-order all-pass network, with a Q_D-value larger than 1, have been presented on-chip in either CMOS or BiCMOS technology nodes.

In this work, a novel analytically-based quasi-arbitrary group delay function synthesis theory using all-pass elements is investigated and microelectronic implementations of second-order all-pass networks sought, in both CMOS and BiCMOS technology nodes, to address the

above research gap. Prototypes of all the proposed designs are fabricated and measured results presented. This will, enable future monolithic integration of ASPs on-chip, making the analogue approach a viable alternative to existing DSP techniques.

3.3   OUTLINE OF THE METHODOLOGY

The research methodology followed in this work is summarized in Figure 3.1 and can be grouped into the: (1) literature review, (2) mathematical network synthesis, (3) circuit design: process tolerance modeling, simulation and optimization, (4) layout design: parasitic extractions, (5) measurements and (6) evaluation of the research questions. These topics are described in more detail below.

1. Literature review: The first stage of the research methodology involves performing a detailed literature study to contextualize the work, establish existing techniques and solutions and identify shortcomings in the SOTA. This is done by firstly, reviewing existing all-pass network synthesis techniques, followed by a review of both their passive and active realizations. The CCII is found to be a versatile analogue building block of all-pass networks, in turn justifying a literature review of high-bandwidth CCIIs. Research gaps are then identified in both the mathematical synthesis and on-chip practical implementation of all-pass networks. Once the research paradigm is contextualized and gaps identified, the research questions are formulated and methodology established.

2. Mathematical network synthesis: The next step in the proposed methodology involves the mathematical synthesis of minimum-order all-pass networks. A solution to the synthesis problem is sought by means of an approximation-based numerical approach. An approximation to the second-order all-pass function is first found which can in turn be used to find an analytical solution to the initial locations of the all-pass function poles and zeros. These locations can then be numerically optimized guaranteeing a minimum-order solution.
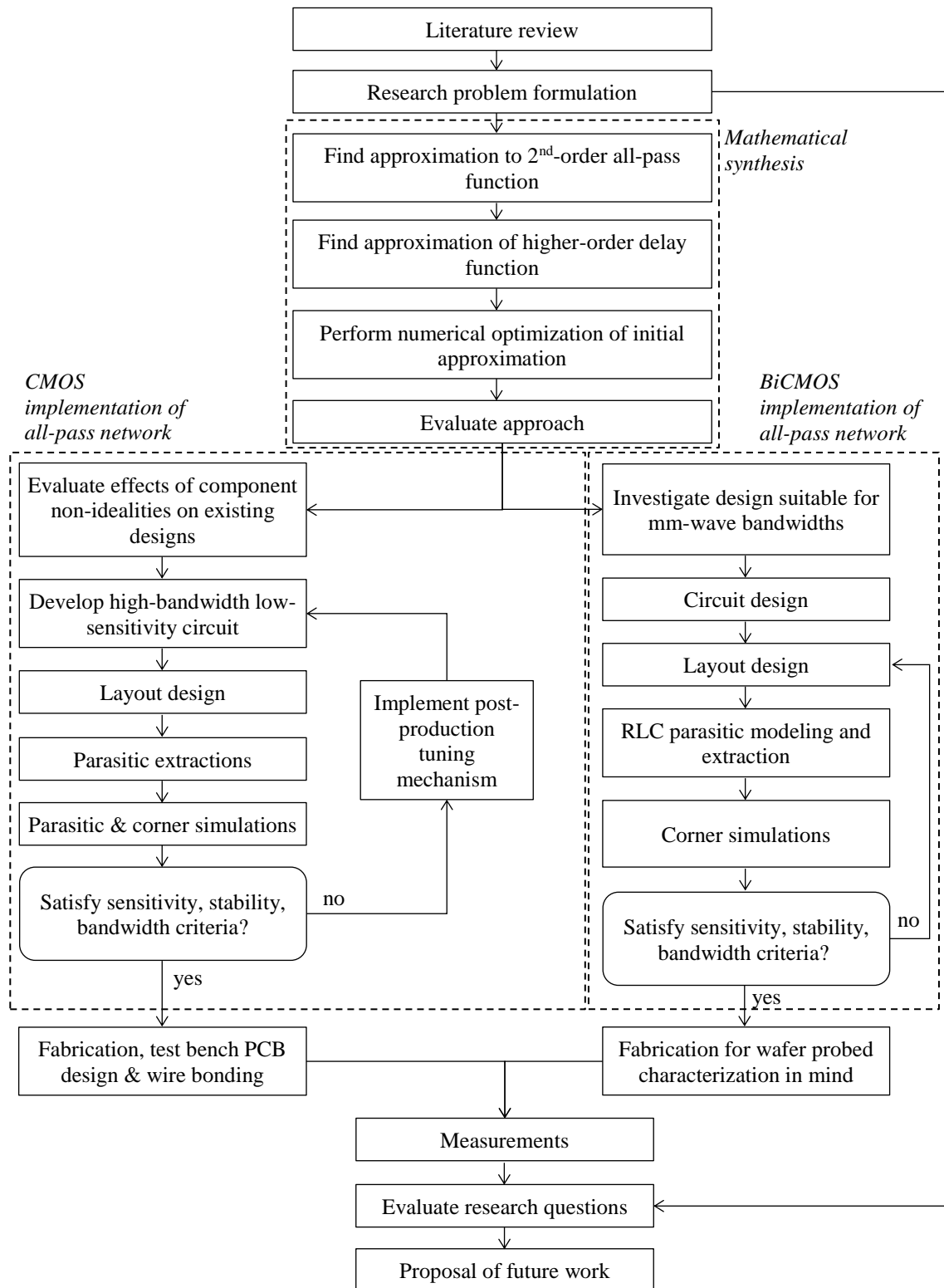
Figure 3.1. Proposed research methodology.

In order to illustrate the proposed synthesis theory and to compare it to SOTA techniques, examples of all-pass functions are synthesized and compared to existing methods.

3. Circuit design: Next, circuit designs of second-order all-pass networks, suitable for both the CMOS and BiCMOS processes, are investigated and proposed. This is done by investigating the effects that component non-idealities and process tolerances have on existing designs and proposing solutions which have reduced sensitivities to these factors. Initial circuit simulations are completed using detailed Berkeley Short-Channel IGFET (BSIM) models supplied by the foundries in the form of PDKs. These models include component non-idealities as well as intrinsic parasitic elements (as opposed to extrinsic layout parasitics) to provide the best estimate to the actual response as possible. The CCII is used as the active element in the CMOS all-pass network design to maximize bandwidth. For this reason, a novel CCII design methodology is also investigated, for the first time incorporating multi-loop feedback stability analysis into the published circuit design. To further increase bandwidths a single transistor BiCMOS design is developed which takes into account layout considerations (non-zero length elements) and component non-idealities. Throughout the design process, measurement considerations are also taken into account for compatibility with a practical test bench setup.

4. Layout design: Next, IC layouts of the schematic-level designs are proposed. Since this research is focused at practical implications of all-pass networks on-chip, this step is particularly crucial. The IC layouts are developed by considering the circuits' dependence on layout effects such as process tolerances, parasitic elements and distributed effects. The effects of process tolerances on key performance metrics are established by performing corner analyses. Since large component-level tolerances are expected in the CMOS process (up to 30% absolute variation [88]), methods of compensating for this dependence, such as post-production tunability, are then investigated. This justifies a corner analysis, as opposed to a Monte Carlo analysis, allowing the extremes and therefore the required tuning ranges to be established. On

the other hand, in the BiCMOS design, smaller component-level variations are expected and therefore the need for a post-production tuning mechanism is not anticipated. Even though a Monte Carlo analysis would be useful, only a corner simulation is viable due to limited PDK data. Parasitic effects are further considered based on layout geometry, using process design kits (PDKs) offered by the various foundries. This step is particularly important for a high-frequency design such as the proposed mm-wave network in BiCMOS. Lastly, distributed effects are also considered by including parameterized S-parameter models in the design for certain layout elements larger than 1/10 of the smallest wavelength of interest.

5. Measurements: After the designs are completed and manufactured, measurements are performed to validate the proposed design flow. A testbench setup is proposed suitable for the desired measurements using available laboratory equipment. Operating frequency ranges are further used to determine a suitable testbench setup and signal excitation path (SMA to wire-bond interface or on-chip probing). This stage also involves the design of test bench PCBs to which the IC is wirebonded. A formal test-plan is then developed to ensure reproducibility of results.

6. Evaluation of the research questions: The measurements of step 5 are used to verify the theoretical simulations and therefore to address the research questions.

3.4   SIMULATION SOFTWARE

The IC6 Cadence Virtuoso software suite is the industry standard in IC design and is used in this research for schematic and layout design, as well as parasitic extractions, process variation and MC simulations. The components of the Cadence Virtuoso software suite, as well as other software used in this work, are detailed in Table 3.1.

Firstly, MATLAB is used to implement the proposed group delay network synthesis theory. To this end, a MATLAB script is written which accepts as input parameters the numerical description of the desired group delay function as well as the maximum allowed minimax

error of approximation. The script then computes the locations of the all-pass poles and zeros as well as the required order of the network. Next, the SLiCAP MATLAB tool is used in the design of second-order all-pass circuits by performing an initial theoretical symbolic analysis of the proposed designs. Such an analysis allows key circuit parameters to be identified and their effect on the circuit's performance quantified. Once the ideal operation of the circuit has been established and meets requirements, the circuit schematic is captured with the Virtuoso Schematic Editor tool, where ideal components are replaced with available on-chip equivalent models, supplied by the PDK. Detailed BSIM model simulations are then performed using the Virtuoso MMSIM package. As per the proposed research methodology of Section 3.3, the Virtuoso Layout Editor is subsequently employed to design the physical layout described by the circuit schematic. The Cadence Assura tool is used to perform a DRC check of the layout to ensure manufacturability. This tool also identifies common electrical design errors, such as floating CMOS gates, as well as lack of adherence to latch-up rules. After the DRC check errors are corrected an LVS check is performed using the same tool. This check ensures that the original schematic is indeed an accurate representation of the layout design.

Table 3.1. Simulation software packages used in this work.

Package name	Description
Virtuoso Schematic Editor	Circuit schematic entry tool.
Virtuoso Layout Editor	Circuit IC layout tool.
Virtuoso Multi Mode Simulation (MMSIM)	The Spectre circuit simulator is used which provides SPICE-level analog and RF simulation for microelectronic designs.
Cadence Assura	Provides design rule check (DRC), layout versus schematic (LVS) as well as pattern density check functionality.
Quantus QRC	Provides RLC parasitic extraction functionality based on the layout geometry.
Matrix Laboratory (MATLAB)	Used for mathematical computations, optimizations and figure plotting.
SLiCAP	Symbolic linear circuit analysis program which is used for the symbolic simulation of circuits.

After the LVS check has been successfully run the Quantus QRC tool is used to extract parasitic RC components. This tool generates a netlist with the parasitic elements included which can then be simulated with the Virtuoso MMSIM tool. Using a combination of the

aforementioned tools, an optimization can be performed and the final camera-ready artwork generated in the GDSII file format, which is the de facto industry standard for IC layout artwork.

3.5   MANUFACTURING PROCESS

Two semiconductor technology processes are used for prototyping purposes in this work: the 0.35 µm C35B4C3 CMOS process from austriamicrosystems (ams AG) as well as the 0.13 µm SiGe BiCMOS 8HP process from GlobalFoundries. These processes are chosen due to their availability as part of funded prototyping runs.

In the 0.35 µm C35B4 CMOS process NMOS and PMOS transistors are available with f_T of 28 GHz and 15 GHz, respectively. Four front-end-of-line metallization layers can be used for signal routing. Table 3.2 summarizes the technology features.

Table 3.2. Summary of relevant 0.35 µm C35B4 CMOS technology features [120].

Device	Features
NMOS transistor	$f_T <= 28$ GHz, $V_{GS(MAX)} = 3.6$ V, $V_{TN0} = 0.46$ V, $K_{PN} = 170$ $\mu A/V^2$
PMOS transistor	$f_T <= 15$ GHz, $V_{GS(MAX)} = 3.6$ V, $V_{TP0} = -0.58$ V, $K_{PP} = 58$ $\mu A/V^2$
MOS varactor	Q-factor > 43, $C \in (1.33{:}4.88)$ fF/μm^2
Metal-insulator-metal (MIM) capacitor	$C = 1.25$ fF/μm^2, $Q > 100$
POLY1-POLY2 capacitor	$C = 0.86$ fF/μm^2
Poly resistors	$R \in (42{:}52)$ $\Omega/\square$ (ohms per square)

The PDK supplied by ams AG for this technology node uses BSIM3v3 models to characterize the devices in Table 3.2 up to a frequency of 6 GHz. The models are physics-based and scalable, that is, model parameters are dependent on the chosen device geometry, such as for instance the width and the length of a transistor. Detailed model parameters are not presented in this document as they are protected by a non-disclosure agreement (NDA).

The SiGe BiCMOS 8HP process offers high performance NPN SiGe heterojunction bipolar transistors (HBTs) with an f_T approaching 240 GHz. N- and P-type field-effect transistors

(FETs) are also available with f_T approaching 100 GHz and 45 GHz, respectively. Four front-end-of-line metallization layers, a thicker MQ layer and two thick RF-wiring layers are further available. Table 3.3 summarizes some of the technology features.

The PDK supplied by GlobalFoundries for this technology process characterizes the devices up to 40 GHz, with an extrapolation made to higher frequencies. The models are physics-based and are scalable. Once again detailed model parameters are not presented in this document as they are protected by a non-disclosure agreement (NDA).

Table 3.3. Summary of relevant 0.18 µm SiGe BiCMOS 8HP technology features [121], [122].

Device	Features
NPN HBT transistor	$f_T <= 240$ GHz, $h_{fe} > 100$ A/A, $BV_{CEO} \approx 1.5$ V
NS resistor	$R \approx 8$ $\Omega/\square$
N+ resistor	$R \approx 72$ $\Omega/\square$
P+Poly resistor	$R \approx 270$ $\Omega/\square$
KQ back-end-of-line (BEOL) resistor	$R \approx 142$ $\Omega/\square$
NMOS varactor	Q-factor > 2 (f > 40 GHz)
MIM capacitor	$C \approx 1$ fF/µm^2, Q-factor > 20
RFline inductor	$Q > 20$ at resonance

3.6   MEASUREMENT SETUP

A detailed test bench measurement setup used in this work for both the CMOS and BiCMOS designs is described below. This ensures transparency and reproducibility of results. In each case, a list of the equipment used followed by a detailed test procedure is provided.

3.6.1   CMOS second-order all-pass network

The CMOS-based second-order all-pass network is measured with a VNA by means of a test PCB SMA interface to which the IC is wirebonded, as detailed below.

3.6.1.1   Equipment

The following list details the equipment used to measure the response of the CMOS second-order all-pass network.

1. 1 x Anritsu ME7828A vector network analyzer (VNA).

2. 2 x R&S HAMEG HMP4040 power supply unit (PSU) or equivalent.

3. 6 x DC PSU power cables (banana connector).

4. 2 x bias-tee (built into VNA).

5. AutoCal connector-level calibration unit.

6. Analog Devices AD5370 DAC card.

7. 500 MHz bandwidth digital oscilloscope.

8. PC with MATLAB used for the genetic optimization.

3.6.1.2   Test procedure

The measurement procedure is described in detail as follows.

1. A test bench PCB is designed to house the IC and the IC is wire-bonded to the PCB as shown in Figure 5.9. The IC pads corresponding to the schematic of Figure 5.5 are shown in the camera-ready layout of Figure A.1.

2. Standard laboratory test conditions are ensured. The equipment is switched on and the experiment paused until the temperature of the equipment reaches a steady-state condition.

3. RF calibration:

    a. VNA settings: the input power is set to -25 dBm. The sweep frequency range is set to 1-1000 MHz, 1001 points, with a default IF bandwidth (1 kHz) (for speedy measurements).

    b. An AutoCal connector-level calibration of the VNA is performed. De-embedding of the wirebonds is not necessary as justified later in this work.

4. Measurement procedure:

    a. The measurement is setup as shown in Figure 5.10. The PSU supplies ± 2 V to V_{DD} and V_{SS} of the device under test (DUT). An input bias-tee is used and

biased at 0 V. Furthermore, the AD5370 DAC card is biased appropriately and connected to the PC running a MATLAB genetic local optimizer as described in more detail in Section 5.2.4. The VNA is also connected to the PC via USB to complete the optimization control loop.

b. A control signal is sent to the DAC card to bias the control voltages at the nominal levels as described in Section 5.2.3.2. The genetic optimization is then performed as described in Section 5.2.4. During the optimization procedure S-parameters are exported for online analysis.

c. Once the local optimum is found the two-port S-parameters are exported.

d. The digital oscilloscope is used to measure both input and output time-domain waveforms for chosen frequencies as shown in later Chapters.

e. After measurements are performed the PSU voltages are disabled and the measurement setup disassembled.

3.6.2   CMOS CCII+

The CMOS-based CCII+ is measured with a VNA by means of a test PCB SMA interface to which the IC is wirebonded, as detailed below.

3.6.2.1 Equipment

The following list details the equipment used to measure the response of the CMOS second-order all-pass network.

1. 1 x Anritsu ME7828A vector network analyzer (VNA).

2. 2 x R&S HAMEG HMP4040 power supply unit (PSU) or equivalent.

3. 5 x DC PSU power cables (banana connector).

4. 2 x bias-tee (built into VNA).

5. AutoCal connector-level calibration unit.

6. 500 MHz bandwidth digital oscilloscope.

3.6.2.2   Test procedure

The measurement procedure is described in detail as follows.

1. A test bench PCB is designed to house the IC and the IC is wire-bonded to the PCB as shown in Figure 5.20. The IC pads corresponding to the schematic of Figure 5.14 are shown in the micrograph of Figure 5.19.

2. Standard laboratory test conditions are ensured. The equipment is switched on and the experiment paused until the temperature of the equipment reaches a steady-state condition.

3. RF calibration:

   a. VNA settings: the input power is set to -25 dBm. The sweep frequency range is set to 1-1500 MHz, 1001 points, with a default IF bandwidth (1 kHz) (for speedy measurements).

   b. An AutoCal connector-level calibration of the VNA is performed. De-embedding of the wirebonds is not necessary as justified later in this work.

4. Measurement procedure:

   a. DC bias voltages are set as shown in Figure 5.14. The input bias-tee is biased at 0 V.

   b. Bias voltage V_{B2} is swept from 0.3 V to 0.75 V and the voltage transfer response plotted.

   c. Bias voltage V_{B2} is set to 0.35 V (which is found to minimize magnitude response peaking) and time-domain measurements performed of the input and output voltage waveforms, using the digital oscilloscope. The VNA is setup to generate the input waveforms at 100 MHz and 490 MHz.

After measurements are performed the PSU voltages are disabled and the measurement setup disassembled.

3.6.3   BiCMOS second-order all-pass network

The BiCMOS-based second-order all-pass network is measured with a VNA by means of a wafer probe station, as detailed below.

3.6.3.1   Equipment

The following list details the equipment used to measure the response of the mm-wave BiCMOS second-order all-pass network.

1. 1 x Anritsu ME7828A VNA.

2. 2 x GGB Picoprobe Model 110H ground-signal-ground (GSG) probes.

3. 1 x R&S HAMEG HMP4040 PSU.

4. 2 x DC PSU power cables (banana connector).

5. 1 x DC needles.

6. 1 x DC needle positioners.

7. Bias-tee (built into VNA).

8. ISS ceramic calibration substrate.

9. Power meter for power calibration.

3.6.3.2   Test procedure

The measurement procedure is described in detail as follows.

1. Probe planarization: the three contacts of the probes are set to the same height and the planarity of the probes is adjusted until they make three equal marks on the contact substrate.

2. Standard laboratory test conditions are ensured. The equipment is switched on and the experiment paused until the temperature of the equipment reaches a steady-state condition.

3. Power calibration: power calibration is performed by applying the VNA in a one-port THRU measurement and measuring the output power level with a power sensor. The calibration is then performed offline in post-processing.

    a. Step 1: The calibration measurement setup as shown in Figure 3.2 is used.
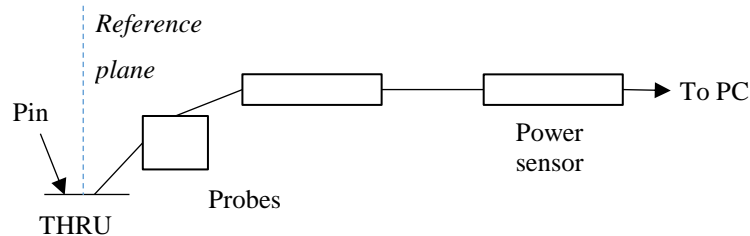
Figure 3.2. GSG VNA Power calibration setup.

    b. A power calibration sweep from -30 dBm to 0 dBm is performed for the following frequencies: 1 GHz, 10 GHz, 20 GHz, 30 GHz, using 1001 points, with a default IF bandwidth (1 kHz).

4. Compression measurement:

    a. The same VNA settings are used as above.

    b. The PSU is connected for operation in dual-rail mode. The voltages are set as shown in Figure A.7. The current limit for V_{DD} and V_{SS} is set to 6 mA and the voltage output disabled.

    c. The PSU is connected to the input bias-tee of the VNA and the bias set to 0.92 V. The voltage output is disabled.

    d. The DC needles are connected to the DUT and then connected to the PSU.

    e. The GSG wafer probes are applied to the DUT.

    f. The PSU output voltage is enabled.

    g. The V_{DD} bias voltage is adjusted to achieve a bias current of 4.6 mA (increasing the bias voltage increases bias current).

    h. A power sweep is performed from -30 dBm to -5 dBm (or until the 1 dB point on the compression plot is visible) for the following frequencies: 1 GHz, 10

GHz, 20 GHz, 30 GHz. For each frequency the appropriate power calibration from step 3 is applied in post-processing.

    i.   The 1 dB compression point for each frequency is determined.

5. RF calibration:

    a.   VNA settings: the input power is set much lower than the 1 dB compression point from step 4. The sweep frequency range is set to 1-50 GHz, 1001 points, with a default IF bandwidth (1 kHz) (for speedy measurements).

    b.   A short-open-load-through (SOLT) calibration is performed using the calibration substrate provided with the VNA. No further de-embedding is performed as justified later in this work.

6. Measurement:

    a.   The same VNA settings as above are used.

    b.   The PSU is connected for operation in dual-rail mode (by reversing the banana connectors – grounds are connected between stages). Bias voltages are set as shown in Figure A.7 for the all-pass measurement. The current limit for V_{DD} is set to 6 mA. The voltage output is disabled.

    c.   The PSU is connected to the input bias-tee of the VNA and the bias voltage set to 0.92 V as in Figure A.7. The voltage output is disabled.

    d.   The DC needles are connected to the DUT and to the PSU.

    e.   The wafer probes are applied to the DUT.

    f.   The PSU voltage output is enabled.

    g.   The bias voltage is fine-tuned to achieve a V_{DD} bias current of 4.6 mA (increasing the bias voltage increases bias current).

    h.   S-parameter measurements are performed.

    i.   After measurements are performed the PSU voltages are disabled and the measurement setup disassembled.

3.7   CONCLUSION

The research methodology has been presented and research paradigm contextualized and justified. Analog signal processing is identified as a promising alternative to DSP techniques and research gaps identified in both existing all-pass network synthesis techniques as well as microelectronic implementations thereof. Based on this a research methodology is then presented to address the gaps. Details of the simulation software used in the design process are further presented and the manufacturing process used to prototype the proposed solutions described. Finally, the measurement setup that is used to verify simulations is presented.

CHAPTER 4   MATHEMATICAL SYNTHESIS

4.1   INTRODUCTION

In this Section, a new numerical synthesis procedure of minimum-order cascaded all-pass networks having a quasi-arbitrary group delay response is presented, making the following contributions to the state-of-the-art (which was detailed in Section 2.2):

1. The proposed method does not require an initial value set. Convergence is therefore not dependent on an appropriate selection.

2. The resulting group delay function is an approximation of the required response to within any arbitrarily specified maximum delay variation across the passband, extending on methods in literature [27], [28], [31]–[33], [36], [58] where the network's order is chosen and not, explicitly, the resulting maximum delay variation.

3. The method is implementation abstracted, in that any theoretical, circuit, or parametrized numerical model description of a second-order all-pass section can be applied in the algorithm.

4. Due to the underlying analytical nature of the approach, rapid convergence is achieved; typically, within 20 iterations for practical cases (as will be shown in the examples). This is an order of magnitude fewer iterations required by more general state-of-the-art optimization methods (genetic algorithm and simulated annealing), as is shown in comparative examples.

To maintain generality, this work only considers second-order all-pass sections with complex poles and zeros, of which a special case is the first-order network centered around zero frequency.

This Chapter is organized as follows:

First, the problem of synthesizing a quasi-arbitrary group delay function is formulated geometrically on the complex s-plane and a derivation of the algorithm is presented. The proposed method is then demonstrated by synthesizing a linear group delay function with a sixth-order theoretical all-pass network. Next, in order to demonstrate the flexibility of the algorithm, Gaussian and higher-order delay functions are synthesized, followed by an example synthesis of a network to reduce the variation in group delay response of a physically measured system (in this case, a fifth-order hairpin resonator bandpass filter (BPF)). Lastly, the proposed method is compared to existing techniques in the literature and the improvements demonstrated. The synthesized all-pass networks are implemented in lumped-element circuit [62] form. In all cases, National Instruments AWR Microwave Office 10 is used as the circuit simulator.

4.2   THEORY OF GROUP DELAY SYNTHESIS

4.2.1   Approach

The group delay (as a function of frequency) of an ideal second-order all-pass network may be expressed in terms of the quadrature all-pass pole/zero pair locations, by the relationship [27]:

$$\tau_{\text{quad,x}}\left(\omega\right) = 2\sigma_x \cdot \left[\frac{1}{\left(\omega + \omega_x\right)^2 + \sigma_x{}^2} + \frac{1}{\left(\omega - \omega_x\right)^2 + \sigma_x{}^2} \right], \qquad (4.1)$$

where σ_x and ω_x represent the values of the $\mathbb{R}$ and $\mathfrak{I}$ components of the poles and zeros respectively, on the $\mathbb{C}$ s-plane. A quasi-arbitrary group delay function can be synthesized indirectly by cascading all-pass sections as described by (4.1) with a cost function τ_c (chosen to have the negative of the required delay curve) to obtain a cascaded response referred to as an error function τ_e. Then, by minimizing the variation of τ_e, the resulting all-pass network's delay τ_s approximates the required response *to within a constant*. Reducing the

variation of an arbitrary continuous group delay function τ_e by means of cascaded all-pass sections can be expressed mathematically as:

$$\tau_e = \min\left\{\Delta_{\max}\left\{\tau_c + \underbrace{\sum_{x=1}^{N}\tau_{\text{quad,x}}(\omega)}_{\tau_s}\right\}\right\},\qquad(4.2)$$

where the order of the all-pass network is $2N$, $\Delta_{\max}$ represents the maximum variation in the passband of interest and min is a numerical minimization algorithm. The maximum (rather than average) variation is minimized to approximate the desired function to within an arbitrary minimax error and therefore to obtain an equi-ripple approximation, which is an optimal result [36]. This is further illustrated in Figure 4.1. The delay cost function τ_c can either represent the negative of the desired function (by reducing the variation $\Delta_{\max}$, an approximation to the desired function is then synthesized) or a practical system that is to be equalized.
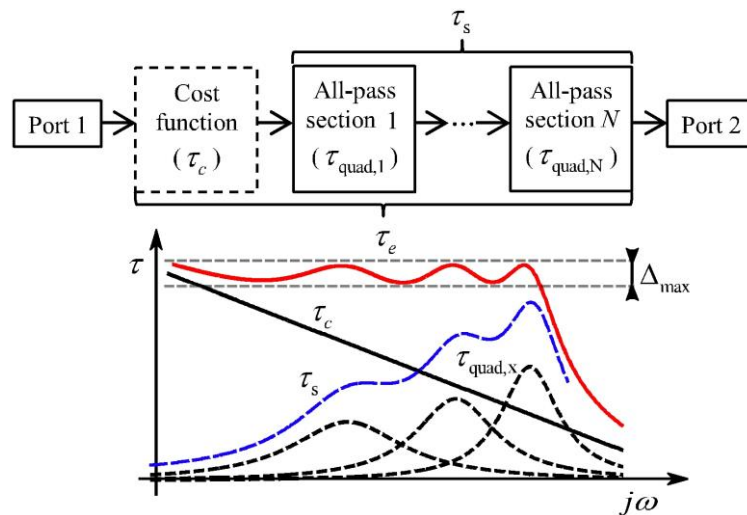
Figure 4.1. All-pass network cascaded with the cost function.

As per the introductory discussion, a theoretical basis is desired for finding the N initial solution sets (σ_x, ω_x), which ensure that (4.2) converges to a global optimum.

The approach in this work simplifies the complexity of the minimization problem described in (4.2) by using a novel geometrical approximation to the ideal all-pass network of (4.1), subsequently leading to an approximate analytical solution to (4.2) in the form of initial all-pass pole/zero solution intervals. These intervals are found in such a way that (4.2) is necessarily monotonic over each interval and, as a result, any gradient-based optimization will always converge to a solution. The end result is either an equi-ripple solution (if at all possible for the given delay function and desired maximum variation) or a solution with the fewest local minima and maxima, both of which are optimal results [36].

After an initial solution is found using the approximation to (4.1), it is replaced with the more accurate all-pass network descriptions of (4.1) and subsequently numerically optimized in the knowledge that, by the preceding step, a global optimum solution will be found. This results in an all-pass network composed of ideal all-pass quadrature pairs. The ideal all-pass quadrature pairs are then individually implemented in the passband of interest by using theoretical, circuit-simulated or otherwise parameterized numerical descriptions of practical all-pass networks [62]. The individual second-order circuit blocks are then optimized to achieve 1:1 equivalence with the theoretical all-pass section they are to represent in the circuit, and then cascaded with the original network without further optimization.

4.2.2   Theoretical derivation

First, a simplification to (4.1) will be found, as mentioned above. Equation (4.1) may be simplified by assuming that $\omega_x \gg BW/2$ (where BW is the bandwidth of interest) and by removing the negative frequency pole/zero pair, without appreciably affecting the group delay curve at positive frequencies. Next, using the additive property of differentiation and the symmetry of the pole/zero pair about the $\Im$ $j\omega$-axis, it can be shown that the group delay of a single pole and a pole/zero pair is equivalent, provided that:

$$p_x = \left\langle -\sigma_{px}, \omega_{px} \right\rangle \rightarrow \begin{cases} p_x = \left\langle -2\sigma_{px}, \omega_{px} \right\rangle \\ z_x = \left\langle +2\sigma_{px}, \omega_{px} \right\rangle \end{cases}, \tag{4.3}$$

where p_x and z_x are the locations of the xth pole and zero pair, respectively. An approximation to (4.1) can therefore be found by approximating the group delay of a single pole in the II quadrant of the s-plane, as described by Theorem 1 below. The derivation is presented in 4.6.

Theorem 1. *Assume that a single pole exists in the II quadrant of the complex s-plane at point P_1. Further assume that τ is the group delay caused by the pole at P_1 at a frequency $\omega = P_2$, where P_2 is any point on the positive $\Im\, j\omega$-axis of the s-plane. Then:*

$$\tau \approx \frac{1}{\overline{P_1 P_2}} \,, \tag{4.4}$$

where the error of the approximation is:

$$\lim_{P_2 \to \mathrm{Im}\{P_1\}} \left[\tau - \frac{1}{\overline{P_1 P_2}} \right] = 0 \,. \tag{4.5}$$

Theorem 1 is derived in Section 4.6.1 and is demonstrated graphically in Figure 4.2 where the error of approximation approaches zero as τ approaches its maximum. This relationship forms the basis of the synthesis algorithm presented in this Chapter.

Having derived an approximation to (4.1) the problem of (4.2) can be solved. To realize a final cascaded system (τ_e) (Figure 4.1) with equi-ripple variation in group delay, appropriate selections for $\tau_{\mathrm{quad,x}}$ must be made. Since τ_c is assumed to be a piecewise smooth function, the introduction of $\tau_{\mathrm{quad,x}}$ to satisfy (4.2) necessarily implies the introduction of new local minima and maxima to τ_e. Furthermore, by the equi-ripple requirement, a local minimum must be followed by a local maximum. The minimization problem can therefore be restated as finding optimal locations for new local minima in τ_e, while simultaneously controlling the peak locations and values of resulting subsequent local maxima, such that the desired maximum p-p delay variation of τ_e is obtained.
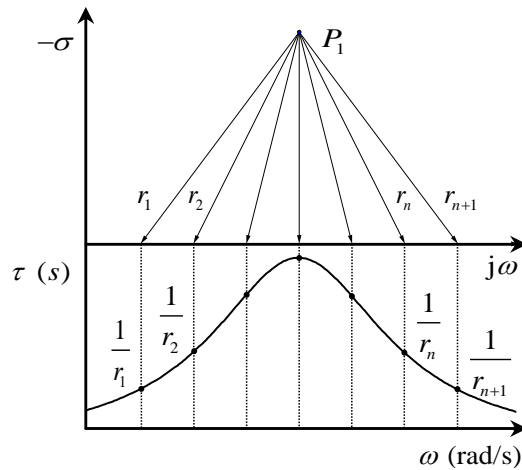
Figure 4.2. Graphical illustration of Theorem 1.

An algorithm is developed for introducing such local minima and maxima with second-order all-pass networks, whereby a sequence of iterations successively introduces new poles (p_x) in the II quadrant of the s-plane, until the desired peak group delay variation in (4.2) is obtained. These poles are then later replaced by quadrature all-pass pairs as per the transformation of (4.3), to preserve the initial system's magnitude response.

To introduce a new local maximum or minimum at a specific frequency of the group delay function τ_e, a pole *must* be placed on a semi-circular curve on the s-plane, as described by the following Theorem which is derived in Section 4.6.

Theorem 2. *Assume that $\tau_e(\omega)$ represents the group delay of some system of interest. Suppose that a local maximum or minimum is required at $\omega = \omega_{mx}$. Then a new pole p_x must be placed in the s-plane anywhere on the curve $\overrightarrow{r_{px}(\beta_x)}$ where β_x is the angle with the +jω-axis and*

$$r_{px} = \sqrt{\frac{\cos(\beta_x)}{c_1}}, \qquad (4.6)$$

where:

$$c_1 = -\frac{\partial}{\partial \omega}\left[\tau_e\left(\omega\right)\right]\Bigg|_{\omega=\omega_{mx}} \tag{4.7}$$

The location of p_x, as expressed in terms of rectangular coordinates, in a right hand coordinate system is equivalently:

$$\overrightarrow{r_{px}\left(\beta_x\right)} = \left\langle\omega_{px}, -\sigma_{px}\right\rangle = \left\langle\cos\left(\beta_x\right).r_{px} + \omega_{mx}, -\sin\left(\beta_x\right).r_{px}\right\rangle.$$

Theorem 2 is derived in Section 4.6.2 and illustrated in Figure 4.3.
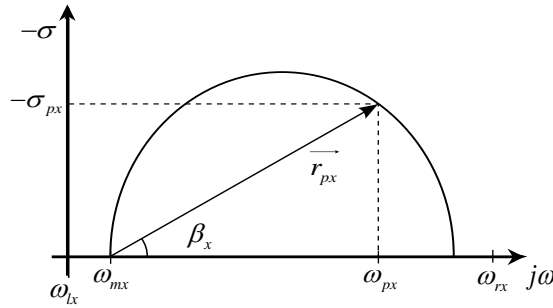
Figure 4.3. Graphical illustration of Theorem 2 on the s-plane.

As β_x is traversed from $\pi/2$ to 0, the maximum peak delay value introduced by p_x first decreases (since the perpendicular distance from p_x to the Im $j\omega$-axis increases) and then increases again (after the apex of the semi-circle is traversed), as per Theorem 1. The introduced delay value therefore begins and ends at ∞. When $\beta_x > \beta_{cx}$, ω_{mx} is a local maximum and as β_x decreases ω_{mx} transforms to a local minimum exactly once (see Section 4.6 for derivation), where β_{cx} can be found by:

$$c_1 = \frac{\left[3\cos^2\left(\beta_{cx}\right)-1\right]\left[c_1\right]^{3/2}}{\left[\cos\left(\beta_{cx}\right)\right]^{3/2}}. \tag{4.8}$$

The foregoing discussion is further illustrated with Figure 4.4.
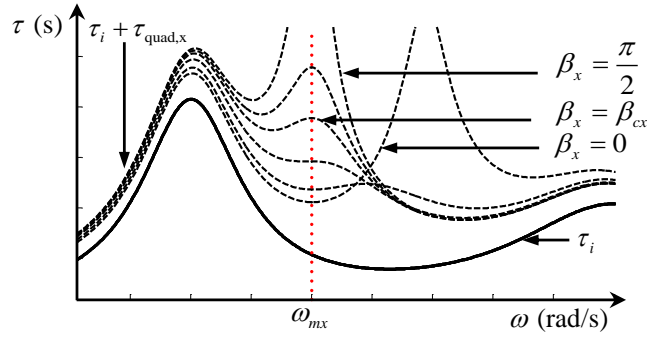
Figure 4.4. Graphical illustration of Theorem 2 in the frequency-time domain.

The next step involves selecting one pole from the infinite set of allowed values for p_x, with $\beta_x < \beta_{cx}$, such that the local maxima preceding (at $\omega = \omega_{lx}$) and succeeding (at $\omega = \omega_{rx}$) the newly introduced local minimum are both equal in magnitude and not greater from the local minimum than by the maximum allowed variation value, Δ_{max}, as further illustrated in Figure 4.5:

$$\tau_e\left(\omega_{lx}\right) - \tau_e\left(\omega_{mx}\right) = \Delta_{max}, \tag{4.9}$$

$$\tau_e\left(\omega_{rx}\right) - \tau_e\left(\omega_{mx}\right) = \Delta_{max}. \tag{4.10}$$
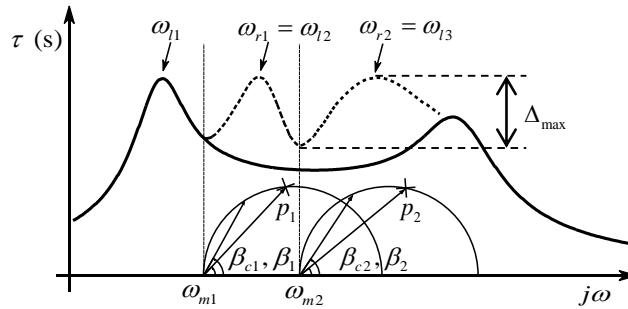
Figure 4.5. Graphical illustration of the group delay synthesis algorithm.

Before solving for ω_{mx} and β_x the behavior of the group delay at any point other than at ω_{mx}, as β_x changes in Figure 4.3, is found to be:

$$\tau(\omega) = \tau_e(\omega) + \left\{\left(\omega - \omega_{mx} - \cos\left(\beta_x\right)r_{px}\right)^2 + \left(\sin\left(\beta_x\right)r_{px}\right)^2\right\}^{-\frac{1}{2}}. \tag{4.11}$$

The solution to (4.9) and (4.10) is split into two parts; first ω_{mx} is found by assuming $\beta = \beta_{cx}$ and solving (4.9); then β_x is found by solving (4.10). Substituting (4.11) into (4.9) it can be shown that:

$$\Delta_{\max} = \tau_e\left(\omega_{lx}\right) - \tau_e\left(\omega_{mx}\right) + \left\{\left(\omega_{lx} - \omega_{mx} - \cos\left(\beta_x\right)r_{px}\right)^2 + \left(\sin\left(\beta_x\right)r_{px}\right)^2\right\}^{-\frac{1}{2}} - \frac{1}{r_{px}}. \quad (4.12)$$

Since (4.12) has an impractical analytical solution, and since τ_e may be numerical in nature, a numerical solution is justified. An interval for ω_{mx} where a unique solution can exist is found as:

$$\omega_{mx} \in \left(\omega_{lx} : \omega_{mx(\max)}\right), \quad (4.13)$$

where $\omega_{mx(\max)}$ is the first zero crossing of c_1 (itself a function of ω_{mx}) greater than ω_{lx}. A simple numerical root-finding algorithm is used to solve for ω_{mx} in the interval of (4.13). The next step is that of finding β_x such that (4.10) is satisfied. Similar to the previous discussion, (4.11) is substituted into (4.10) and β_x is found using numerical root-finding methods in the interval:

$$\beta_x \in \left(\beta_{cx} : 0\right). \quad (4.14)$$

This step concludes solving (4.9) and (4.10) for the xth pole, p_x. Through a sequence of iterations, τ_c is traversed over the entire bandwidth starting from the band edge with the larger group delay, resulting in the desired variation over the entire band of interest. Each new p_x reduces the delay variation to the specified maximum value, over a portion of the bandwidth, $\omega \in \left[\omega_{lx} : \omega_{rx}\right]$. This procedure is illustrated in Figure 4.5, where two poles are introduced, starting at the lower frequency band edge.

The synthesis unit step is summarized in the flow diagram of Figure 4.6.
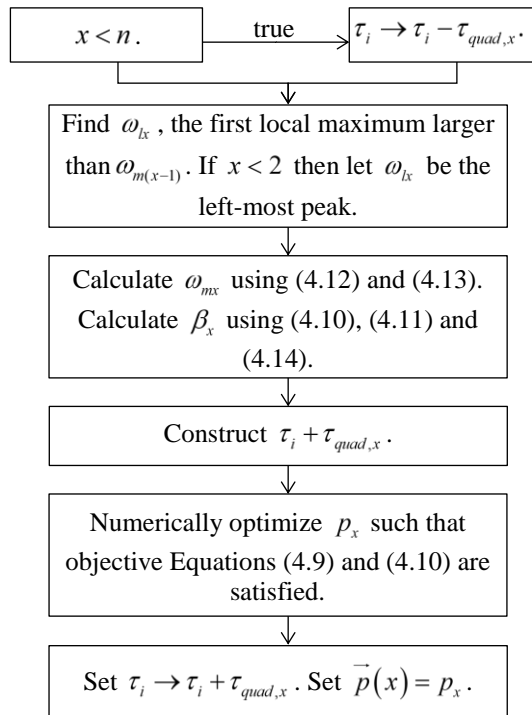
Figure 4.6. The synthesis unit step (S(x)).

Vector $\vec{p}$ is the collection of all n poles p_x calculated up to the current point in the progression of the algorithm. For brevity, only traversal starting at the lower frequency band edge is described. The unit step is used iteratively by the main algorithm of Figure 4.8, both to introduce new poles and also to re-synthesize previously calculated poles. To simplify notation, the unit step in Figure 4.6 is represented by the functional block of Figure 4.7. In Figure 4.8, q is set as the maximum allowed fractional variation between iterations of p_x, allowing for a successful termination of the algorithm.

If the left- and right-most peaks of the group delay are approximately equal, equations for finding ω_{mx} and β_x, have, in certain cases, no solution for the desired maximum variation. This problem can be averted by raising/pre-distorting one band edge with an additional pole before proceeding with the algorithm of Figure 4.8.
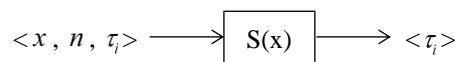
Figure 4.7. Synthesis unit step function block diagram.
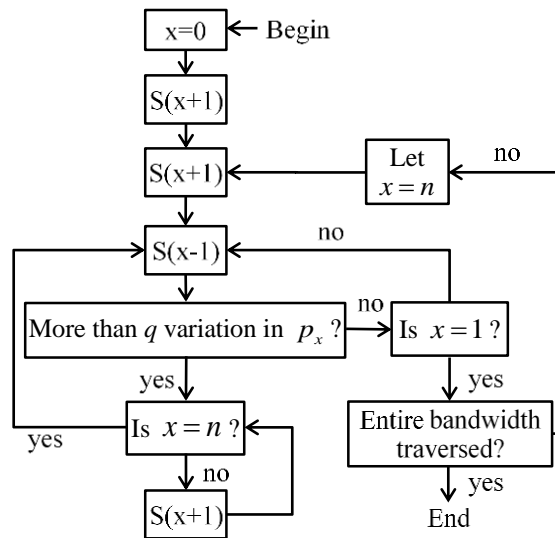
Figure 4.8. Proposed numerical group delay synthesis algorithm.

Finally, each individual $\tau_{\text{quad,x}}$ may be replaced by a more accurate parametric or numerical description of the delay element, provided that the delay function can be approximated by (4.1), the equivalence of which is established through gradient-based optimization of the final all-pass pole/zero locations.

4.3   APPLICATION EXAMPLES

4.3.1   Example 1: Synthesis of a linear group delay function

The proposed algorithm of Figure 4.8 is demonstrated by synthesizing a network with a linear increase in group delay from 2 ns to 4 ns, as may be required when implementing a Fourier transform [10]. This can be done by reducing the variation of the negative of the desired response (cost function) over the desired passband, as illustrated in Figure 4.9.

The initial peak group delay variation is 2000 ps. The desired final variation of the error function is set to 300 ps. From Figure 4.9, $\omega_{l1} = 1.256 \times 10^{10}$ rad/s and using (4.12) and (4.13) it is found that $\omega_{m1} = 1.618 \times 10^{10}$ rad/s with $\beta_{c1} = 0.9553$ rad. Then with (4.10), (4.11) and (4.14) $\beta_1 = 0.4286$ rad.
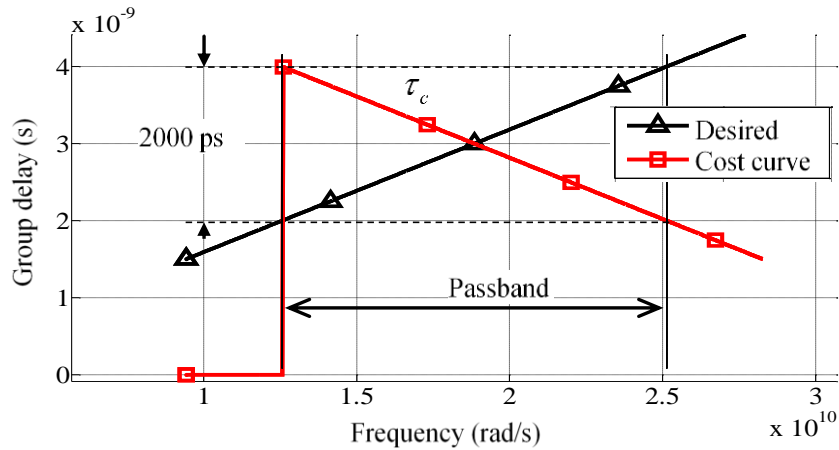
Figure 4.9. Desired linear group delay and cost function.

The initial starting location for p_1 is therefore $\left(-\sigma_1, \ \omega_1\right) = \left(0.994, 17.790\right) \times 10^9$. This value is then optimized numerically using a simple monotonic gradient-based optimizer (no longer using the approximation of Theorem 1 in (4.2)) and found to be $\left(-\sigma_1, \ \omega_1\right) = \left(-\sigma_1, \ \omega_1\right) = \left(0.995, 18.397\right) \times 10^9$. The aforementioned results are summarized in Table 4.1.

Table 4.1. First five iterations of the group delay equalization algorithm.

Variable of interest	Iteration (x)				
	1 (p$_1$)	2 (p$_2$)	3 (p$_1$)	4 (p$_2$)	5 (p$_3$)
$\omega_{lx}/10^{10}$ rad/s	1.2557	1.8067	1.2576	1.8085	2.1878
$\omega_{mx}/10^{10}$ rad/s	1.6180	2.0293	1.6387	2.0878	2.3519
β_{cx} / rad/s	0.9553	1.0035	0.9860	0.9892	1.0051
β_x / rad/s	0.4286	0.5410	0.4636	0.5310	0.5830
$p_{,\text{initial}}/10^9$	(0.994, 17.790)	(0.703, 21.132)	(1.145, 17.782)	(0.766, 21.765)	(0.581, 24.155)
$p_{,\text{optimised}}/10^9$	(0.995, 18.397)	(0.697, 21.561)	(1.188, 18.698)	(0.764, 22.151)	(0.581, 24.482)

The first unit step is concluded with the desired variation satisfied over the interval $\omega \in \left[\omega_{l1} : \omega_{r1}\right] = \left[1.26 : 1.82\right] \times 10^{10}$ rad/s, as shown in Figure 4.10. The unit step procedure is repeated as per the algorithm of Figure 4.8. The second unit step is concluded with the desired delay variation satisfied over the interval $\omega \in \left[1.83 : 2.14\right] \times 10^{10}$ rad/s, as shown in Figure 4.11.
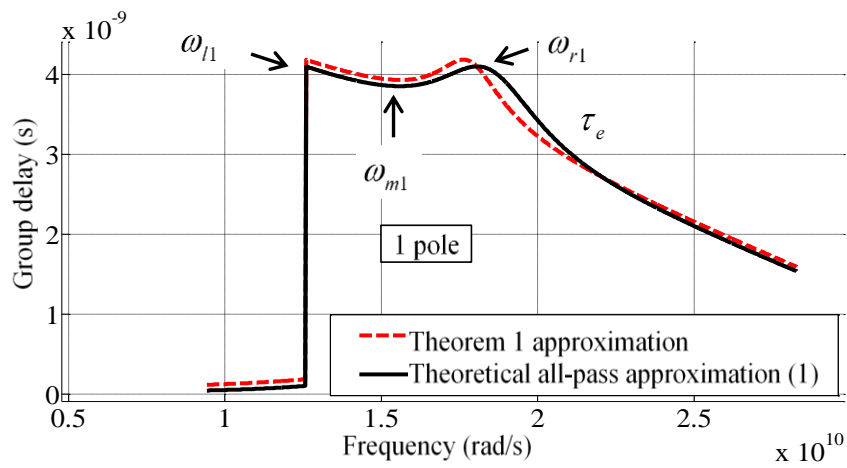
Figure 4.10. Iteration 1 of the delay synthesis algorithm.
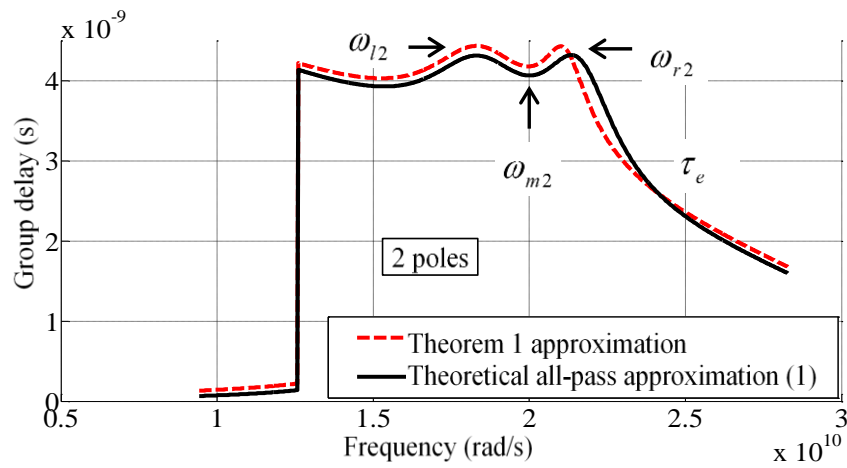
Figure 4.11. Iteration 2 of the delay synthesis algorithm.
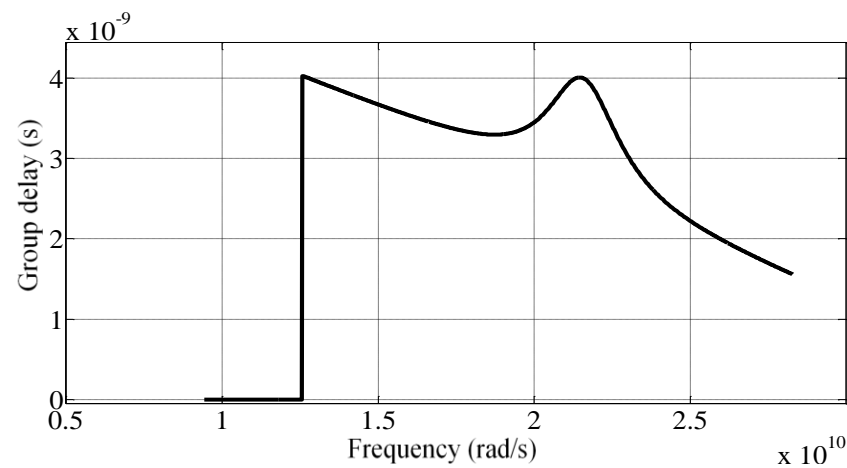
Figure 4.12. Iteration 3 of the delay synthesis algorithm – removal of pole p_1.

The next step involves removing a previously synthesized pole and recalculating its position as per the modification $\tau_i \to \tau_i - \tau_{quad,x}$ shown in Figure 4.6. Figure 4.12 shows the group delay profile after this removal is performed.

The unit step procedure is repeated and the erased pole re-computed, as shown in Table 4.1, under iteration three. A comparison of iterations one and three in Figure 4.10 and Figure 4.12 reveals the improvement in the approximation of the location of pole p_1. The group delay synthesis algorithm repeats in this manner until the values of p_1, p_2 and p_3 converge (the threshold is set at $q = 3\%$, which is arbitrarily chosen).

For brevity, all iterations are not shown in detail. The algorithm ends after 9 iterations with the final cascaded error group delay shown in Figure 4.13.
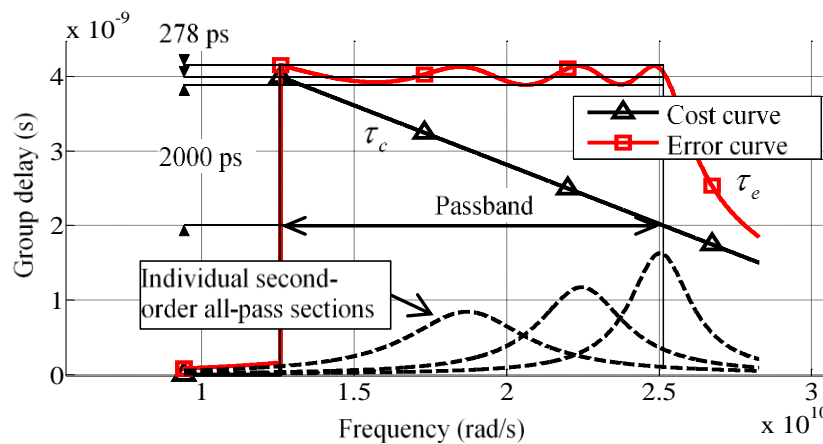
Figure 4.13. Cost and error delay functions.

By removing the cost function response (τ_c), the resulting synthesized linear group delay function (τ_s) is shown in Figure 4.14. The constant offset (representing a constant delay for all frequencies) is not important in ASP applications and is ignored throughout this work. Convergence of the three poles is shown in Figure 4.15.
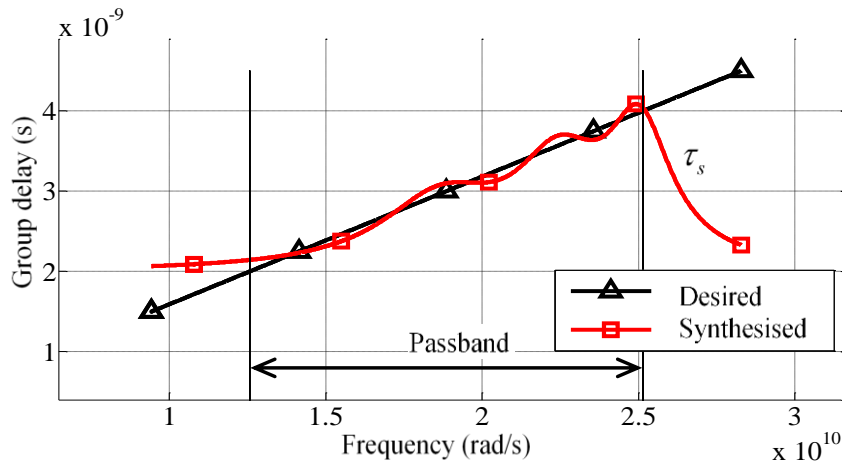
Figure 4.14. Desired and synthesized linear group delay functions.
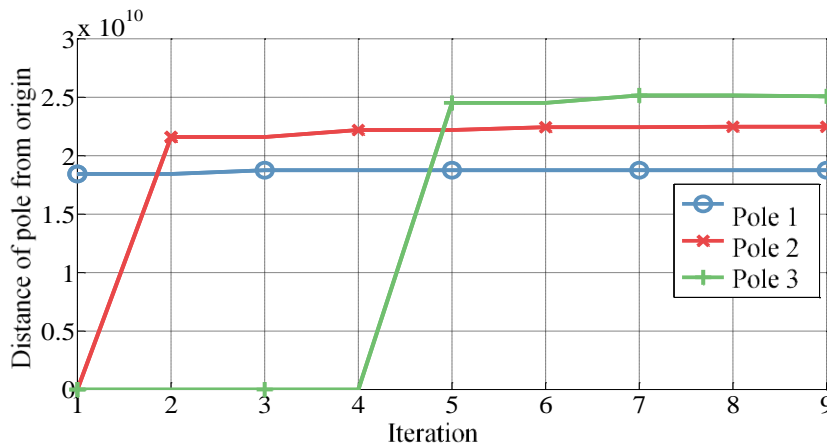
Figure 4.15. Convergence of the pole components – measured as the distance of each respective pole from the origin of the s-plane.

The resulting error function variation is 278 ps over the entire bandwidth whereas the specified variation was 300 ps. Agreement can be further improved by specifying a threshold criterion more stringent than $q = 3\%$, although this would lead to a network of higher order than six. In order to synthesize the desired linear all-pass network, the obtained poles (as shown in Figure 4.16) can be used to calculate the all-pass quadrature pair locations by using the transformation in (4.3).
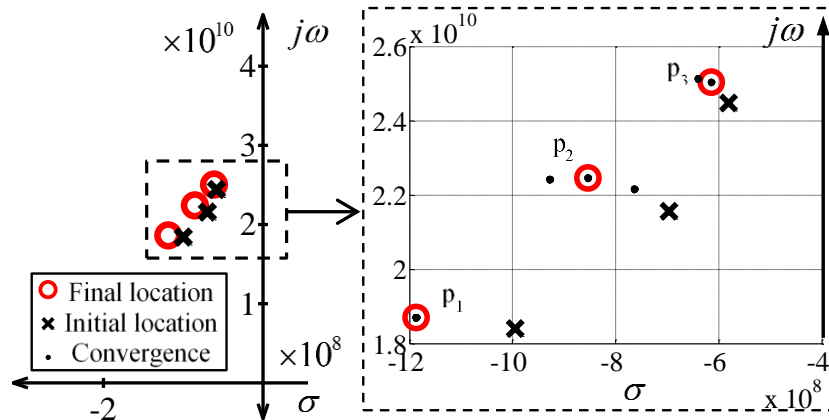
Figure 4.16. Pole locations used to synthesize the sixth-order linear delay approximation of Figure 4.14. Pole locations: p_1 = (-11.879, 1.870), p_2 = (-8.535, 2.245), p_3 = (-6.141, 2.503).

4.3.2   Example 2: Synthesis of Gaussian and quadratic delay functions

More complicated examples are now synthesized, namely Gaussian and quadratic delay functions, with peak delays of 2 ns and 1 ns respectively, as shown in Figure 4.17(a) and Figure 4.18(a). The cost functions (τ_c) are constructed and the desired variation of the error function is set to 130 ps and 80 ps respectively, resulting in the equi-ripple error curves (τ_e) of Figure 4.17(b) and Figure 4.18(b). Synthesized pole locations are shown in Figure 4.17(c) and Figure 4.18(c), while their convergence to the final solution is further illustrated in Figure 4.17(d) and Figure 4.18(d). Finally, the resulting Gaussian delay approximation is shown in Figure 4.17(a) and the quadratic approximation in Figure 4.18(a).

(a)

(b)

(c)

(d)
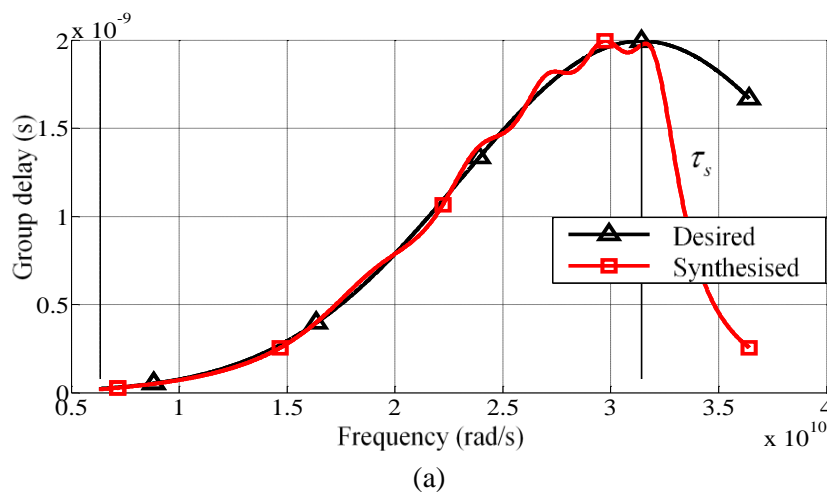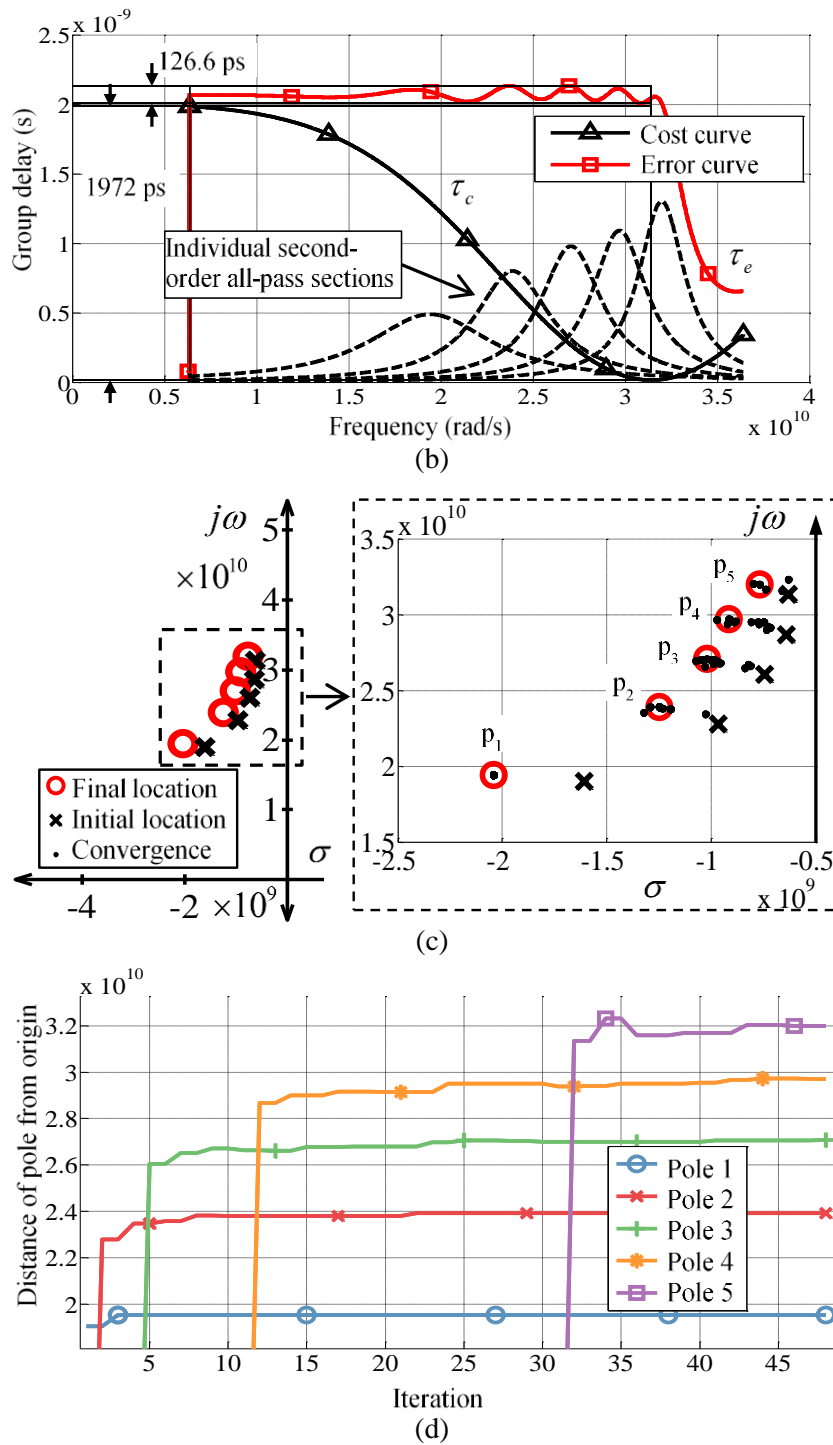
Figure 4.17. (a) Desired Gaussian delay function and the synthesized 10th-order approximation thereof. (b) Group delay cost and error functions. (c) Pole locations: $p_1 = (-2.040, 1.942)$, $p_2 = (-1.247, 2.389)$, $p_3 = (-1.020, 2.706)$, $p_4 = (-0.915, 2.969)$, $p_5 = (-0.767, 3.198)$. (d) Convergence of the pole components – measured as the distance of each respective pole from the origin of the s-plane.
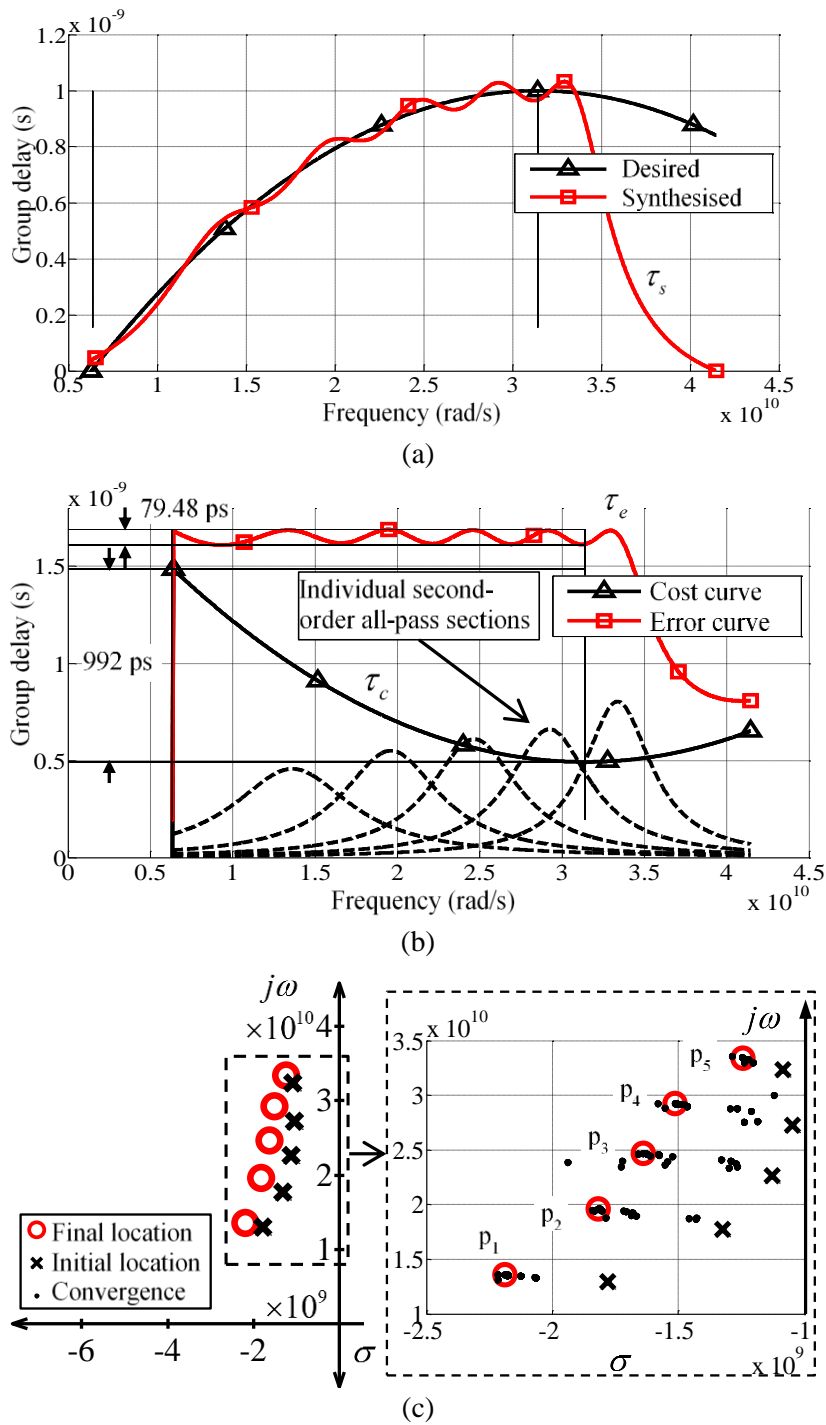
(a)

(b)

(c)

Figure 4.18. (a) Desired quadratic delay function superimposed onto the synthesized 10th-order approximation. (b) Group delay cost and error functions. (c) Pole locations: $p_1 = (-2.188, 1.362)$, $p_2 = (-1.817, 1.960)$, $p_3 = (-1.639, 2.467)$, $p_4 = (-1.513, 2.923)$, $p_5 = (-1.244, 3.337)$.

(d)
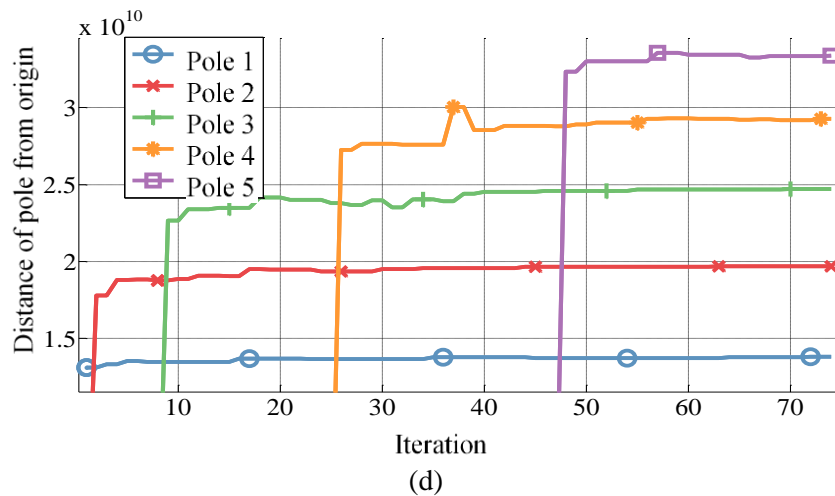
Figure 4.18 (cont.). (d) Convergence of the pole components – measured as the distance of each pole from the origin of the s-plane.

To illustrate the numerical process of synthesizing the Gaussian delay response, iterations that produce new poles are shown in Figure 4.19. This illustrates the systematic approach of the proposed algorithm, where the error delay curve is effectively "stitched-up" into its final form by the inclusion of each subsequent second-order all-pass section and the preceding response "flattened out" by re-computing previous poles. The accuracy of the approximation of Theorem 1 is confirmed by the correspondence of the dotted group delay curves (representing an approximate group delay using Theorem 1) to the solid curves.
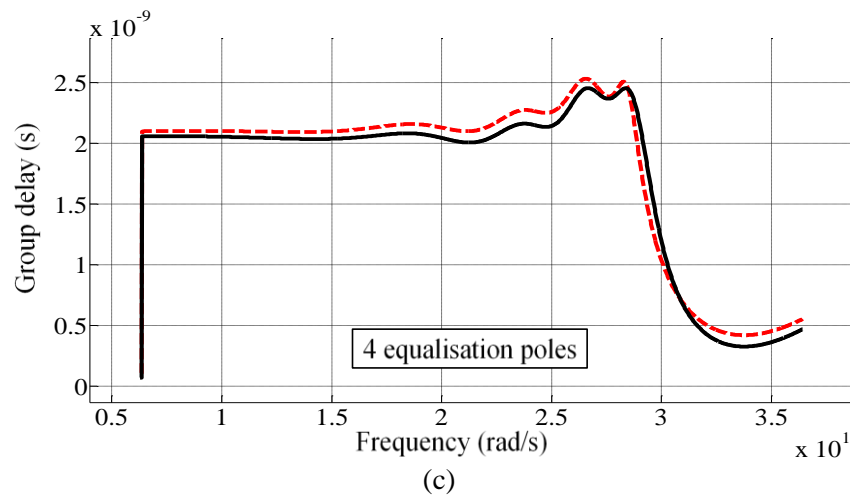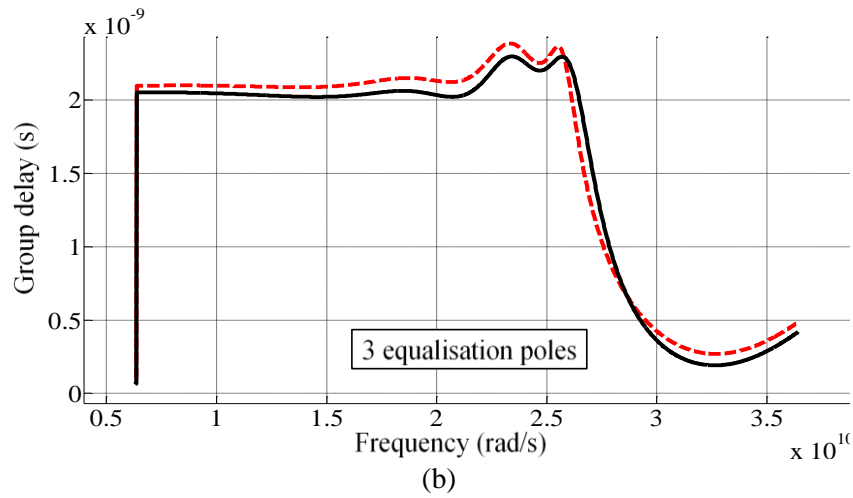

(a)
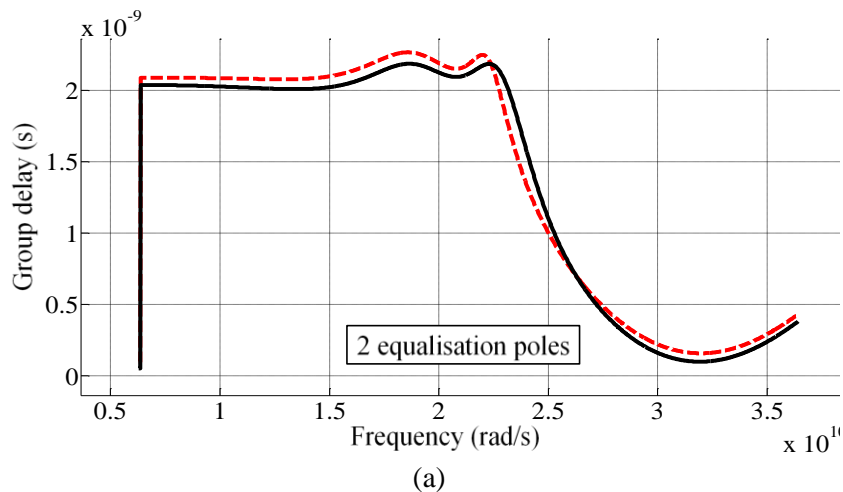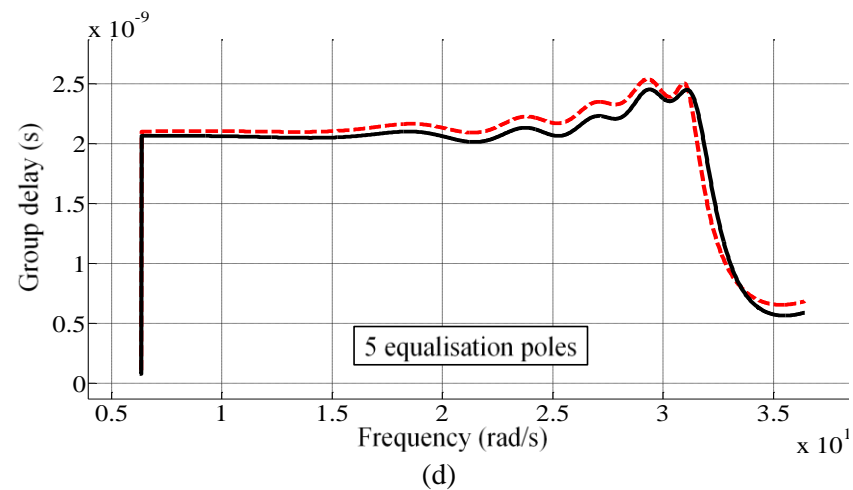
Figure 4.19. Progression of the group delay synthesis algorithm at iterations: (a) 2; (b) 5; (c) 12; and (d) 32 as per Figure 4.17 (d). Solid curves are obtained using (4.1), dotted curves using the approximation of Theorem 1.

4.3.3   Example 3: Equalization of measured BPF group delay response

The proposed algorithm is now demonstrated by reducing the delay variation of physical S-parameter measurements; in this case, on a fifth-order coupled resonator BPF with a fractional bandwidth of 6.8% and magnitude and group delay responses as shown in Figure 4.21(a-b). The generated ideal equalization poles are implemented separately in a circuit solver using lumped-element second-order all-pass sections [62] as shown in Figure 4.20.
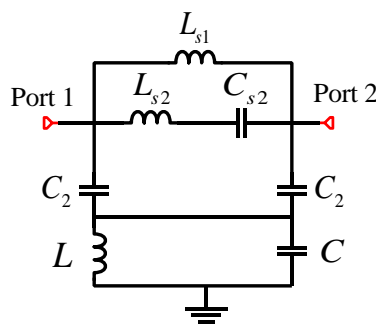
Figure 4.20. Lumped-element second-order all-pass section.

The initial delay variation is measured as 4026 ps and a desired maximum variation of 1100 ps is set for the error function τ_e. This requirement results in a three-section equalizing all-pass network with parameter values as summarized in Table 4.2. A Q-factor of 400 is assumed for the inductors and capacitors. Values for the circuit elements are calculated from the theoretical second-order all-pass sections [35], [39], [62]. All of the individually tuned all-pass sections are cascaded (without further optimization) to obtain the final all-pass network as shown in Figure 4.1. The equalized curves are shown in Figure 4.21(a-b).

The magnitude and precision of component values required in Table 4.2 rule out implementation with discrete surface mount devices. This is true, in general, for microwave filters operating at C-band frequencies [123]. However, the required values are feasible on-chip using MIM capacitors and spiral inductors [88]. Furthermore, in the microelectronic realization, the Q-factors of the inductors may be achieved by either active enhancement of on-chip coil inductors [124] or implementation of active inductors with gm-C type

impedance inversion of fixed MIM capacitors [73], [75]. In both cases, careful control of process tolerances (on-chip MIM capacitors have absolute and relative tolerances of ±25% and ±0.1% respectively) is necessary to ensure the desired group delay response [88]. One approach is by means of post-production tunable CMOS varactors.

A resulting ripple of 1058 ps is obtained in Figure 4.21(b). This is equivalent to a group delay variation reduction of 74%.
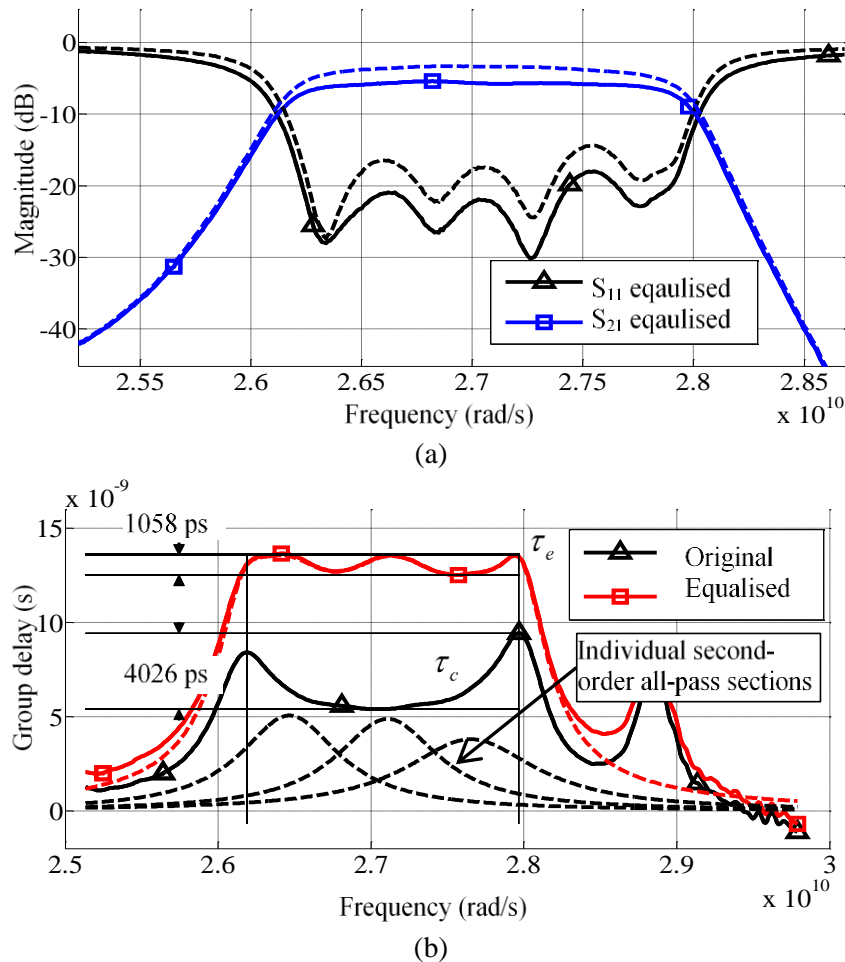

Figure 4.21. (a) Magnitude response of the initial (dotted curve) and equalized (solid curve) filter using a sixth-order all-pass network. (b) Original group delay measurement.

Table 4.2. Parameter values for the three-element equalizing network.

Parameter	Section $\tau_{quad,x}$		
	1	2	3
L_{s1} / (pH)	354.4	352.4	433.3
L_{s2} / (pH)	766.1	757.2	936.1
C_{s2} / (pF)	1.273	1.226	0.9533
C_2 / (fF)	70.92	70.12	86.61
L / (nH)	3.178	3.071	2.376
C / (fF)	308.1	303.9	378.6

Synthesized pole locations are shown in Figure 4.22(a), while their convergence to the final solution is further illustrated in Figure 4.22(b).
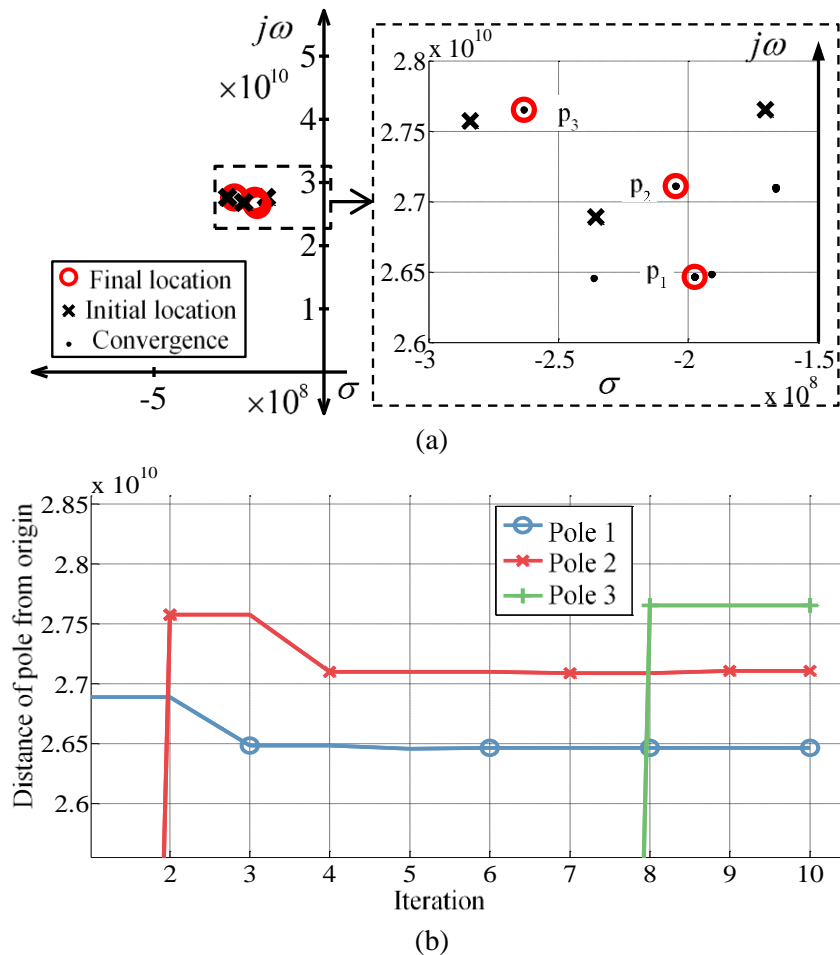

(a)

(b)

Figure 4.22. Equalization pole locations: p_1 = (-1.975, 2.647), p_2 = (-2.049, 2.711), p_3 = (-2.632, 2.765). (b) Convergence of the equalization pole components – measured as the distance of each respective pole from the origin of the s-plane.

The reduced group delay is achieved at the cost of a 2.3 dB insertion loss increase. This increase is, however constant across the band, which is characteristic of dissipative resistive losses attributed to the finite Q-factor of the lumped elements. The effect of a finite Q-factor on the insertion loss and resulting group delay variation of the cascaded system is further illustrated in Figure 4.23. The sixth-order equalizing all-pass network is synthesized in only 10 iterations due to the natural reduction in group delay variation experienced in practical filters because of finite resonator Q-factors.
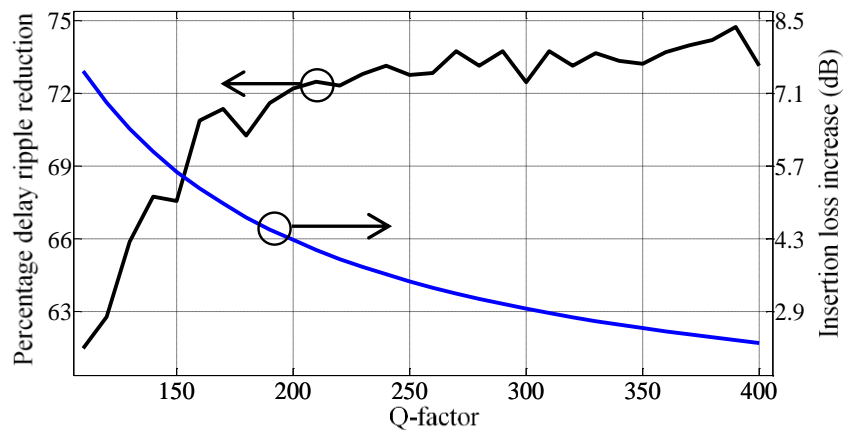

Figure 4.23. Percentage delay variation reduction and insertion loss increase of the cascaded system as a function of the Q-factor.

4.4   COMPARISON OF PROPOSED METHOD WITH EXISTING APPROACHES IN THE LITERATURE

The Gaussian delay function synthesis (Figure 4.17(b)) and practical BPF delay equalization (Figure 4.21(b)) examples described above are repeated here, this time using existing approaches in literature, in order to illustrate the shortcomings stated in the introduction.

The methods in [10], [27], [28], [38], [58], require as input the desired number of poles as well as their starting locations. Here the starting values of the ω_x components are evenly sub-divided across the bandwidth of interest while the starting σ_x components are all set such as to create a peak delay equal to the maximum initial delay variation of the cost

function. This choice can be justified by observing the pole locations synthesized in earlier examples.

Using this initial pole placement, the theoretical treatment presented in [27] is first applied to find an all-pass network for the Gaussian cost function (Figure 4.17(b)). No convergence could be obtained for any number of poles without reducing the bandwidth by 16.6% (upper-frequency edge is removed). Figure 4.24(a) shows the resulting group delay when five all-pass poles are used over this reduced bandwidth. A ripple of 786 ps is achieved over the originally specified passband as opposed to the 126.6 ps demonstrated in 4.3.2 using the proposed method.
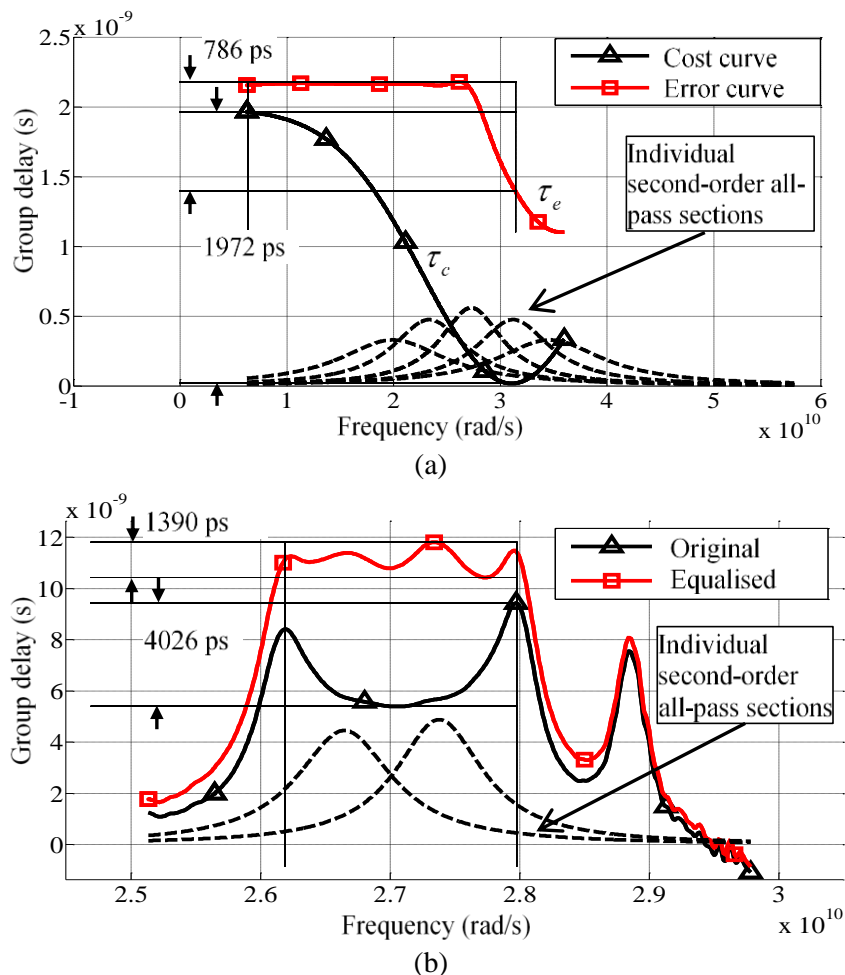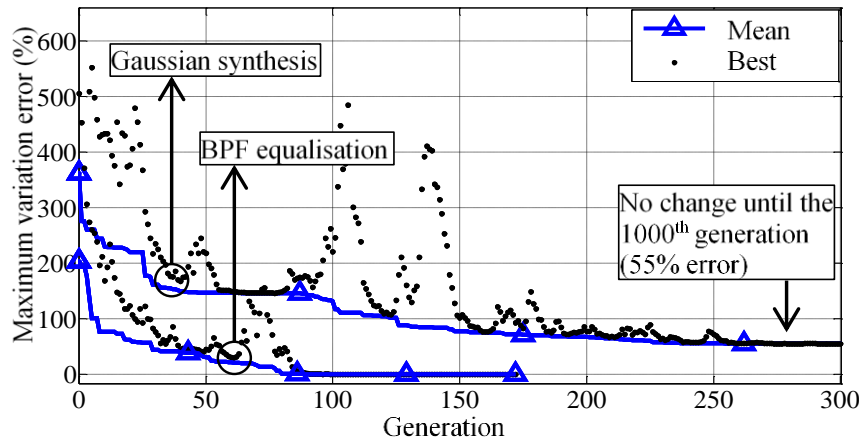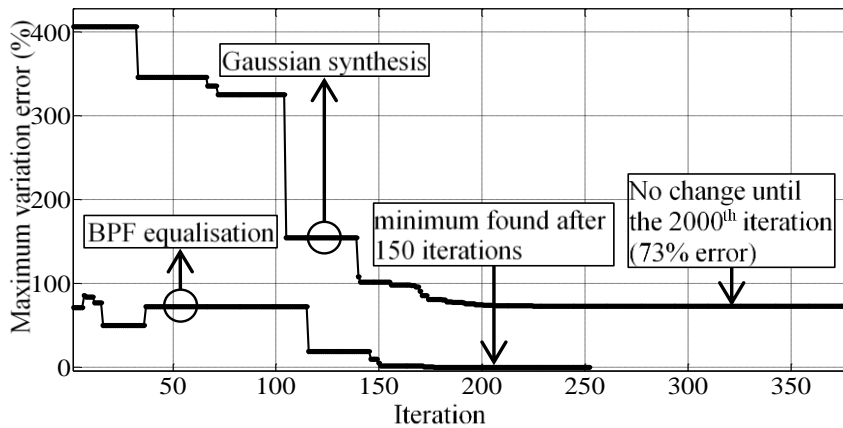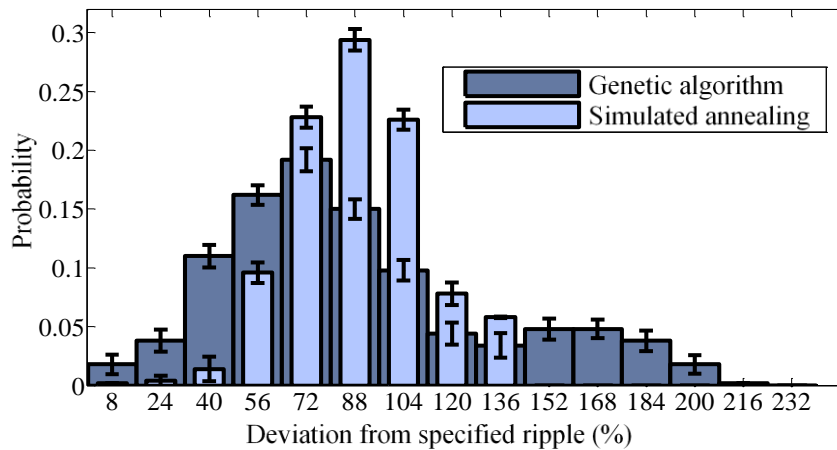

Figure 4.24. (a) Cost and error group delay functions (Gaussian). (b) Initial and equalized group delay responses (practical BPF).

(c)

(d)

(e)

Figure 4.24 (cont.). (c) Genetic algorithm applied to the minimization of the Gaussian and BPF delay error functions. (d) Simulated annealing applied to the minimization of the Gaussian and BPF delay error functions. (e) Probability distribution of the deviation from the specified passband ripple for the Gaussian synthesis problem.
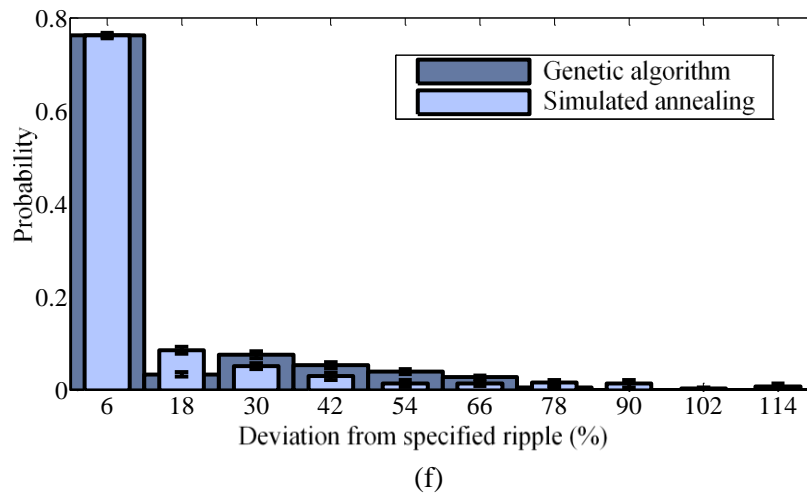
(f)

Figure 4.24 (cont.). (f) Probability distribution of the deviation from the specified passband ripple for the BPF delay synthesis problem.

In a similar manner an equalizing network is synthesized for the BPF delay response (Figure 4.21(b)), using two all-pass poles (no convergence could be obtained for three poles) as shown in Figure 4.24(b).

The above two examples illustrate the following shortcomings associated with the approach in [27]:

1. There is no explicit control over the resulting maximum error delay variation.

2. Often the method does not converge to a solution (depending on the delay cost function and specified number of all-pass sections). In the preceding examples, this prevented the synthesis of the Gaussian delay to within the desired error ripple of 130 ps and the equalization of the BPF to within the specified ripple of 1100 ps, since the bandwidth and number of all-pass sections had to be adjusted to ensure convergence.

Numerical approaches aimed at finding optimal solutions in a large search space have gained prevalence in the modern group-delay synthesis literature [10], [28], [36]–[38], [58]. Here two such approaches which are well suited to the synthesis problem are investigated, namely the genetic algorithm and simulated annealing technique. The maximum ripple of the error function is plotted versus the genetic algorithm generation in Figure 4.24(c) and versus the

simulated annealing iteration in Figure 4.24(d), for a single run. A limit of 1000 generations and 2000 iterations are imposed on the two numerical algorithms. A solution is found to the BPF delay equalization problem after 100 generations (genetic algorithm) and 150 iterations (simulated annealing). On the other hand, an optimal Gaussian group delay response is not successfully synthesized – no improvement occurs after the 300th generation and the 400th iteration respectively.

It is important to note that separate runs converge to different end results, due to the random initial seed values intrinsic to the aforementioned algorithms. Probability distributions of maximum error ripples, for the aforementioned synthesis problems, as shown in Figure 4.24(e) and Figure 4.24(f), are computed from a sample of 500 separate runs with different seed values. The standard deviation is shown for each bin. A limit of 1000 generations and 2000 iterations is imposed on the two numerical algorithms.

These results illustrate that convergence to a global optimum is a matter of finite probability and cannot be guaranteed. Figure 4.24(e) and Figure 4.24(f) further show the large number of local minima that typically exist in the search space. For example, in the Gaussian synthesis problem, 10 parameters are optimized. If each parameter is assigned to a sub-set of 100 points about the initial starting value then 10^{20} function evaluations are required to cover the entire search space. A simulation of 1000 generations requires roughly 20000 function evaluations, which covers only 2×10^{-16} of the search space. Further, the search space partitioning could be too course and fail to find the solution altogether.

Only 1.8% of the genetic algorithm solutions for the Gaussian synthesis problem are optimal. On the other hand, an optimum solution is found to the simpler BPF equalization problem 76% of the time. This might still be insufficient for certain applications, such as the adaptive delay equalization of a practical system.

The proposed method of this work overcomes the aforementioned limitations by, in each case, converging to an optimal solution with an order of magnitude fewer iterations.

4.5   CONCLUSION

An analytically-based numerical method for synthesizing quasi-arbitrary group delay functions using minimum-order cascaded all-pass networks, is presented. The method is compatible with any numerically describable implementation of a second-order all-pass delay network and is shown to converge to a globally optimum, equi-ripple solution, requiring an order of magnitude fewer iterations than state-of-the-art methods. It is also shown that current methods relying on the genetic and simulated annealing algorithms do not always converge to a global optimum (as the proposed method does). As proof of concept, linear, quadratic and Gaussian delay functions are synthesized, to within any arbitrarily specified maximum error of approximation. The proposed method is further demonstrated by reducing the group delay variation of a physically measured fifth-order hairpin resonator BPF with 6.8% relative bandwidth, by 72%.

4.6   MATHEMATICAL DERIVATIONS

4.6.1   Derivation of Theorem 1

Consider a single pole in the s-plane as shown in Figure 4.25. Point P_1 represents the location of the pole of interest. Points P_2 and P_4 are arbitrary locations on the Im $j\omega-$axis separated by a differential distance $\Delta\omega$. The group delay of a linear system described by an S-parameter matrix can be calculated from the S_{21} parameter as follows:

$$\tau \triangleq -\frac{\partial \measuredangle S_{21}}{\partial \omega}. \tag{4.15}$$

Equation (4.15) can be expressed as:

$$\begin{aligned} \tau &= -\lim_{\Delta\omega\to 0} \frac{\Delta\measuredangle S_{21}}{\Delta\omega} = -\lim_{\Delta\omega\to 0} \frac{\Delta\theta_{21}}{\Delta\omega} \\ &\approx -\frac{-\left(\theta_2 - \theta_1\right)}{\Delta\omega}, \end{aligned} \tag{4.16}$$

allowing the relationship between τ and the geometrical representation of Figure 4.25.
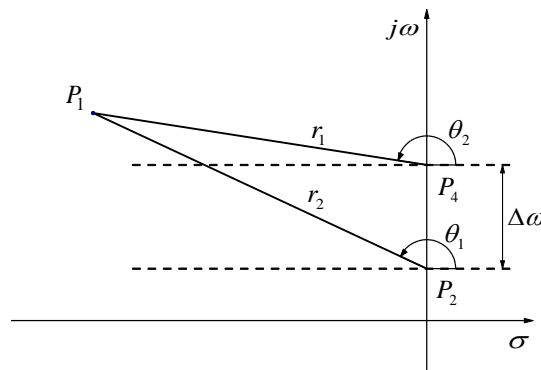

Figure 4.25. Geometry of the group delay response of a single pole.

As ω is traversed from P_2 to P_4 by the differential $\Delta\omega$, the angle $\angle S_{21}$ increases by a corresponding differential amount. The ratio of these two differential changes is the group delay of S_{21}. The distance of P_1 from P_2, determines the resulting differential ratio, or group delay. With the aid of Figure 4.26 and using the approximation presented in (4.16), the relationship between $\Delta\theta$ and $\overline{P_2 P_3}$ is first found as follows:
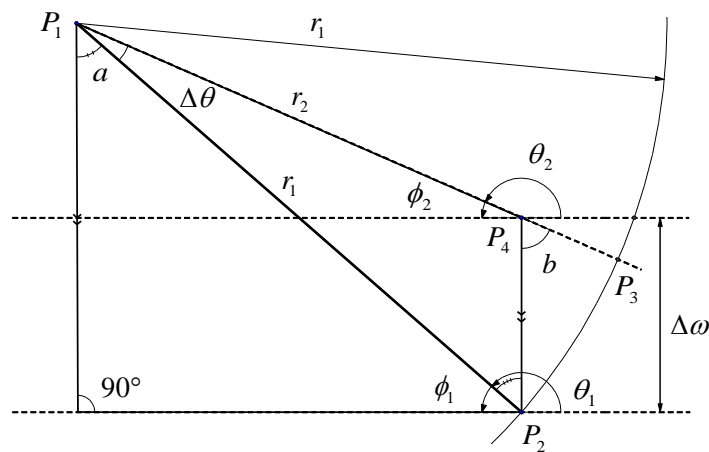

Figure 4.26. A geometrical representation of the group delay caused by a single pole.

$$\Delta\omega \to 0 \Rightarrow r_1 \to r_2$$
$$\Rightarrow r_1 \approx r_2 \approx r$$
$$P_2 P_3 = 2\pi r \frac{\Delta\theta}{2\pi}$$

$$\therefore P_2 P_3 = r\Delta\theta$$

$$\text{Since } \lim_{\Delta\omega \to 0} P_2 P_3 = \overline{P_2 P_3}$$

$$\Rightarrow \overline{P_2 P_3} \approx r\Delta\theta. \tag{4.17}$$

Next, a relationship between $\overline{P_2 P_3}$ and $\Delta\omega$ can be found:

$$\frac{\sin\left(\measuredangle b\right)}{\overline{P_2 P_3}} = \frac{\sin\left(\dfrac{\pi - \Delta\theta}{2}\right)}{\Delta\omega} \quad \text{(Sin rule)}$$

$$\therefore \Delta\omega = \frac{\overline{P_2 P_3}\sin\left(\dfrac{\pi - \Delta\theta}{2}\right)}{\sin\left(\measuredangle a + \Delta\theta\right)}. \tag{4.18}$$

In order to impose the required simplicity of this approximation, it is assumed that $\measuredangle a \approx \pi/2 \text{ rad}$. This implies that the approximation is only valid for $\omega \approx \omega_{px}$ (where ω_{px} is the imaginary component of the pole p_x at point P_1 in Figure 4.26). The validity of the approximation is only important near this point – as justified by the requirement of (4.9) and (4.10). Therefore:

$$\overline{P_2 P_3} \approx \Delta\omega \Rightarrow \Delta\omega \approx \Delta\theta.r$$

$$\therefore \frac{\Delta\theta}{\Delta\omega} = \frac{1}{r}.$$

From (4.16):

$$\tau_{px} \approx \frac{1}{r}. \tag{4.19}$$

4.6.2   Derivation of Theorem 2

Let $\omega = \omega_{mx}$ be the location of the xth desired local minimum of τ_e on the $j\omega$-axis. The frequency $\omega = \omega_{mx}$ is a local minimum or maximum if, and only if,

$$\left. \frac{\partial}{\partial \omega} \left[\tau_e (\omega) \right] \right|_{\omega = \omega_{mx}} = - \left. \frac{\partial}{\partial \omega} \left[\tau_{quad,x} (\omega) \right] \right|_{\omega = \omega_{mx}}. \tag{4.20}$$

Using Theorem 1 the right-hand side of (4.20) can be replaced with $\tau_{quad,x}(\omega) = 1/r_{px}$, where $r_{px} = \overline{P_1 P_2}$. Let:

$$\left. \frac{\partial}{\partial \omega} \left[\tau_e (\omega) \right] \right|_{\omega = \omega_{mx}} = -c_1, \tag{4.21}$$

where c_1 is a constant. From Pythagoras it is know that:

$$r_{px}^2 = \sigma_{px}^2 + \left(\omega_{px}^2 - \omega \right)^2. \tag{4.22}$$

It therefore follows from (4.20) - (4.22) that:

$$-\frac{\partial}{\partial \omega} \left[\frac{1}{r_{px}} \right] = - \left[-\frac{1}{2} \left\{ \sigma_{px}^2 + \left(\omega_{px}^2 - \omega \right)^2 \right\}^{-\frac{3}{2}} .2. \left(\omega - \omega_{px} \right) \right]$$

$$\therefore - \left. \frac{\partial}{\partial \omega} \left[\frac{1}{r_{px}} \right] \right|_{\omega = \omega_{mx}} = - \left(\omega_{px} - \omega_{mx} \right). \left\{ \sigma_{px}^2 + \left(\omega_{px}^2 - \omega_{mx} \right)^2 \right\}^{-\frac{3}{2}}$$

$$\therefore c_1 = \left(\omega_{px} - \omega_{mx} \right). \left\{ \sigma_{px}^2 + \left(\omega_{px}^2 - \omega_{mx} \right)^2 \right\}^{-\frac{3}{2}}. \tag{4.23}$$

This equation can be rewritten as:

$$r_{px} = \left\{ \frac{\omega_{px} - \omega_{mx}}{c_1} \right\}^{\frac{1}{3}}. \tag{4.24}$$

In order to simplify this result, polar co-ordinates are introduced, where $\sphericalangle\beta_x$ is defined as shown in Figure 4.27.
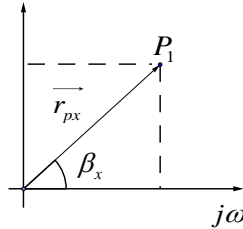

Figure 4.27. Definition of $\sphericalangle\beta_x$.

Equation (4.24) may now be written as:

$$r_{px}{}^3.c_1 = \cos\left(\beta_x\right).r_{px}$$

$$r_{px} = \left\{0, \sqrt{\frac{\cos(\beta_x)}{c_1}}\right\}.$$

(4.25)

The first solution is trivial.

4.6.3   Derivation of (4.8)

A turning point is a local minimum if and only if $\tau''(\omega)$ is continuous at ω_{mx} and if both the following conditions are met (second derivative test): $\tau'(\omega)\big|_{\omega=\omega_{mx}} = 0$ and $\tau''(\omega)\big|_{\omega=\omega_{mx}} > 0.$ Since the first condition has already been satisfied (by Theorem 2) the following restriction must be applied to ensure that the second condition is met:

$$\frac{\partial^2}{\partial\omega^2}\left[\left\{\sigma_{px}{}^2+\left(\omega-\omega_{px}\right)^2\right\}^{-\frac{1}{2}}\right] > c_2, \text{ where:}$$

$$c_2 = -\frac{\partial^2}{\partial\omega^2}\left[\tau_e\left(\omega\right)\right]\Bigg|_{\omega=\omega_{mx}}.$$

(4.26)

This may be re-written as:

$$\frac{2\omega_{mx}^{2} - 4\omega_{mx}\omega_{px} + 2\omega_{px}^{2} - \sigma_{px}^{2}}{\left[\omega_{mx}^{2} - 2\omega_{mx}\omega_{px} + \omega_{px}^{2} + \sigma_{px}^{2}\right]^{5/2}} > c_{2}. \tag{4.27}$$

This expression can be simplified by using the polar coordinate representation:

$$\frac{\left[3\cos^{2}\left(\beta_{cx}\right) - 1\right].c_{1}^{3/2}}{\left[\cos\left(\beta_{cx}\right)\right]^{3/2}} > c_{2}. \tag{4.28}$$

CHAPTER 5   ON-CHIP   IMPLEMENTATION OF ALL-PASS NETWORKS

5.1   INTRODUCTION

Having presented an all-pass network synthesis theory in the previous Chapter, physical implementations of second-order all-pass sections (the building blocks of all-pass networks) are discussed here, in both the CMOS and BiCMOS technology nodes, as justified in the methodology. As mentioned earlier, CCIIs are used to construct the CMOS second-order all-pass network. Improvements to the SOTA CCII design methodology is however necessary, as will be shown. Therefore, an improved design methodology and physical implementation of a suitable CCII is also presented in this Chapter. Finally, the bandwidth of the CMOS all-pass network is extended by considering a BiCMOS implementation.

5.2   CMOS ALL-PASS NETWORK

5.2.1   Introduction

In this work, the synthesis approach presented in [45], based on partial fraction decomposition, is applied to the general CCII configuration in [46] to synthesize an inductorless second-order all-pass network. The proposed all-pass network can be tuned post-production to concurrently account for non-unity A_v and A_i of the CCII, as well as a non-zero R_X, which is a unique contribution to the state-of-the-art. The effects of remaining CCII non-idealities (as defined in Section 2.4.1) on the all-pass response are further considered and it is shown that the proposed design is sufficiently insensitive to them for practical consideration in this work. The proposed design is implemented in the 0.35 µm CMOS technology node using CCIIs based on [99], with the surrounding RC network

implemented in the form of varactors and NMOS transistors operating in the triode region to provide post-production tunability. A post-production *automated* tuning method is further proposed, whereby measured data from a VNA is used in a real-time genetic local optimizer on a PC (which controls various bias voltages using a DAC card) to tune the physical all-pass network. It is shown that this step is crucial to realizing a practical system, leading to the first ever measured results of an active on-chip CMOS dispersive second-order all-pass network, suitable for ASP applications (Q_D-value > 1). It is important to emphasize again, as was done in Section 2.2, that *dispersive* all-pass networks and not flat delay networks such as those used in true time delay networks are required for ASP applications.

This Section is structured as follows:

1. An improved second-order all-pass network is synthesized, and discrepancies between the resulting response and the ideal response caused by CCII non-idealities compared across existing designs in literature.

2. The all-pass network from step 1 is implemented in a 0.35 µm CMOS process from ams AG with a post-production tunability mechanism. Corner analyses and Monte Carlo simulations are presented to justify the need for post-production tunability.

3. The manufactured IC is fine-tuned post-production using an automated genetic optimizer, with measured results presented.

4. A high-precision CCII design used in the construction of the all-pass network is presented after this Section in Section 5.3.

5.2.2 Second-order all-pass network synthesis method

A generalized second-order transfer function can be realized with the CCII-based circuit configuration presented in [46], as shown in Figure 5.1.

The transfer function, considering non-unity A_i and A_v, is [46]:

$$\frac{V_o}{V_i} = \frac{A_i A_v \cdot \left[Y_1 - Y_3 \cdot A_i A_v \right]}{Y_2 - Y_4 \cdot A_i A_v} \cdot \tag{5.1}$$
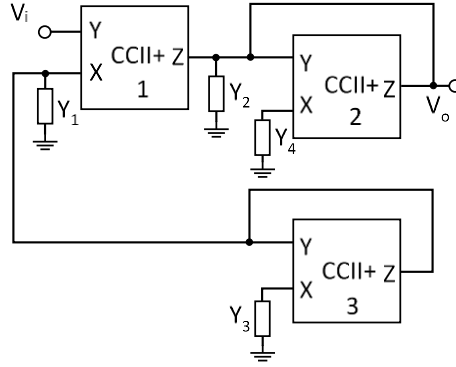

Figure 5.1. CCII-based general second-order transfer function implementation [46].

Extending on the work in [46], the desired second-order all-pass transfer function can be written as:

$$\frac{V_o}{V_i} = \frac{s_n^{\,2} - s_n / Q_D + 1}{s_n^{\,2} + s_n / Q_D + 1} = \frac{Y_N}{Y_D} \tag{5.2}$$

where $s_n = s / \omega_0$, $Q_D = \Delta\tau . \omega_0 / 4$, $\omega_0 = 2\pi f_0$ is the center frequency of the second-order all-pass delay function, and $\Delta\tau$ is the corresponding peak-to-nominal group delay. Following the approach in [45], from (5.2):

$$\frac{Y_N}{s_n \left(s_n + 1 \right)} = \frac{s_n^{\,2} - s_n / Q_D + 1}{s_n \left(s_n + 1 \right)} \cdot \tag{5.3}$$

After division and partial fraction decomposition this becomes

$$\frac{Y_N}{s_n + 1} = s_n + 1 - \frac{s_n \left(2 + 1 / Q_D \right)}{s_n + 1} \cdot \tag{5.4}$$

Similarly, for the denominator, it can be stated that

$$\frac{Y_D}{s_n+1}=s_n+1-\frac{s_n\left(2-1/Q_D\right)}{s_n+1}.\tag{5.5}$$

Dividing (5.4) by (5.5), introducing a real scaling constant α (Ω^{-1}) in both the numerator and denominator and equating with (5.1), the following set of equations can be derived:

$$A_iA_vY_1=\left(s_n+1\right)\alpha,\tag{5.6}$$

$$Y_2=\left(s_n+1\right)\alpha,\tag{5.7}$$

$$A_i^2A_v^2Y_3=\frac{s_n\alpha\left(2+1/Q_D\right)}{s_n+1},\tag{5.8}$$

$$Y_4A_iA_v=\frac{s_n\alpha\left(2-1/Q_D\right)}{s_n+1}.\tag{5.9}$$

It is trivial to show that Y_1 and Y_2 can be realized as a parallel RC network, and Y_3 and Y_4 as a series RC network, with:

$$
\begin{aligned}
R_1&=\frac{A_iA_v}{\alpha}, & C_1&=\frac{\alpha}{A_iA_v\omega_0}, \\
R_2&=\frac{1}{\alpha}, & C_2&=\frac{\alpha}{\omega_0}, \\
R_3&=\frac{A_i^2A_v^2}{\alpha\left(2+1/Q_D\right)}, & C_3&=\frac{\alpha\left(2+1/Q_D\right)}{\omega_0A_i^2A_v^2}, \\
R_4&=\frac{A_iA_v}{\alpha\left(2-1/Q_D\right)}, & C_4&=\frac{\alpha\left(2-1/Q_D\right)}{\omega_0A_iA_v},
\end{aligned}
\tag{5.10}
$$

where the subscripts of Y correspond to that of the R and C components. The proposed *method* and resulting proposed *RC network* presents an extension to the design of [46]. The proposed novel design schematic is shown later in Section 5.2.3.2.

In the preceding derivation, CCII non-idealities other than A_i and A_v (specifically the critical parameter R_X) are not considered, since equating terms in (5.4) and (5.5) with (5.1) becomes

impossible without applying an approximation. It is evident in Figure 5.1 that R_X of both the 2nd and 3rd CCIIs can be incorporated into the series R_4 and R_3, respectively, thus effectively compensating for its non-ideal effect. Perfect compensation for R_X in the 1st CCII is not possible; however, as will be shown later in this Section, this is not necessary for many practical cases ($R_X < 30\ \Omega$). Perfect compensation of practical R_Z values in all the CCIIs is also not possible, however this is not necessary as will be also shown later in this Section. Furthermore, the same CCII non-idealities have been assumed for all three CCII's, as can be assumed for a monolithic implementation.

Next, the effects that CCII non-idealities have on the proposed method, as well as on other all-pass networks in the literature [42], [47], [48], [50]–[54], are investigated and compared. In each case, the transfer function of the all-pass network is derived, including the effect of the non-ideal constituent CCII parameters, and the numerical transfer response computed for variable values of different CCII non-idealities. In all cases, identical non-ideality is assumed for all CCIIs in the second-order network, with only one non-ideality varied at a time. Moreover, the approach of simplifying the analysis by only considering some more important non-idealities for each CCII individually based on its configuration is not taken and instead all non-idealities are considered for each CCII equally. This is the more complete approach.

It is also important to consider, in this comparison, to what degree the synthesis of the network (selection of R and C values) can be adapted to *a priori* known CCII non-idealities. It is evident from their respective transfer functions that non-ideal A_i can be compensated for in this approach, as well as in [42], [47], [51], [52], [54] through appropriate selection of R and C values. Non-ideal A_v can be compensated for in this work and in [53] and non-zero R_X in circuits [51]–[54] (partially in this work as described earlier). In all cases R_Z and R_Y cannot be perfectly compensated for, but as will be shown in this Section below, this is often not necessary for practical cases.

From each numerical transmission response calculation over frequency, two performance metrics are extracted. These are the magnitude response variation ($\Delta|H|$) and the group delay similarity, defined as:

$$\tau_s = \int_0^{2\omega_0} \left| \tau_{ideal} - \tau_{net} \right| d\omega \Big/ \int_0^{2\omega_0} \left| \tau_{ideal} \right| d\omega, \tag{5.11}$$

where τ_{ideal} is the delay response with ideal CCIIs and τ_{net} the non-ideal network response. A $\tau_s = 0$ indicates that the two responses are identical whereas a $\tau_s > 0$ indicates dissimilarity between the group delays. This definition captures both $\Delta\tau$ and ω_0 deviations, as well as deviations from the ideal group delay curve *shape*. This is important to consider, as CCII non-idealities can disrupt the network response to such an extent that the network's delay no longer resembles that of a second-order all-pass network.

Finally, as all the considered networks are underdetermined (fewer bounding equations than *R* and *C* unknowns) the following component choices are made to ensure a fair comparison between the circuits, as shown in Table 5.1. In all cases $C_1 = C_2 = C$.

Table 5.1. Component choices for inter-circuit comparison.

All-pass network	Imposed conditions	Design eq. 1	Design eq. 2	Design eq. 3
[47]	$R_2 = R_3$ $= R_4 = R$	$R_2 = \beta$	$R_1 = \beta / \left(Q_D{}^2 \cdot A_i \right)$	$C = \dfrac{\left(A_i{}^2 \cdot Q_D \right)}{\left(\omega_0 \cdot \beta \right)}$
[48]	--	$R_1 = \beta$	$R_2 = \beta / Q_D{}^2$	$C = Q_D / \left(\omega_0 \cdot \beta \right)$
[50]	$R_1 = R_2$	$R_1 = \beta$	$R_3 = \beta / Q_D{}^2$	$C = Q_D / \left(\omega_0 \cdot \beta \right)$
[51]	--	$R_1 = \beta$	$R_2 = A_i \beta / Q_D{}^2$	$C = Q_D / \left(\omega_0 \cdot \beta \right)$
[52]	$R_2 = R_3$	$R_1 = \beta$	$R_2 = A_i \beta / Q_D$	$C = Q_D / \left(\omega_0 \cdot \beta \right)$
[53]	$R_2 = 2R_3$	$R_3 = \beta$	$R_1 = 4\beta / Q_D{}^2$	$C = Q_D / \left(2\omega_0 \cdot \beta \right)$
[54]	$R_2 = R_3$	$R_1 = \beta$	$R_2 = A_i \beta / Q_D$	$C = Q_D / \left(\omega_0 \cdot \beta \right)$

The value for β is chosen as 10^3 (as this leads to a realizable resistance on-chip), Q_D is set as 2 and ω_0 as $2.\pi.200.10^6$ (corresponding to design choices made later in this Section). It is, however, reasonable to expect the general conclusions using these parameters to hold for other comparable design choices as well.

For the proposed circuit, as well as the network in [8], the design choices $C_1 = \alpha/\omega_0$, $C_2 = \alpha(2 + 1/Q)/\omega_0$, $C_3 = 2\alpha/\omega_0 Q$, $R_1 = 1/\alpha$, $R_2 = 1/(\alpha \cdot (2 + 1/Q))$, and $R_3 = Q/2\alpha$ are

made. In both cases α is chosen as 10^{-3}, to ensure agreement to the component values in Table 5.1 for the other networks.

Figure 5.2 shows $\Delta|H|$ and τ_s over the various CCII non-idealities. In each sub-figure, the solid curve represents the best-case response, where all the non-idealities are known beforehand and compensated for (if possible) in the selection of R and C values. The dotted curves, on the other hand, represent the response without any compensation for non-idealities (even if compensation is possible). This allows the sensitivity of the various methods to be compared, since in some cases compensation is possible and in other cases it is not possible.

From Figure 5.2(a-b), it can be seen that the proposed circuit and the design in [53] are the only designs in which non-unity A_v can be compensated for, regardless of the chosen Q_D, as is evident in Table 5.1. For $\Delta|H| < 0.5\ dB$, designs [42], [48], [50] require A_v precisions within 1.4 % of unity, designs [51], [52], [54] A_v precisions within 2.7 % of unity, and design [47] an A_v precision within 6 % of unity (making it the design most resilient to CCII A_v variation without explicit compensation applied). Achieving even this level of precision, however, requires a high-precision CCII with feedback control which complicates the design and reduces the achievable bandwidth [87]. This point is further illustrated later with a Monte Carlo analysis. Figure 5.2(c-d) shows the benefit of the proposed design over the circuit in [53], since it allows for ideal compensation of A_i variation, regardless of the chosen Q_D, as is evident in Table 5.1. In [53], A_i precision within 3 % of unity is required for $\Delta|H| < 0.5\ dB$ – a value that is difficult to guarantee with CMOS CCIIs – again illustrated later with a Monte Carlo analysis.

In Figure 5.2(e-f), the proposed method is shown to require $R_X < 30\ \Omega$ for $\Delta|H| < 0.5\ dB$, which is easily achievable with most CCIIs presented in the literature. In contrast, design [47] requires $R_X < 13\ \Omega$ and design [42] $R_X < 4\ \Omega$ for $\Delta|H| < 0.5\ dB$. This, again, necessitates high-precision CCIIs. Even though designs [51]–[54] can compensate for non-zero R_X, as was shown in Figure 5.2(a-d), they still require high-precision CCIIs to achieve the required A_v and A_i precisions.
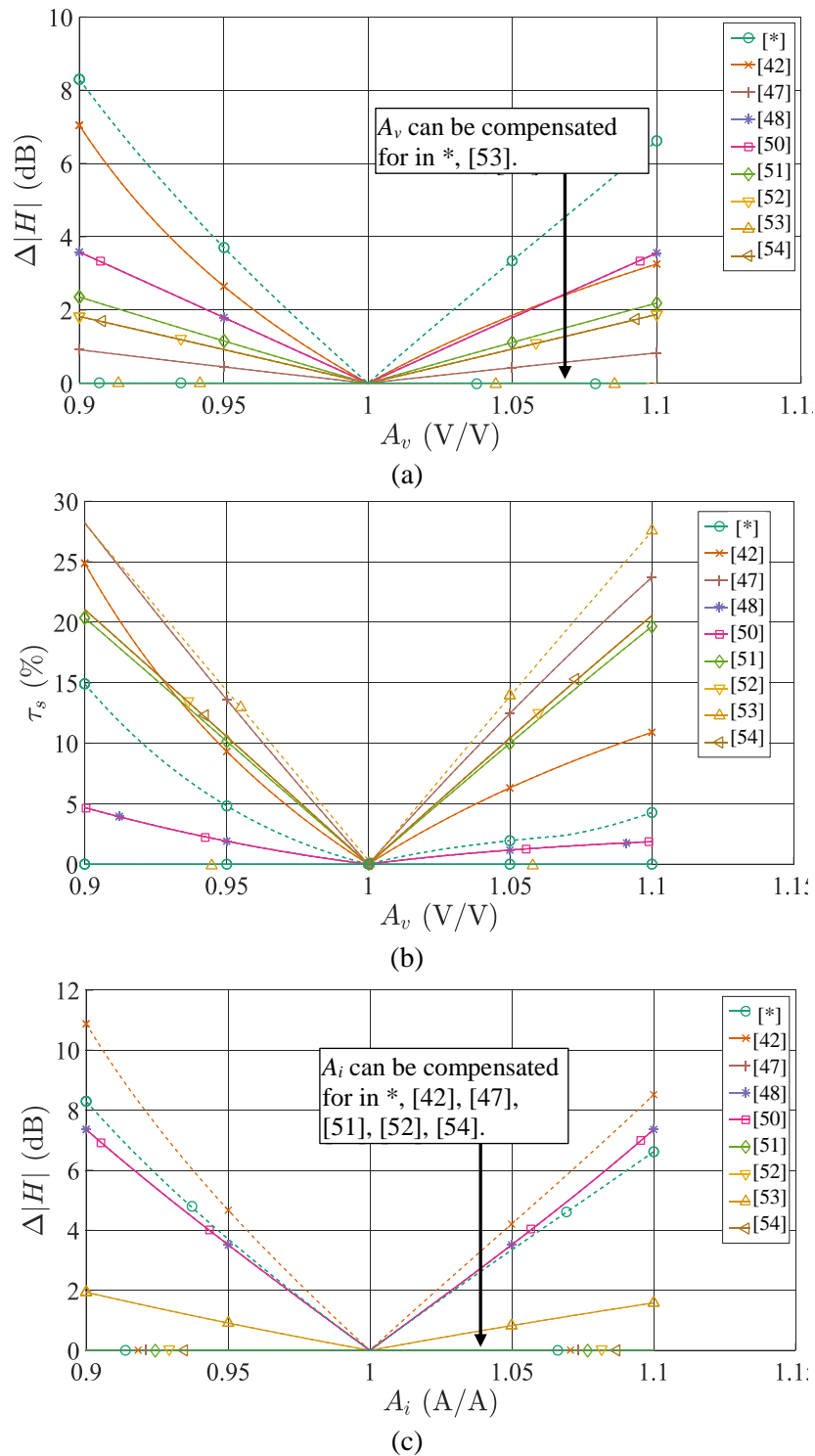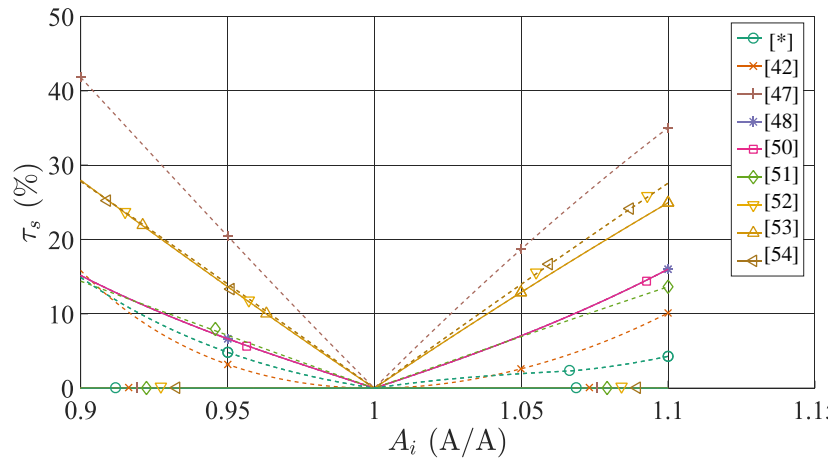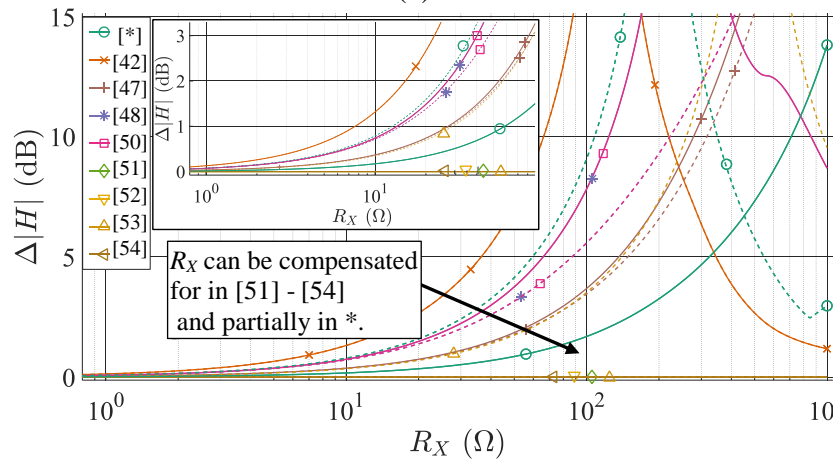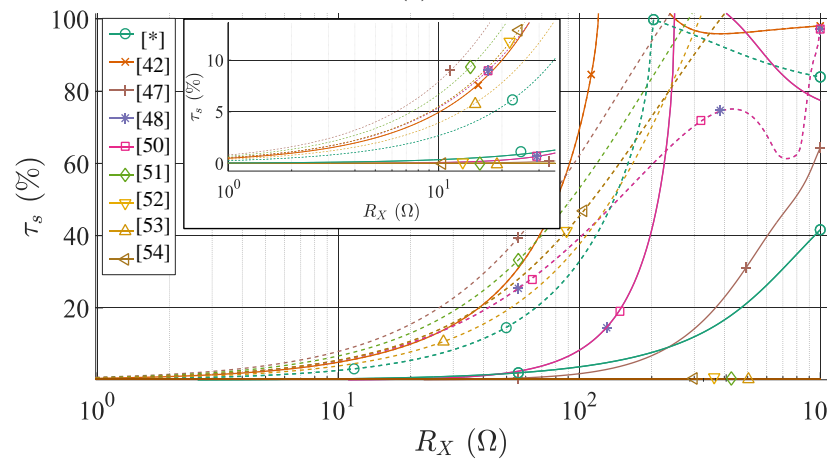
Figure 5.2. Effects of CCII non-idealities on second-order all-pass network response characteristics: (a) non-unity A_v vs. $\Delta|H|$, (b) non-unity A_v vs. τ_s, (c) non-unity A_i vs. $\Delta|H|$.

(d)

(e)

(f)

Figure 5.2 (cont.). Effects of CCII non-idealities on second-order all-pass network response characteristics: (d) non-unity A_i vs. τ_s, (e) non-zero R_X vs. $\Delta|H|$, (f) non-zero R_X vs. τ_s.
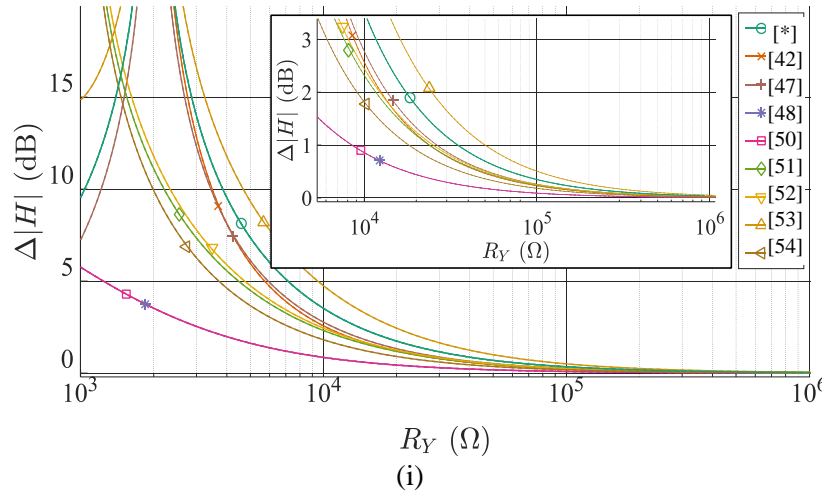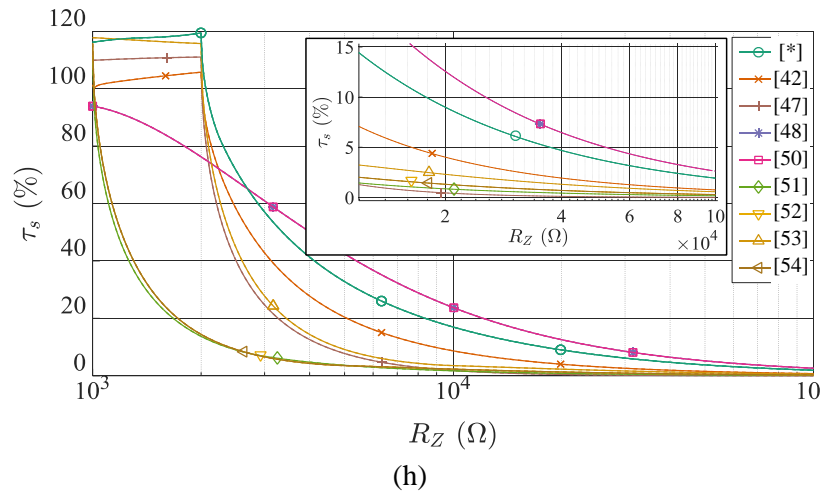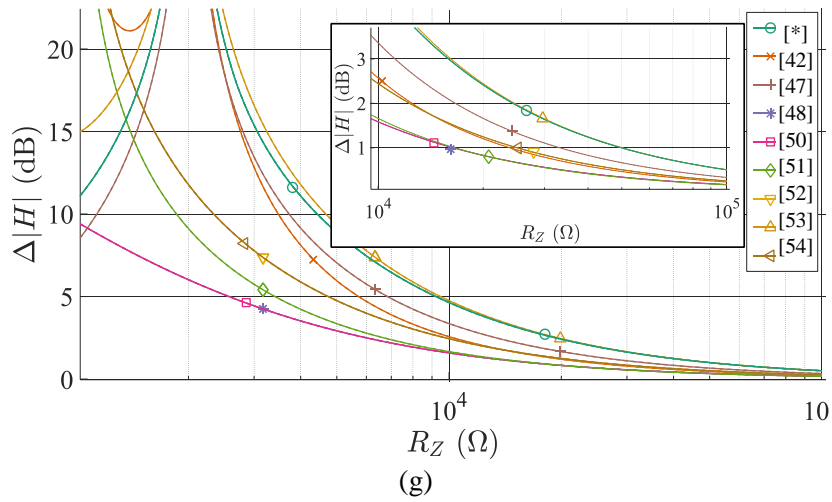
Figure 5.2 (cont.). Effects of CCII non-idealities on second-order all-pass network response characteristics: (g) finite R_Z vs. $\Delta|H|$, (h) finite R_Z vs. τ_s, (i) finite R_Y vs. $\Delta|H|$.
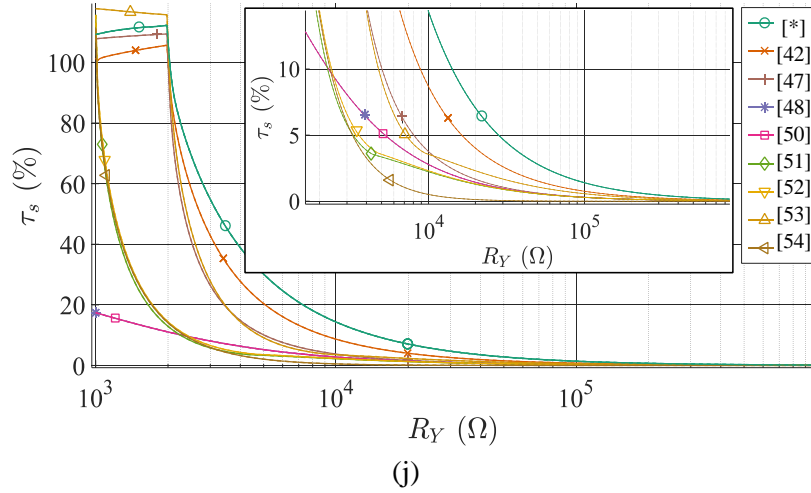
Figure 5.2 (cont.). Effects of CCII non-idealities on second-order all-pass network response characteristics: (j) finite R_Y vs. τ_s. In all cases * denotes the network proposed in this work.

Figure 5.2(g-h) indicates that for $\Delta|H| < 0.5\,dB$, the proposed design requires $R_Z > 100\,\text{k}\Omega$, circuit [47] an $R_Z > 66\,\text{k}\Omega$, designs [42], [52], [54] an $R_Z > 50\,\text{k}\Omega$, and designs [48], [50], [51] an $R_Z > 33\,\text{k}\Omega$. An $R_Z \approx 100\,\text{k}\Omega$ can be achieved with a cascode current mirror. Even though current conveying bandwidth is reduced by a low R_Z, its effect on CCII bandwidth is much smaller compared to the bandwidth reduction incurred by the feedback required in high-precision CCII designs.

Figure 5.2(i-j) indicates that a $\Delta|H| < 0.5\,dB$ requires $R_Y > 100\,\text{k}\Omega$ for all the designs, a value achievable with many CCIIs. The proposed network distinguishes itself from published literature in that it can compensate for known a priori non-ideal A_i and A_v *simultaneously* while at the same time displaying an acceptable sensitivity to R_X, R_Y and R_Z without the need for bandwidth-limiting high-precision CCIIs. This design thus presents a meaningful contribution to the SOTA. Figure 5.2 also illustrates the importance of post-production tunability of the R and C components to account for *a postiori* knowledge of CCII non-idealities. This need is exacerbated by other circuit non-idealities not captured by the above analysis, such as component value variation due to process tolerances and parasitic capacitances.

Finally, as was mentioned earlier, varying Q_D (and therefore $\Delta\tau$) has no effect on the $|H|$ and τ_s versus A_i and A_v plots for the proposed design. On the other hand, varying Q_D influences

the response *of the proposed compensated network* in terms of R_X, R_Y and R_Z as shown in Figure 5.3. There is clearly scope for trade-off consideration (between R_X, R_Y and R_Z) in the choice of Q_D given a requirement for $\Delta|H|$ (for instance RY can be increased by decreasing the bias current through the current mirror). Such a tradeoff is, on the other hand, difficult to achieve in passive soft-substrate designs without the use of active enhancement techniques.
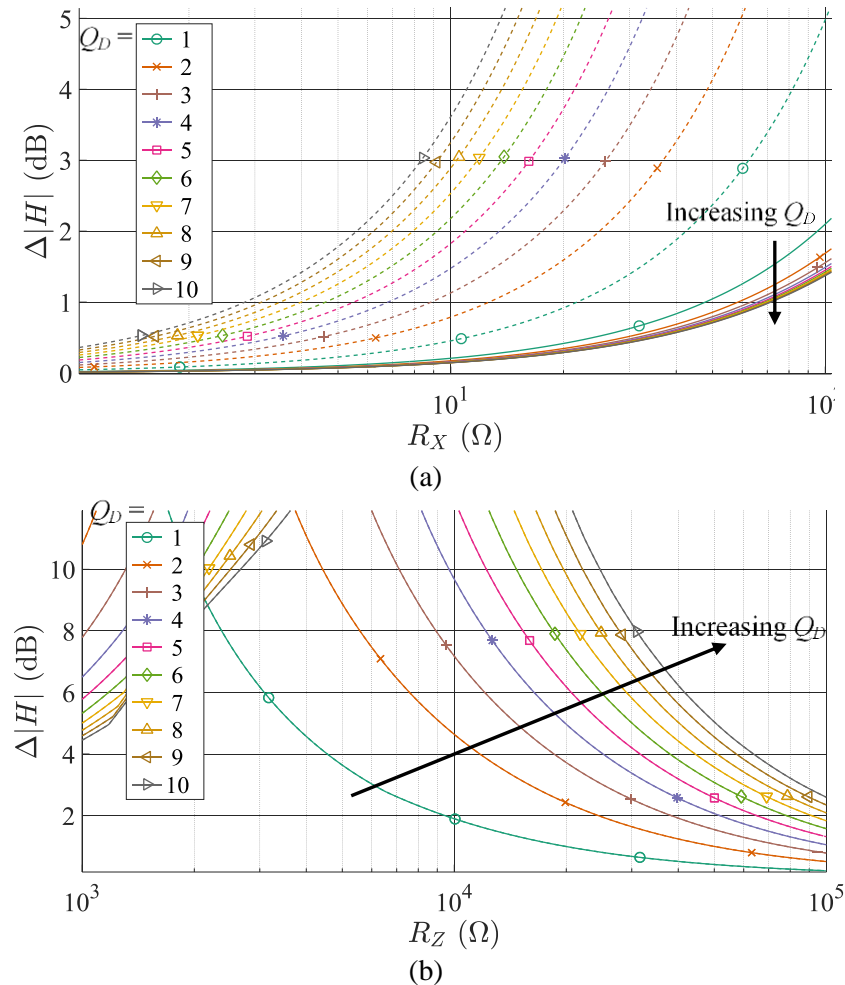

Figure 5.3. Tradeoff between the second-order all-pass Q_D-value and $\Delta|H|$ for various CCII R_X and R_Z non-idealities. The solid and dotted curves represent compensated and uncompensated cases for R_X, respectively, as defined earlier in the text preceding Figure 5.2.
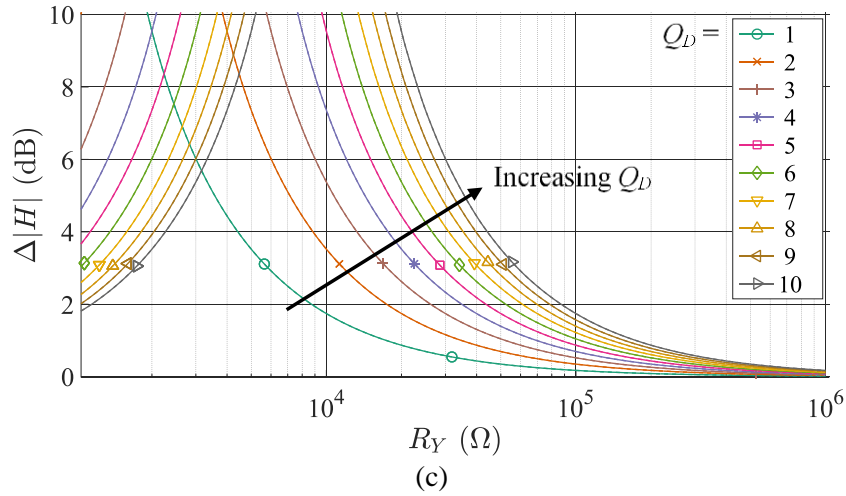
(c)

Figure 5.3 (cont.). Tradeoff between the second-order all-pass Q_D-value and $\Delta|H|$ for various CCII R_Y non-idealities.

5.2.3 CMOS implementation of second-order all-pass network

Having proposed a second-order all-pass network suitable for monolithic integration in CMOS, and verifying its improvement on the state-of-the-art w.r.t. resilience to CCII parametric variation, the detailed implementation of the all-pass network of Figure 5.1 is presented in a 0.35 µm CMOS technology node, using the synthesis in Section 5.2.2.

5.2.3.1 CCII implementation

The high-bandwidth, low-R_X CCII presented in [99] with the characteristics $A_i \approx 0.976$, $A_v \approx 0.96$, $R_X < 20\ \Omega$, $R_Y \approx 25\ \text{k}\Omega$ and $R_Z \approx 35\ \text{k}\Omega$, is used as the basis of the CCII design in this work. In addition to the schematic in [99] a cascode current mirror is used at port Z (as opposed to a single current mirror stage) and different transistor aspect ratios are used in some instances. Even though higher-precision CCIIs have been reported, their design is complicated by necessary stability analyses and compensation networks, as shown in [87] and further discussed in Section 5.3. Furthermore, the higher precision comes at the expense of lower bandwidth. As per the discussion in Section 5.2.2, the non-idealities of the chosen CCII lie within acceptable bounds and a bandwidth-precision tradeoff is not necessary. A minimum *theoretical* magnitude variation of ~2 dB can be achieved after compensation using the proposed network, which is attributed mostly to R_Y and R_Z as per Figure 5.2(e-j).

The CCII used in this work is presented in Figure 5.4 (with bias currents shown) and transistor aspect ratios given in Table 5.2. The design as presented in [99] is used and optimized in simulation to minimize the deviation of A_v and A_i from unity, minimize R_X and maximize R_Z and R_Y.

Table 5.2. Transistor sizes chosen for devices in Figure 5.4.

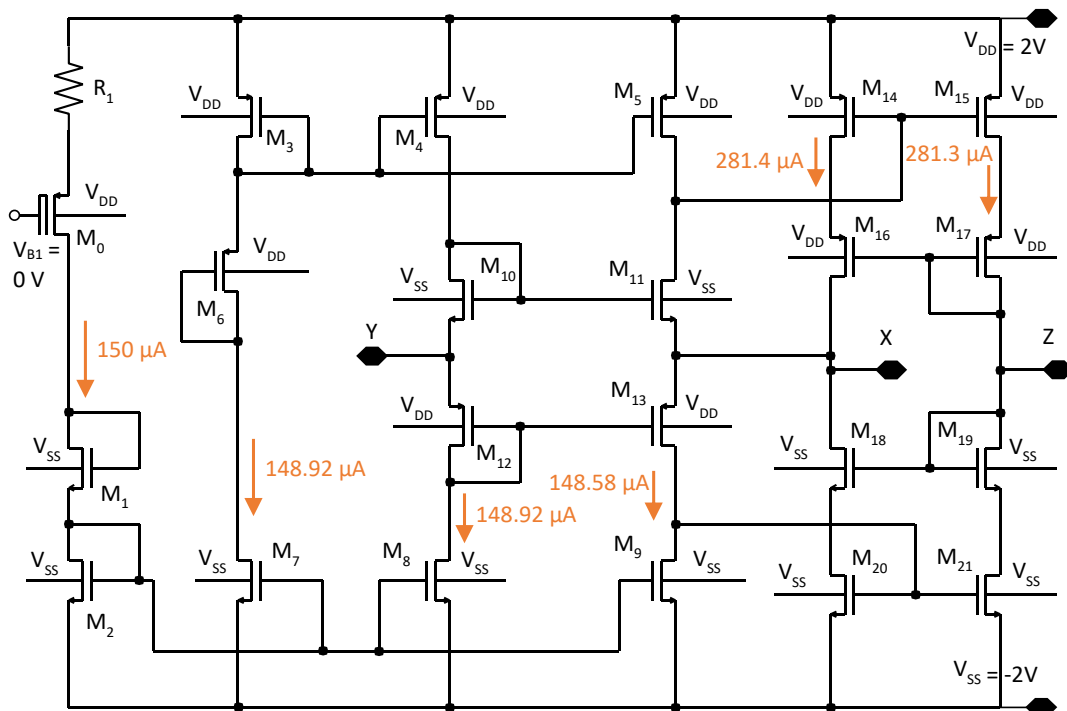| Device	W/L (µm/µm)	Device	W/L (µm/µm)	Device	W/L (µm/µm)
M0	11/0.5	M8	10/1.95	M16	40/0.35
M1	10/1.95	M9	10/1.95	M17	40/0.35
M2	5/1.95	M10	20/0.35	M18	10/0.35
M3	18.9/0.75	M11	20/0.35	M19	10/0.35
M4	18.9/0.75	M12	30/0.35	M20	10/0.35
M5	18.9/0.75	M13	30/0.35	M21	10/0.35
M6	25/1.5	M14	30/0.35		
M7	10/1.95	M15	30/0.35		

Figure 5.4. CCII+ used for the second-order all-pass network proposed in this work.

Resistor R_1 is chosen as 333 Ω to ensure an M_0 bias current of 150 µA with V_{B1} set to 0 V. The bias currents were in turn determined by the aforementioned optimization. To establish a range of possible operating conditions of the CCII, a Monte Carlo analysis is performed on the circuit, leading to the performance characteristic range as shown in Table 5.3.

Table 5.3. Performance characteristics of the CCII in Figure 5.4.

Parameter	Mean	Standard deviation	Minimum	Maximum
A_v (V/V)	1.083	9.814×10^{-3}	1.062	1.114
A_i (A/A)	0.992	6.306×10^{-3}	0.968	1.009
R_X (Ω)	18.66	8.611	12	161
R_Z (kΩ)	70	26	23	125
R_Y (kΩ)	26	2.4	18	30

These simulation results agree well with values presented in the literature [99]. Resistance R_X is well within maximum bounds for successful compensation in most scenarios, as only 3% of the simulated 500 samples have R_X of greater than 20 Ω and only 1.4 % have R_X of greater than 30 Ω. The maximum achievable -3 dB transmission bandwidth (limited by the voltage transfer between ports Y and X) for the nominal corner is ~ 500 MHz.

5.2.3.2   All-pass implementation

Having designed the CCII as in Figure 5.4, the remainder of the proposed all-pass network is implemented as shown in Figure 5.5. In this work $f_0 = 200$ MHz and $\Delta\tau = 7$ ns are chosen, leading to a Q_D of 2.2, which is larger than 1. Component values are then calculated using (5.10), with α chosen as 10^{-3}. Using the mean A_v and A_i from Table 5.3, the synthesis results in $R_1 = 1074$ Ω, $R_2 = 1$ kΩ, $R_3 = 470$ Ω, $R_4 = 695$ Ω, $C_1 = 0.74$ pF, $C_2 = 0.80$ pF, $C_3 = 1.69$ pF, $C_4 = 1.15$ pF. All resistors are implemented as NMOS devices operating in the triode region, where the bias voltages $V_{R1} - V_{R4}$ are chosen such that the effective channel resistance corresponds to $R_1 - R_4$ above. The NMOS devices in Figure 5.5 are chosen as (10 µm/0.5 µm) for $M_{1,2}$ and (20 µm/0.5 µm) for $M_{3,4}$, requiring nominal bias voltages of $V_{R1} = 1.943$ V, $V_{R2} = 1.974$ V, $V_{R3} = 1.952$ V, $V_{R4} = 1.832$ V.

Capacitors are implemented using accumulation-mode MOS varactors [125] which are tuned with gate bias voltages $V_{C1} - V_{C4}$ such that the effective capacitance values correspond to $C_1 - C_4$ above. The peak capacitances of the varactors are chosen as $C_{1p} = C_{2p} = 1.8$ pF and $C_{3p} = C_{4p} = 2.88$ pF, allowing for a sufficient tuning range around the nominal values. Nominal bias voltages $V_{C1} = -0.028$ V, $V_{C2} = -0.063$ V, $V_{C3} = 0.107$ V, $V_{C4} = 0.270$ V are then required. An 8-bit DAC is required to set and later to fine-tune the bias voltages. In this

work, the DAC is implemented off-chip due to prototyping space restrictions. Lastly, remaining bias values are set as in Figure 5.4.
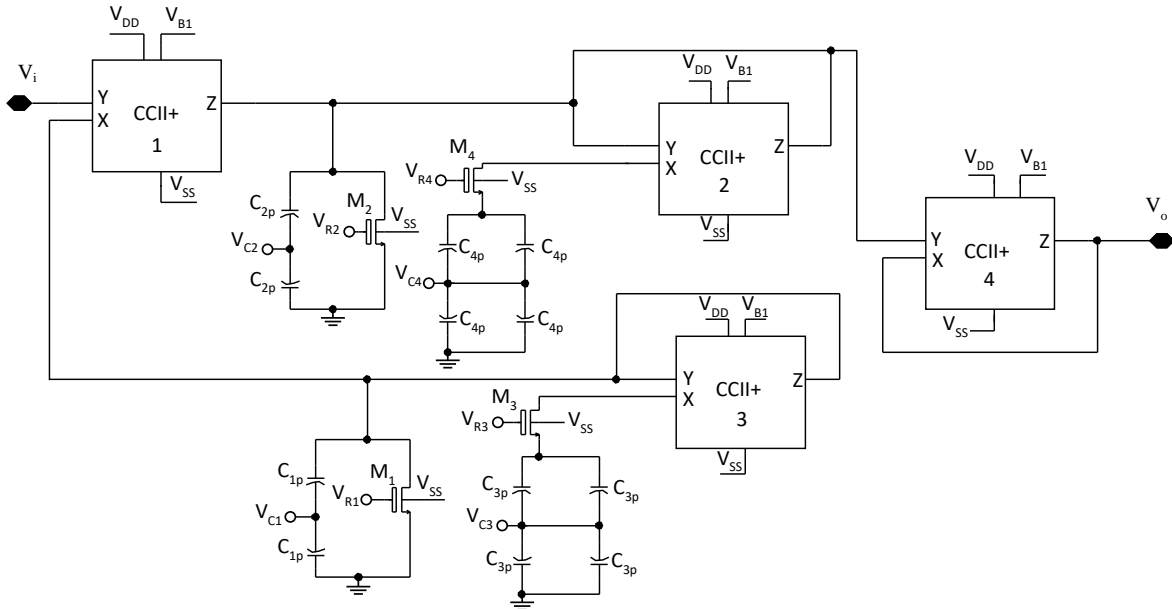

Figure 5.5. Circuit schematic of the proposed second-order all-pass network with post-production tunability compensation.

An additional voltage buffer stage (4th CCII+) is added to make the circuit capable of driving a 50 Ω load impedance, as is required for the VNA measurement. A high-precision CCII capable of driving a 50 Ω load as described later in Section 5.3 [87] is used for this purpose.

After initial circuit synthesis and layout, a simulation of the proposed all-pass network is performed using accurate non-ideal device models from the ams AG foundry as well as extracted layout RC parasitics, as shown in Figure 5.6. To compensate for the resulting change in circuit response, bias voltages are optimized *for the nominal corner* (in simulation) using a global optimizer resulting in the solid black curve, as indicated. The optimized bias voltages are $V_{R1} = 2.009$ V, $V_{R2} = 1.976$ V, $V_{R3} = 2.052$ V, $V_{R4} = 1.925$ V, $V_{C1} = 0.027$ V, $V_{C2} = 0.129$ V, $V_{C3} = 0.156$ V, $V_{C4} = 0.399$ V, $V_{B1} = -0.020$ V. These values will serve as the initial values for the automated tuning procedure during measurement. A corner analysis is also performed as indicated by the legend in Figure 5.6(a – b). The temperature is further swept from $0 – 85$ ^{0}C for each corner (as indicated by the dotted gray curves). This serves to

illustrate the large variation in response characteristics that can be expected without post-production tuning. It is important to note that in this work only the room temperature case is considered when comparing simulated and measured responses and the temperature sweep does not serve to explain discrepancies. The temperature sweep should therefore be seen as additional information, which could be useful to the reader.

A Monte Carlo simulation of the full circuit, analyzing $\Delta|H|$ and the average magnitude over the passband ($\overline{|H|}$) over 500 samples, indicates that a magnitude flatness of 5.6 dB is most likely, which further justifies the proposed post-production tuning approach.
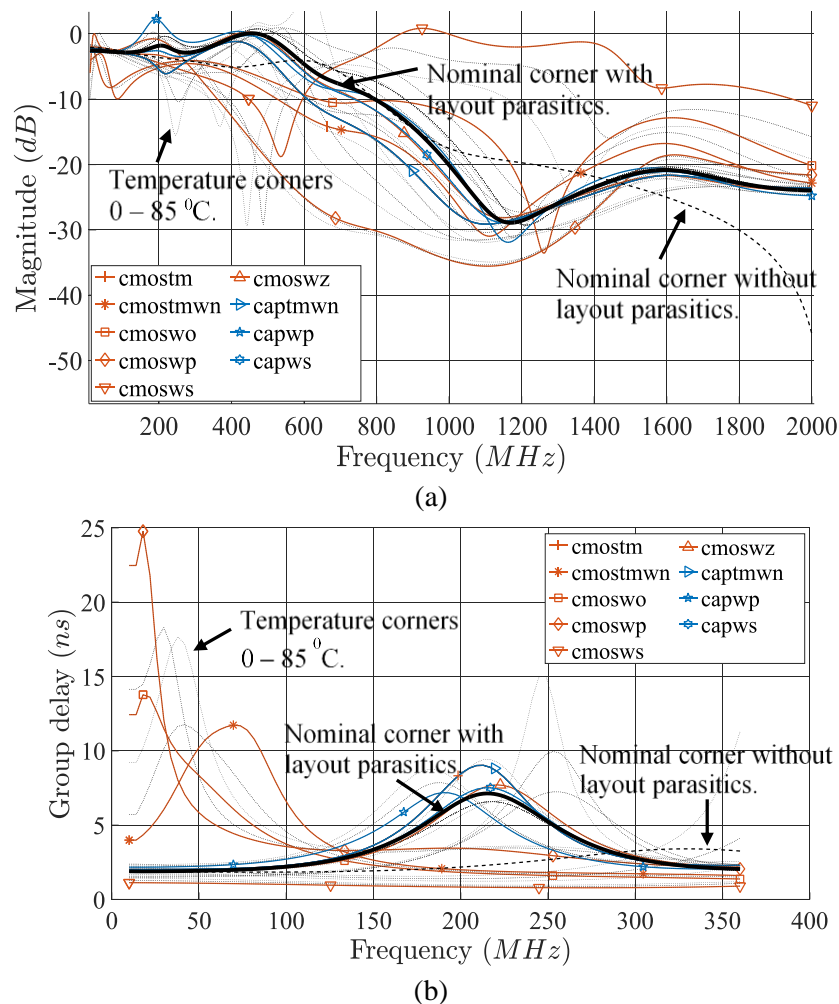

Figure 5.6. Corner simulations of the proposed second-order all-pass network.
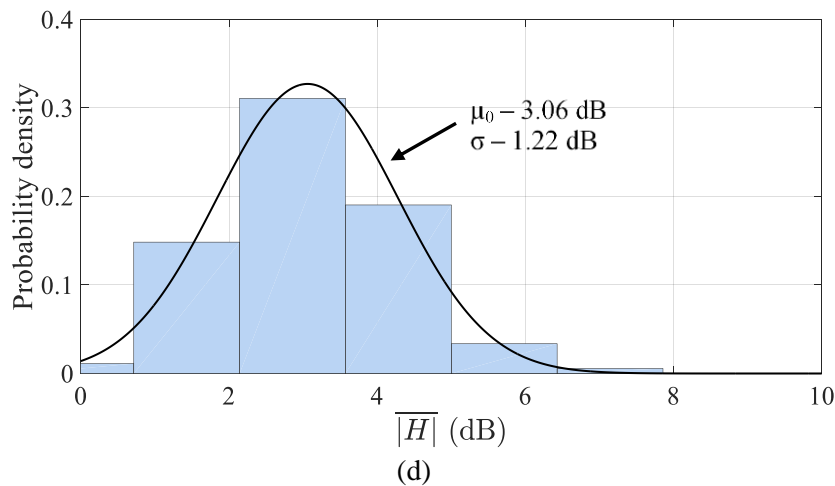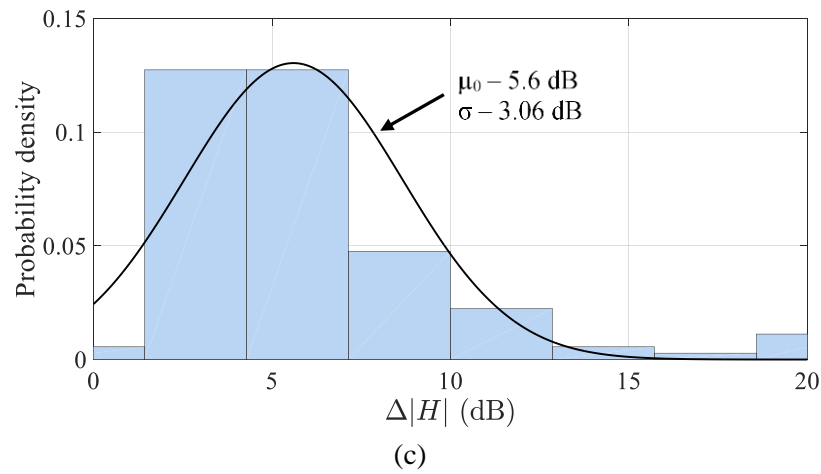
(c)

(d)

Figure 5.6 (cont.). Monte Carlo simulations of the proposed second-order all-pass network.

A description of the corner designations used in the legend of Figure 5.6 is given in Table 5.4 below.

Table 5.4. Process corner designations used in the ams AG 0.35 µm CMOS technology node.

Corner designation	Description
cmostm	Typical mean (nominal)
cmostmwn, captmwn	Typical worst noise
cmoswo	Worst case one (fast NMOS and slow PMOS)
cmoswp, capwp	Worst case power (fast NMOS and fast PMOS)
cmosws, capws	Worst case speed (slow NMOS and slow PMOS)
cmoswz	Worst case zero (slow NMOS and fast PMOS)

The power consumption of the second-order all-pass network is simulated to be 37 mW (15 mW without the voltage buffer). A noise simulation is also performed, as shown in Figure 5.7, indicating a nominal output noise of 22 nV/√Hz, peaking at 62 nV/√Hz at resonance.
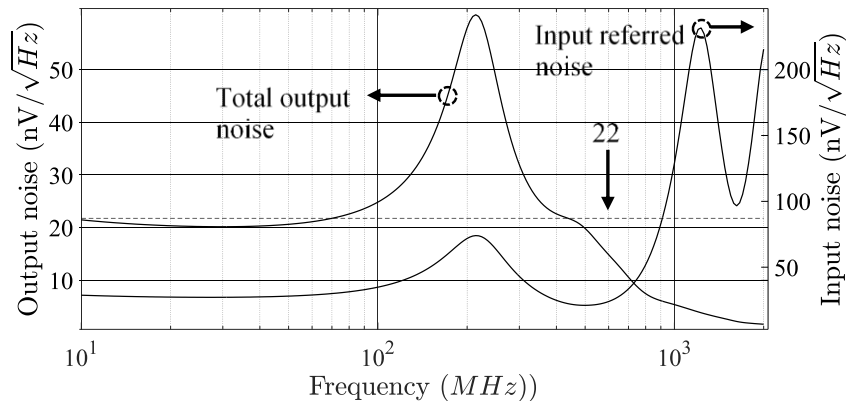

Figure 5.7. Noise simulation of the all-pass network.

5.2.4   Measurement Results and post-production automated tuning

The second-order all-pass network of Figure 5.5 is manufactured using the C35B4 0.35 µm CMOS process from ams AG, as shown in the micrograph of Figure 5.8. The camera ready-layout submitted for manufacturing is shown in Figure A.1 - Figure A.6.
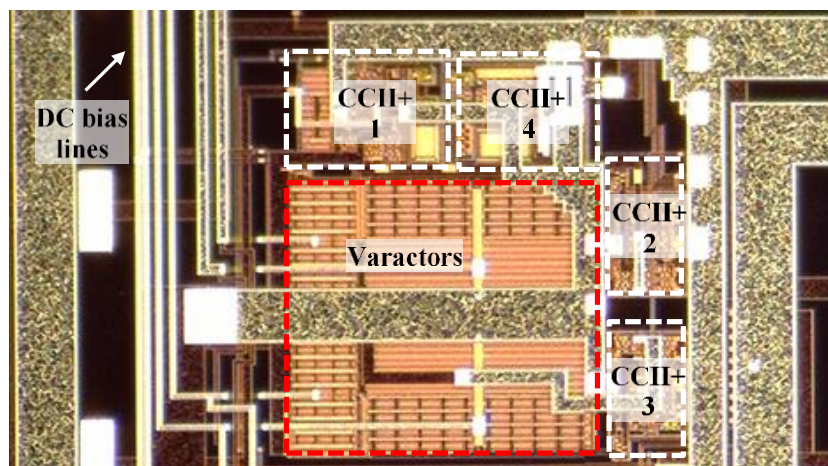
Figure 5.8. Micrograph showing the top view of the all-pass network.

A PCB is also designed to house the IC and supply the necessary bias voltages and RF test signals, as shown in Figure 5.9. Bondwires are used to connect the IC to the PCB pads. Power supply and DC bias voltages are supplied by the IDC connector as shown.

(a)

Figure 5.9. (a) Layout of test PCB for housing and testing the proposed all-pass network.

(b)

Figure 5.9 (cont.). (b) Test PCB for housing and testing the proposed all-pass network.

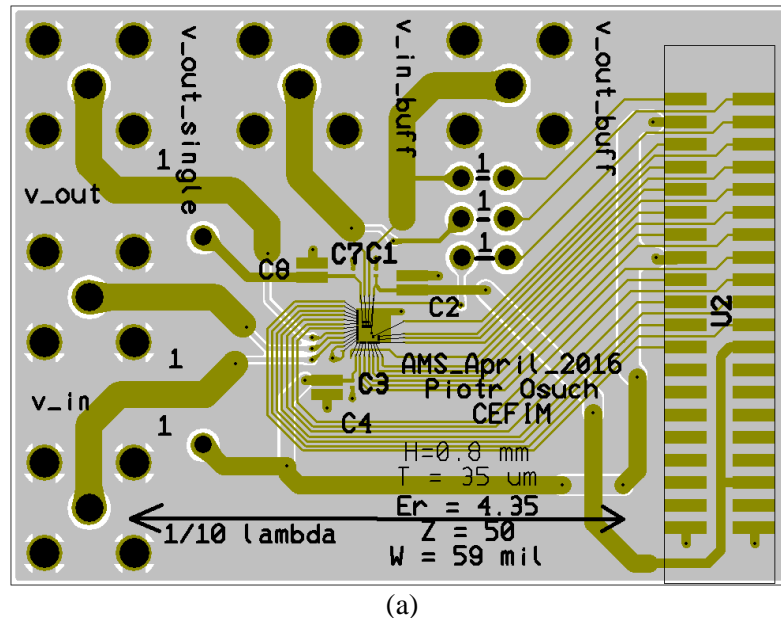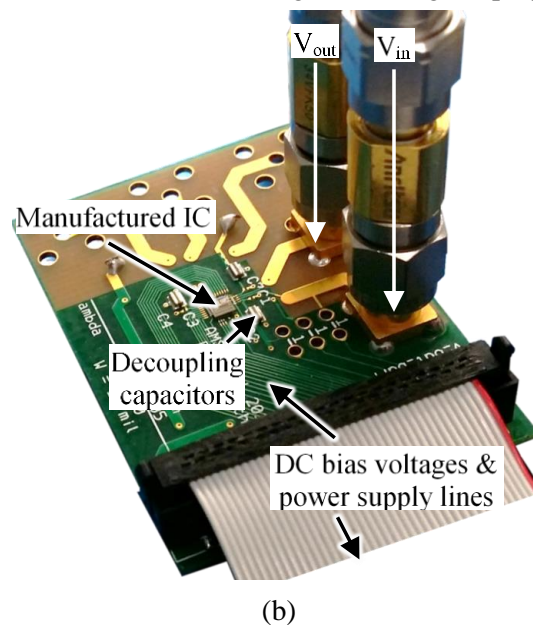The V_{in} and V_{out} terminals are routed to SMA connectors, which are connected to an Anritsu MS4640A VNA for measurements. To ensure wideband operation, decoupling capacitors are placed as close to the IC as possible (less than $\lambda/10$ at 500 MHz).

The measurement and post-production automated tuning setup is shown in Figure 5.10.

Figure 5.10. Post-production measurement and genetic optimization setup.

The pre-calibrated VNA captures 1001-point two-port S-parameter data and sends it to a PC (through a USB interface) running a genetic optimizer algorithm in MATLAB®. A genetic optimizer is chosen because the objective function has multiple local minima and the number of optimization variables is large. The optimizer code extracts the transmission magnitude and group delay responses from the S-parameters, and calculates the required bias voltage settings for the next measurement iteration. These values are programmed to an Analog Devices AD5370 DAC card connected to the PC via USB, and then applied to the DUT. Only the 8 most significant bits of the 16-bit DAC card are used. The measurement setup is placed in a thermally stable room to prevent DAC voltage drift. Initial tuning values as found in Section 5.2.3.2 for the nominal corner are used in the optimization algorithm. A 500 MHz bandwidth digital oscilloscope is also used to monitor the input and output excitations.

The metrics used to calculate the overall objective function are the group delay similarity, τ_s and $\Delta|H|$, as defined earlier as well as the deviation from the desired ω_0 and τ, $\Delta\omega_0$ and $\Delta\tau$ respectively. The overall objective function is then a simple linear superposition of these components.

Measured results are shown in Figure 5.11 versus the genetic optimizer iteration. As expected from the corner simulations, the initial response is far from the desired all-pass response with a magnitude variation of ~15 dB. After running the optimizer for 4500 iterations, the magnitude variation is reduced to 3.1 dB with a -3 dB voltage transfer bandwidth of 280 MHz. The optimized group delay $f_0 = 73$ MHz and $\Delta\tau = 10$ ns. From (2.2) $Q_D = \frac{10 \, ns \, .2\pi \, .73 \, MHz}{4} = 1.15$. The discrepancy in the optimized values and the initial design goals (Section 5.2.3.2) is caused by process variations as discussed next. It is worth noting that nine bias voltages are optimized which means that an optimal solution is found with 2.5 sweep points per optimized variable. The speed of the optimization can further be improved; however, this is not a key consideration in this work.

Finally, in an effort to compare measured results with simulations the simulation corner which best corresponds to the measured responses is identified as shown in Figure 5.11.
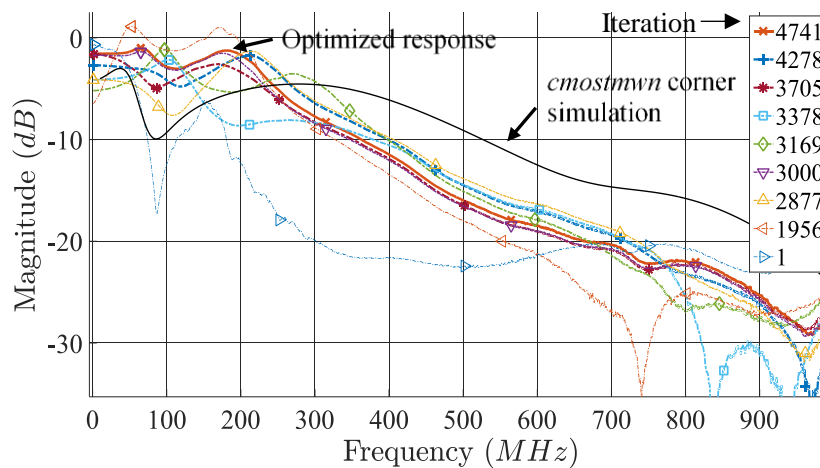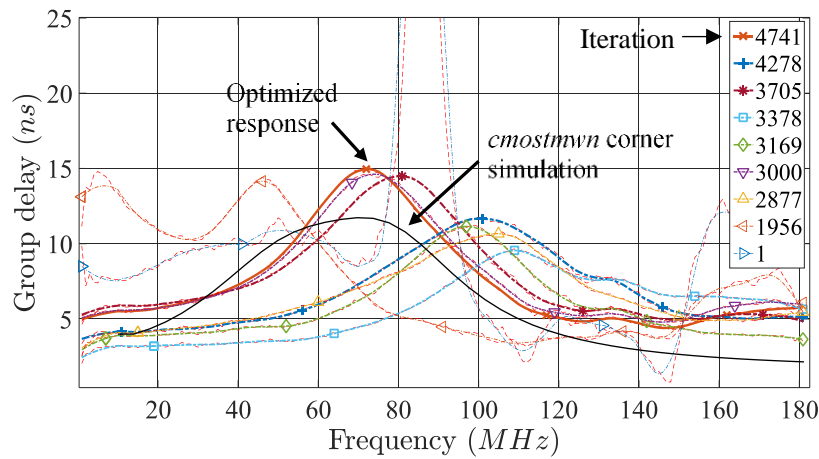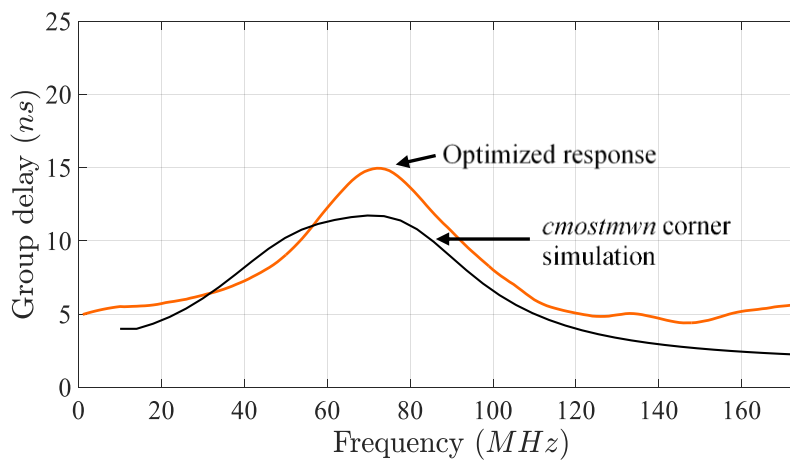

(a)
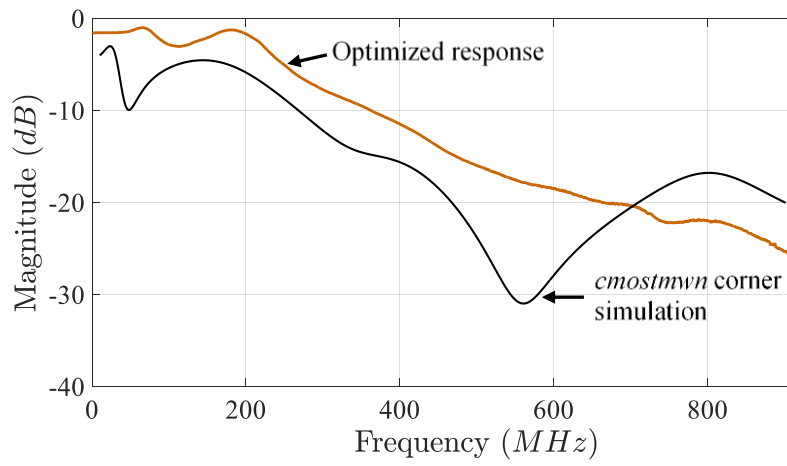
(b)

Figure 5.11. (a – b) Measured voltage transfer curve of the proposed CCII.

(c)

(d)

Figure 5.11 (cont.). (c – d) Measured and simulated group delay curves.
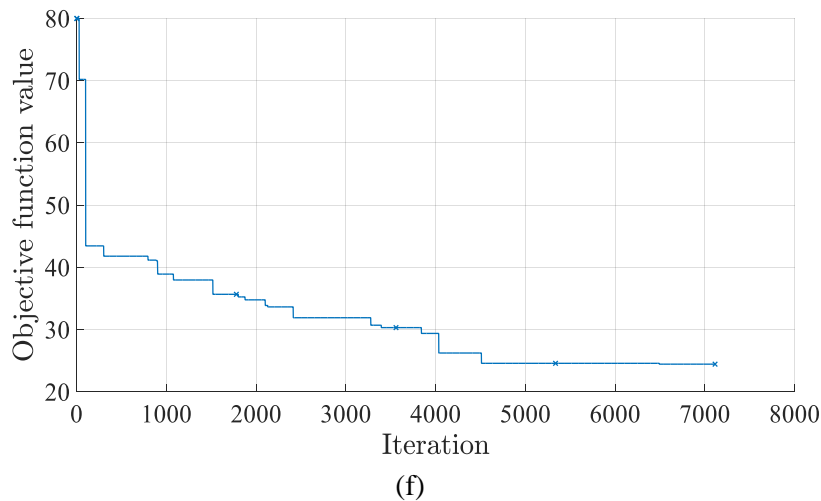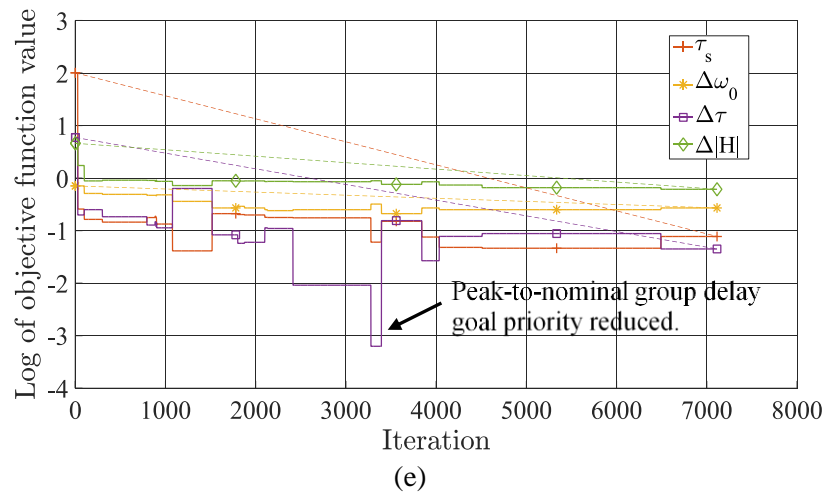
(e)

(f)

Figure 5.11 (cont.). (e) Various metrics used in the calculation of the objective function versus the iteration, (f) overall objective function versus the iteration.

This is not a trivial task since the simulated response is fine-tuned for the nominal corner whereas the measured results are fine-tuned for an unknown corner. Therefore, the most-probable corner is identified by choosing a corner simulation which best matches all the measured results in the optimization flow. The alternative approach of first fine-tuning each corner case in simulation and then comparing fine-tuned corners to the fine-tuned measured result is not chosen for two reasons. Firstly, the high degree of tunability of the design would allow the optimizer to achieve similar responses for each corner, making a meaningful comparison impossible. Secondly, this approach defeats the purpose of the corner analysis which is simply intended to gauge the spread in expected results as well as the need for post-production tunability and not to *predict* the measured response. This particular IC falls

roughly into the *cmostmwn* corner (as shown in Figure 5.11) – resulting in a much lower bandwidth than predicted by nominal operating conditions.
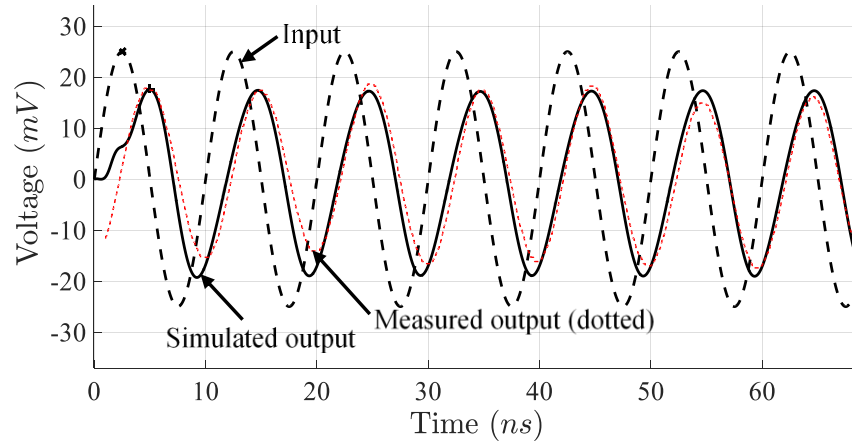

Figure 5.12. Measured input versus output sinusoid signals at 100 MHz.

A time-domain measurement of the optimized circuit's output for an input sinusoid of 100 MHz is further shown in Figure 5.12 (filtered with a digital low-pass filter to remove noise). The observed attenuation corresponds to the attenuation seen in the measured (as well as simulated) magnitude responses shown earlier. The input to the all-pass network must satisfy the small-signal condition to ensure that all active devices operate within permissible ranges. The input signal level is limited by the NMOS resistors (to ensure operation in the triode region) as well as the accumulation-mode capacitors, which experience non-linear effects if the input signal is large.

5.2.5   Conclusion

A novel on-chip active second-order all-pass network is proposed, with post-production tunability to account for CCII non-idealities as well as CMOS process tolerances. The method is implemented in a 0.35 µm CMOS prototype design, and subjected to automated post-production tuning using a genetic local optimizer to realize a practical all-pass network. This represents the first measured results of an active inductorless on-chip second-order all-pass network with a Q_D-value larger than 1 and therefore the first all-pass network suitable for implementing ASPs on chip, as described earlier in Chapter 2. This benefit of the

proposed design is contrasted with other performance metrics and compared to existing implementations in Table 5.5. The centre frequency of the second-order all-pass delay function is denoted by f_0 and the passband magnitude variation by $\Delta|H|$.

Table 5.5. Comparison with published measured work.

Ref.	Order	Q_D	f_0 (GHz)	-3 dB BW (GHz)	Technology	# of L	Size (mm^2)	Power (mW)	$\Delta|H|$ (dB)
This work[*]	2nd	1.15	0.073	0.280	0.35 µm CMOS	0	0.0625	15	3.1
[79]	2nd	~0.19 (0.59)[3]	3	4	0.25 µm CMOS	0	0.085	30	1.5 (> 25)
[63]	2nd	~0.08 (0.52)	6	13	0.13 µm CMOS	1	0.0627	18.5	0.5 (> 13)
[80]	2nd	~0.071	6.5	16.5	0.09 µm CMOS	0	-	-	-
[126]	2nd	~0.049 (0.61)	6.3	12	0.13 µm CMOS	1	-	16.5	1.5 (> 10)
[127]	2nd	~0.047	6	7.5	SiGe BiCMOS HBT ($f_T = 95$ GHz)	1	0.49[1]	121	~ 1
[81]	2nd ($f_0 = 0$)[2]	0	0	12.2	0.16 µm CMOS	0	0.15	450	1.4
[128]	0th ($f_0 = 0$)[2]	0	0	10	SiGeRF HBT ($f_T = 80$ GHz)	2	0.4197	38.8	2 – 2.5
[82]	2nd ($f_0 = 0$)[2]	0	0	4.38	0.18 µm CMOS	0	0.0512	16.79	-
[83]	2nd ($f_0 = 0$)[2]	0	0	> 3	0.13 µm CMOS	0	0.29	112	~ 0.75

[*] - Post-production tunable, [1] - Including pads, [2] - constant delay with frequency, [3] - values in brackets are computed over the entire bandwidth with the associated magnitude variation also shown in brackets.

5.3 SECOND GENERATION CURRENT CONVEYOR WITH CONTROLLABLE FEEDBACK

As indicated in Section 5.2.3.2, a high-precision CCII was used in the design of the proposed second-order all-pass network as a voltage buffer stage. In principle this CCII could have been used throughout the all-pass network to improve precision and therefore reduce passband peak to nominal variation for larger Q_D-values. However, this comes at the expense of lower bandwidth. In this work the hybrid approach of using CCIIs with varying levels of precision is demonstrated instead, with lower precision CCII's compensated for in the

second-order all-pass network and the output buffer explicitly compensated for through internal feedback control. The high-precision CCII is described in this Section and is itself a versatile analogue processing block which can find various applications in future high-speed ASPs.

5.3.1   Introduction

In this Section, a practical CCII design methodology is proposed incorporating explicit multi-loop stability analysis with accurate device models, process variation, and layout parasitics. An optimization based synthesis technique is presented and implemented in 0.35 µm CMOS, improving on the state-of-the-art CCII+ in [98] by optimizing transistor aspect ratios and introducing a post-production tunable RC compensation feedback network to reduce gain peaking and to tune the phase margin in order to ensure stability despite process variation. A measured operating bandwidth exceeding 500 MHz, a measured transfer error lower than 1% and a simulated R_X lower than 5 Ω is achieved. A multi-loop feedback analysis methodology is applied, which is based on the single-loop true return ratio approach [59], as well as multi-loop feedback theory [60]. To the author's knowledge this is the first time that the true return ratio approach is combined with multi-loop feedback theory and applied to the design of CCIIs. To verify the efficacy of the approach, the stability analysis is applied to the high-precision CCII+ proposed in [94] with two feedback loops, and it is shown (both through analysis and experimental verification) that omitting the stability analysis can lead to oscillation. The proposed multi-loop stability analysis is compatible with numerical simulation methods and can be used in the analysis of future high speed CCII designs.

This Section is structured as follows:

1. A multi-loop cut-based numerical feedback analysis technique is applied, combining single-loop [59] and multi-loop [60], [61] feedback analysis in a practical microelectronic CCII design, which is a novel contribution to the SOTA.

2. An improved approach to designing high-precision CCII+ circuits (based on the circuit proposed in [98]) is presented, and a post-production tunable feedback network implemented using a practical design procedure, optimizing bandwidth while at the same time preserving precision, a low R_X, high R_Z and R_Y. The proposed post-production tunable feedback network is analyzed using the proposed feedback analysis method.

3. The high-precision design in [94] is analyzed for stability and implemented experimentally. It is clearly demonstrated, through measurement results, that omission of stability analysis can lead to oscillation.

The proposed design methodology incorporating multi-loop stability analysis is verified by measured results of a CCII with post-production tunable feedback control. Prototypes are manufactured in the 0.35 µm CMOS technology node from ams AG.

5.3.2 Multi-loop feedback stability considerations in analogue circuits

Existing single- and multi-loop loop analysis techniques were summarized in Section 2.5. In this work, the single-loop cut approach is extended to multi-loop systems using Mason's multi-loop feedback theory [60]. As an example, the CCII+ in [94] can be represented by the signal flow graph in Figure 5.13. The original circuit of [94] is shown later in Figure 5.23 with nodes numbered according to Figure 5.13. All feedback loops, as well as the source and sink, are indicated. Mason has shown that the denominator of the transfer gain (denoted by Δ) is given by [60]

$$\Delta = 1 - \sum_m P_{m1} + \sum_m P_{m2} - \sum_m P_{m3} + ..., \tag{5.12}$$

where P_{mr} is the gain product of the mth possible combination of r non-touching loops where $r > 0$. For example, in Figure 5.13, 10 feedback loops can be identified as:

$$
\begin{aligned}
T_1 &= bc & T_2 &= ed \\
T_3 &= fg & T_4 &= jk \\
T_5 &= hi & T_6 &= gec \\
T_7 &= kic & T_8 &= dfb \\
T_9 &= hjb & T_{10} &= fkid \\
T_{11} &= jgeh
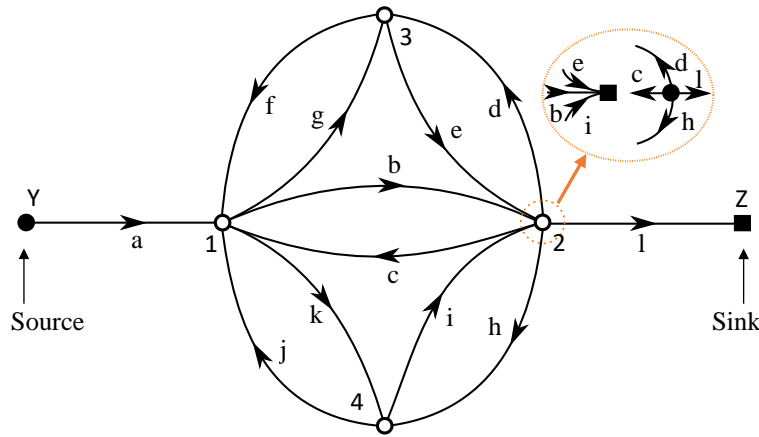\end{aligned}
\tag{5.13}
$$

Figure 5.13. A possible signal flow graph representation of the CCII in [94].

From (5.12):

$$
\Delta = 1 - \sum_{i=1}^{11} T_i + (T_2 T_4 + T_3 T_5)\,,
\tag{5.14}
$$

which can be re-written in the factorized form [60]

$$
\Delta = \prod_n 1 - T_n{'}\,,
\tag{5.15}
$$

where $T_n{'}$ is the loop gain of the nth loop with all lower-numbered loops ($T_1 \dots T_{n-1}$) split.

To illustrate the practical implication of (5.15) node 2 is chosen in Figure 5.13 as the starting node in the analysis. The node is split (or cut) by subdividing it into source and sink nodes (as shown in the figure inset) and the loop gain is computed. Next, leaving the previous loop split, the next loop is found and its gain computed. In this analysis, only node 1 is not affected

(has a non-zero gain) after the previous split of node 2. For instance, computing the gain of the feedback loop from node 3 (by splitting it into a source and sink, as above) after node 1 has been cut will result in a zero gain. Therefore, for this example:

$$\Delta = (1 - T_1')(1 - T_2') = 1 - T_1' - T_2' + T_1'T_2',$$
(5.16)

which has the same form as (5.14) as expected. In general, $T_n' \neq T_n$.

Equation (5.15) is therefore compatible with the aforementioned double-injection cut technique and Δ can be obtained with relative ease, even for complicated cases (in the above example only two cuts were sufficient to isolate 10 loops). It is also compatible with existing numerical approaches, where parts of the signal flow graph might be black box models where the feedback structure is unknown (such as device parasitics of transistor models). Moreover, if a loop or part of a loop is considered twice (which is particularly likely in the black-box scenario) then the computed gain is simply 0 and from (5.15) results in multiplication of Δ by unity.

Finally, the stability of the network can be determined by solving for the poles of Δ. Alternatively, the effective open-loop gain can be found and a Nyquist plot constructed by:

$$\Delta = 1 + A(s)F(s)$$
$$A(s)F(s) = \Delta - 1,$$
(5.17)

where A and F are the effective forward and feedback loop gains.

The proposed approach for analyzing multi-loop feedback stability therefore consists of the following steps:

1. Identify *any* potential feedback loop and introduce a break-point anywhere in the loop. Choosing loops which touch other loops is desirable as this reduces the analysis time.

2. Apply the single loop feedback double injection analysis technique to calculate the return ratio T_n'.

3. Identify any remaining feedback loop that still exists with the previously identified loop(s) cut.

4. Repeat steps 2 and 3 until no more feedback loops can be found.

5. Calculate Δ using (5.15) and the effective open-loop gain using (5.17) if desired.

Having formulated a multi-loop feedback analysis technique compatible with the envisaged numerical circuit design practices, the design and analysis of CCIIs can proceed.

5.3.3   Practical implementation of a high-precision, high bandwidth CCII+

5.3.3.1   Theoretical design and simulations

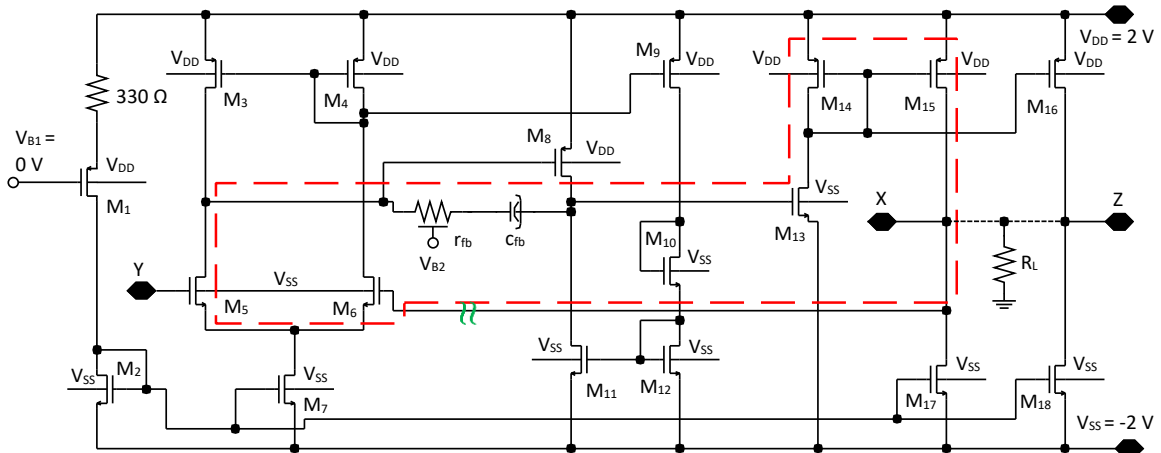The CCII+ circuit proposed in [98], shown in Figure 5.14, is used as the basis for the proposed design approach.

Figure 5.14. Circuit schematic of the improved high-precision, high-bandwidth and low R_X CCII+, based on the circuit proposed in [98].

The circuit input is a differential voltage follower stage (M_3-M_7) which mirrors the voltage from port Y to X. A simple current mirror then conveys the current from port X to Z. Transistor M_8 ensures that M_5 operates in saturation and together with M_9 ensures the same DC V_{DS} across M_5 and M_6 (thereby reducing the voltage following error). M_{13} ensures that

V_{DS} across M_8 is similar to that of M_9, further reducing any voltage difference between the two legs of the differential voltage mirror. A high gain (and, subsequently, narrow band) AC feedback path is formed by the common source pair M_8 and M_{13}. The remaining transistors act as biasing current sources. As a novel extension on [98], the fixed feedback is modified by adding a post-production tunable RC Miller network, with r_{fb} and c_{fb} selected to reduce this feedback gain and increase the bandwidth, at the expense of degrading precision and R_X. Finally, an *external* load impedance R_L of 50 Ω terminates both ports X and Z (indicated by the dotted line in Figure 5.14), to represent either external test equipment or a subsequent SoC stage. This is dictated by the available two-port measurement setup, as detailed later. The proposed method is; however, more general, and ports X and Z are treated separately in the analytical analysis and simulations, to prevent loss of generality. The driving source is assumed to have negligible output impedance, as is common in CCII design.

To further increase the bandwidth and control the resulting tradeoffs, a numerical optimization-based design methodology is implemented, as is sometimes employed in literature [129]. Design Equations (5.18) - (5.22) are derived to serve as a basis for finding initial design values as well as to guide the numerical optimizations. These equations do not consider parasitic effects, which will only be accounted for in later simulation (using accurate device models supplied by the foundry) during the optimization stage. The following design choices are made in the proposed circuit: $R_L = 50\ \Omega$, $g_{m,vm} = g_{m,5} = g_{m,6}$, $g_{mb,vm} = g_{mb,5} = g_{mb,6}$, $g_{m,m} = g_{m,14} = g_{m,15} = g_{m,16}$, $g_{m,fb} = g_{m,13}$, $g_{m,b2} = g_{m,17} = g_{m,18}$, $g_{o,b2} = g_{o,17} = g_{o,18}$, $g_{o,b1} = g_{o,7}$, $g_{o,vm} = g_{o,5} = g_{o,6}$, $g_{o,m} = g_{o,14} = g_{o,15} = g_{o,16}$, $g_{o,fb} = g_{o,13}$, $c_{gs,vm} = c_{gs,5} = c_{gs,6}$. Here g_m, g_{mb} and g_o represent the transconductance, bulk transconductance and output conductance of the transistors, respectively. C_{gs} represents the gate-source capacitance. Using the CCII definitions from (2.6) and definitions from Section 2.4.1 it then follows that:

$$A_v\big|_{f=0} \approx \frac{R_L g_{m,8} g_{m,fb} g_{m,m} g_{m,vm}}{\left[\begin{array}{l} g_{o,8} g_{o,vm}\left(g_{m.m} + g_{o,fb}\right)\left(\dfrac{1}{2} + \left(g_{o,b2} + g_{o,m}\right)R_L\right) \\ + R_L g_{m,8} g_{m,m} g_{m,fb} g_{m,vm} \end{array}\right]} \ , \tag{5.18}$$

$$A_i\big|_{f=0} \approx \frac{g_{m,8}g_{m,fb}g_{m,m}g_{m,vm}}{\left(R_L g_{o,b2} + R_L g_{o,m} + 1\right)\begin{pmatrix} g_{o,8}g_{o,vm}\left(g_{o,b2} + g_{o,m}\right) \\ \left(g_{m,m} + g_{o,fb}\right) + \\ g_{m,8}g_{m,m}g_{m,fb}g_{m,vm} \end{pmatrix}} , \tag{5.19}$$

$$Z_X \approx \frac{s \cdot c_{fb}\left(g_{m,8} + g_{o,vm}\right)\left(g_{m,vm} + g_{o,vm}\right)}{g_{m,fb}g_{m,vm}\left(s \cdot \left(r_{fb}c_{fb}g_{m,8} - c_{fb}\right) + g_{m,8}\right)\left(g_{m,vm} + g_{o,vm}\right)}, \tag{5.20}$$

$$Z_Y \approx \frac{s \cdot 4c_{gs,vm} + 2\left(2g_{m,vm} + \left(2g_{mb,vm} + g_{o,b1}\right)\right)}{s \cdot c_{gs,vm}\left(2g_{mb,vm} + g_{o,b1}\right)}, \tag{5.21}$$

$$R_Z \approx \frac{1}{g_{o,b2} + g_{o,m}} . \tag{5.22}$$

Based on these design equations, the following parametric choices are important:

1. To reduce the transfer error, R_X and increase R_Z, the parameters $g_{o,vm}$, $g_{o,8}$, $g_{o,m}$, $g_{o,b2}$, $g_{o,fb}$ must be minimized.

2. To increase Z_Y at higher frequencies, $g_{b,vm}$ and $g_{o,b1}$ must be minimized.

3. To reduce R_X and also minimize the transfer error (both in α and β), $g_{m,fb}$, $g_{m,m}$, $g_{m,vm}$ are maximized.

Initial design values are chosen with these considerations in mind, as shown in Table 5.6.

Table 5.6. Initial transistor sizes chosen for devices in Figure 5.14.

Device	W/L (µm/µm)	Device	W/L (µm/µm)	Device	W/L (µm/µm)
M1	5/0.5	M7	5/0.35	M13	30/0.35
M2	5/0.35	M8	5/0.35	M14	20/0.35
M3	5/0.35	M9	5/0.35	M15	20/0.35
M4	5/0.35	M10	5/0.35	M16	20/0.35
M5	20/0.35	M11	5/0.35	M17	10/0.35
M6	20/0.35	M12	5/0.35	M18	10/0.35

These choices are made for the typical corner. Initial values for r_{fb} and c_{fb} are set to 1 kΩ and 100 fF, respectively. The resulting transfer curves for $R_L = 50$ Ω and initial design values for the typical process corner are shown in Figure 5.15(a – b). A temperature sweep for each corner is also presented. It is important to note that in this work only the room temperature case is considered when comparing simulated and measured responses and the temperature sweep does not serve to explain discrepancies. The temperature sweep should therefore be seen as additional information, which could be useful to the reader. A description of the corner designations is shown in Table 5.4. Accurate Foundry PDK device models are used in the simulation, and not the simplified models of the analytical analysis in (5.18) – (5.22).
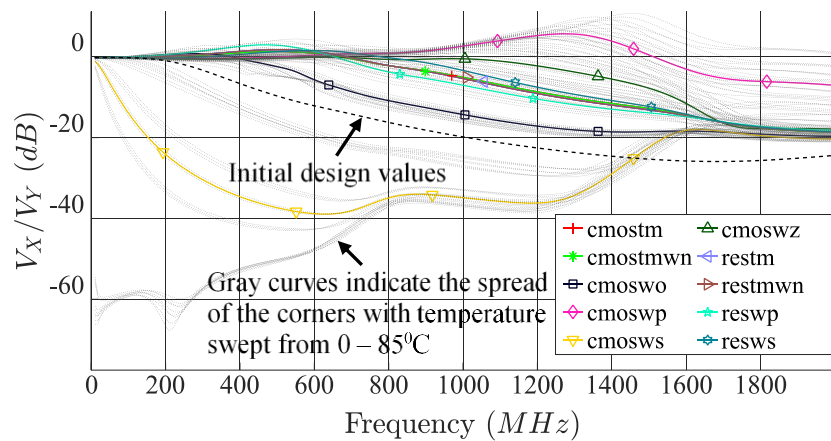
(a)
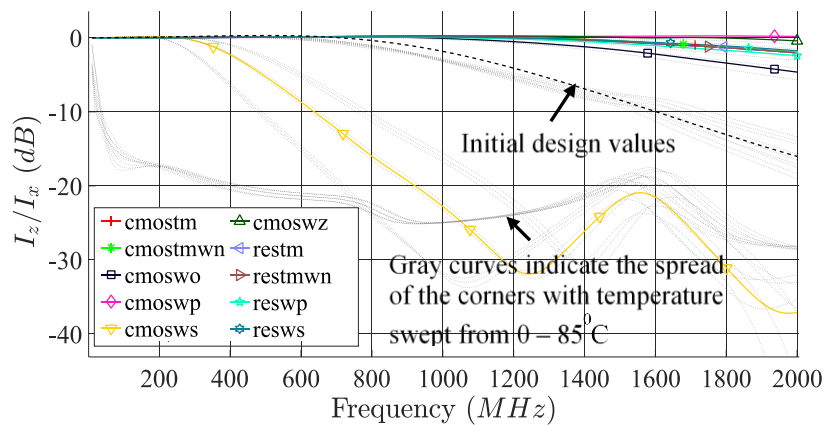
(b)

Figure 5.15. (a) Simulated voltage transfer curves between ports Y and X, and (b) simulated current transfer curves between ports X and Z for various process corners (fixed R_L of 50 Ω).
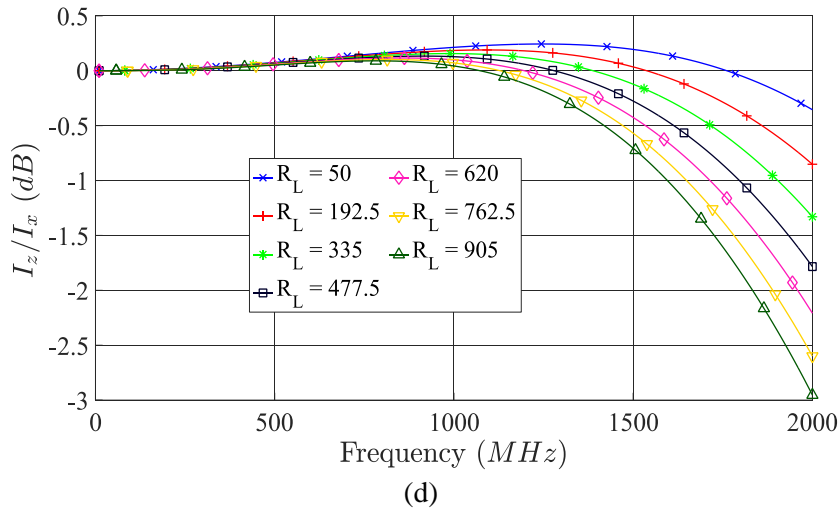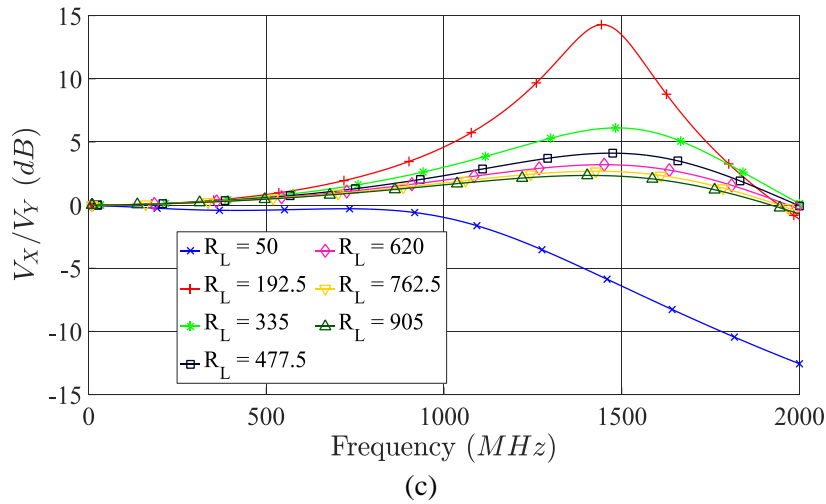
(c)

(d)

Figure 5.15 (cont.). (c) Simulated voltage transfer curves for changing R_L. (d) Simulated current transfer curves for changing R_L.

Next, a robust global optimization is run until the proposed design goals as shown in Table 5.7 are met.

Goal 1 sets an overvoltage requirement of at least 300 mV for each transistor to enforce linear transfer, which is the highest priority. Goal 2 requires transistors that form part of the RF path to be biased for optimal f_T , which corresponds to approximately 100 µA per µm gate width, as detailed in the process documentation. Goals 3 – 5 and 6 – 8 aim to reduce transfer errors and increase bandwidth. The -3 dB bandwidth is determined relative to the values of α and β at $f = 0$ Hz.

Table 5.7. Numerical optimizer goals.

Goal	Quantity	Requirement	Weight
1	$V_{OV(i)} = \left|V_{DS}(M_i)\right| - V_{DSAT}(M_i) \forall i$	$> 300m$	5
2	$I_{fT(i)} = \left|I_{DS}(M_i) - W_{Mi} \cdot 100\mu\right|$	$< 10\%$	2
	$i \in \{5,6,13,14,15,16\}$		
3	$\beta\big|_{f=0\ \text{Hz}} = \overline{(V_X / V_Y)}_{f=0\ \text{Hz}}$	$< 0.05\%$	1
4	$\beta_{BW} = BW_{-3dB}(V_X / V_Y)$	$> 700\ \text{MHz}$	2
5	$\beta_{ripp} = \Delta_{\max}(\left|V_X / V_Y\right|)$	$< 3\ \text{dB}$	1
6	$\alpha\big|_{f=0\ \text{Hz}} = \overline{(I_Z / I_X)}_{f=0\ \text{Hz}}$	$< 0.1\%$	1
7	$\alpha_{BW} = BW_{-3dB}(I_Z / I_X)$	$> 1\ \text{GHz}$	2
8	$\alpha_{ripp} = \Delta_{\max}(\left|I_Z / I_X\right|)$	$< 3\ \text{dB}$	1

After running the optimization, the resulting transfer curves are shown in Figure 5.15(a – b) for various process corners ($R_L = 50\ \Omega$), where the *cmostm*, *restm* pair refer to the nominal corner. Figure 5.15(c – d) illustrates the behavior of the CCII+ for other R_L loads. Including the load impedance in the optimization is critical for wideband applications, with high load impedance leading to wider voltage transfer bandwidth at the expense of slightly reduced current transfer bandwidth. In this example, the load impedance is restricted to 50 Ω, as required by the external test equipment interface.

Optimized design values are shown in Table 5.8, with optimal values for r_{fb} and c_{fb} found as 1.2 kΩ and 360 fF, respectively. The resistance value is used as a "middle ground" value later in the layout design, such that it can be tuned either way post-production.

Table 5.8. Optimized transistor sizes for the devices in Figure 5.14.

Device	W/L (µm/µm)	Device	W/L (µm/µm)	Device	W/L (µm/µm)
M1	12.8/0.5	M7	35/1.25	M13	50/0.5
M2	5/0.35	M8	15/0.35	M14	60/0.35
M3	10/0.35	M9	10/0.35	M15	60/0.35
M4	10/0.35	M10	15/0.35	M16	60/0.35
M5	45/0.35	M11	15/0.35	M17	5/0.35
M6	45/0.35	M12	15/0.35	M18	5/0.35

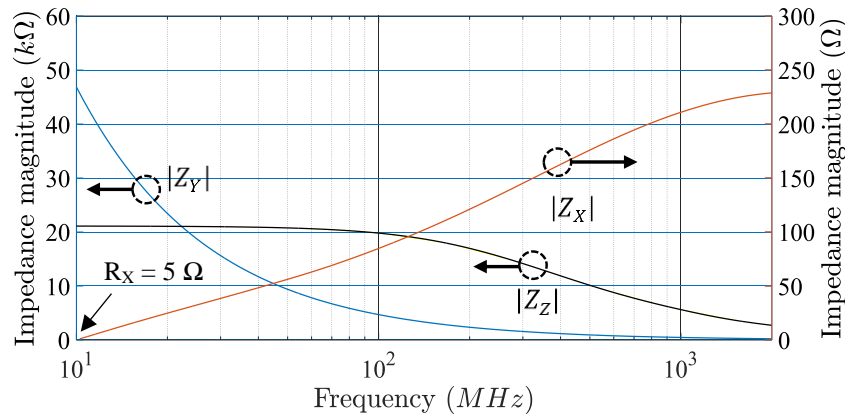The resulting impedance magnitudes at the various ports are further shown in Figure 5.16.

Figure 5.16. CCII+ port impedance magnitudes as a function of frequency.

Next, the stability of the optimized CCII+ is investigated and suitable values for r_{fb}, c_{fb}, which may be used by the post-production tuning mechanism, determined.

Applying the stability analysis procedure presented in Section 5.3.2, a single feedback path can be found that breaks all loops, as indicated by the red dotted line in Figure 5.14. This is verified by breaking other, smaller loops, and observing no change in the Nyquist plot, shown next. The loop is cut at the gate of M6 (as indicated in Figure 5.14) and the return ratio is calculated for $r_{fb} \in (100 : 1.5\text{k})\ \Omega$ and $c_{fb} = 360$ fF. This results in the Nyquist plot in Figure 5.17(a) and matching root-locus plot of the closed-loop gain in Figure 5.17(b), obtained independently by calculating the roots of the transfer function (not using Δ). This further confirms that the chosen loop does indeed break all feedback loops in this case. The indicated poles move into the RHP for $r_{fb} < 300\ \Omega$. Therefore, for $c_{fb} = 360$ fF, $r_{fb} > 300\ \Omega$ ensures a stable design, confirming that the proposed optimized design is indeed stable. This check can be incorporated into the numerical optimization procedure above as an additional high-priority goal (as opposed to running the check after the optimization), but is presented here separately for the sake of clarity. The effects of varying the load resistance are further investigated for $r_{fb} = 1.2$ kΩ in Figure 5.17(c – d) and show that the design is unstable for load impedances larger than 480 Ω. This again shows that high-precision CCIIs employing feedback control must be designed and optimized with the expected load in place to achieve stability.
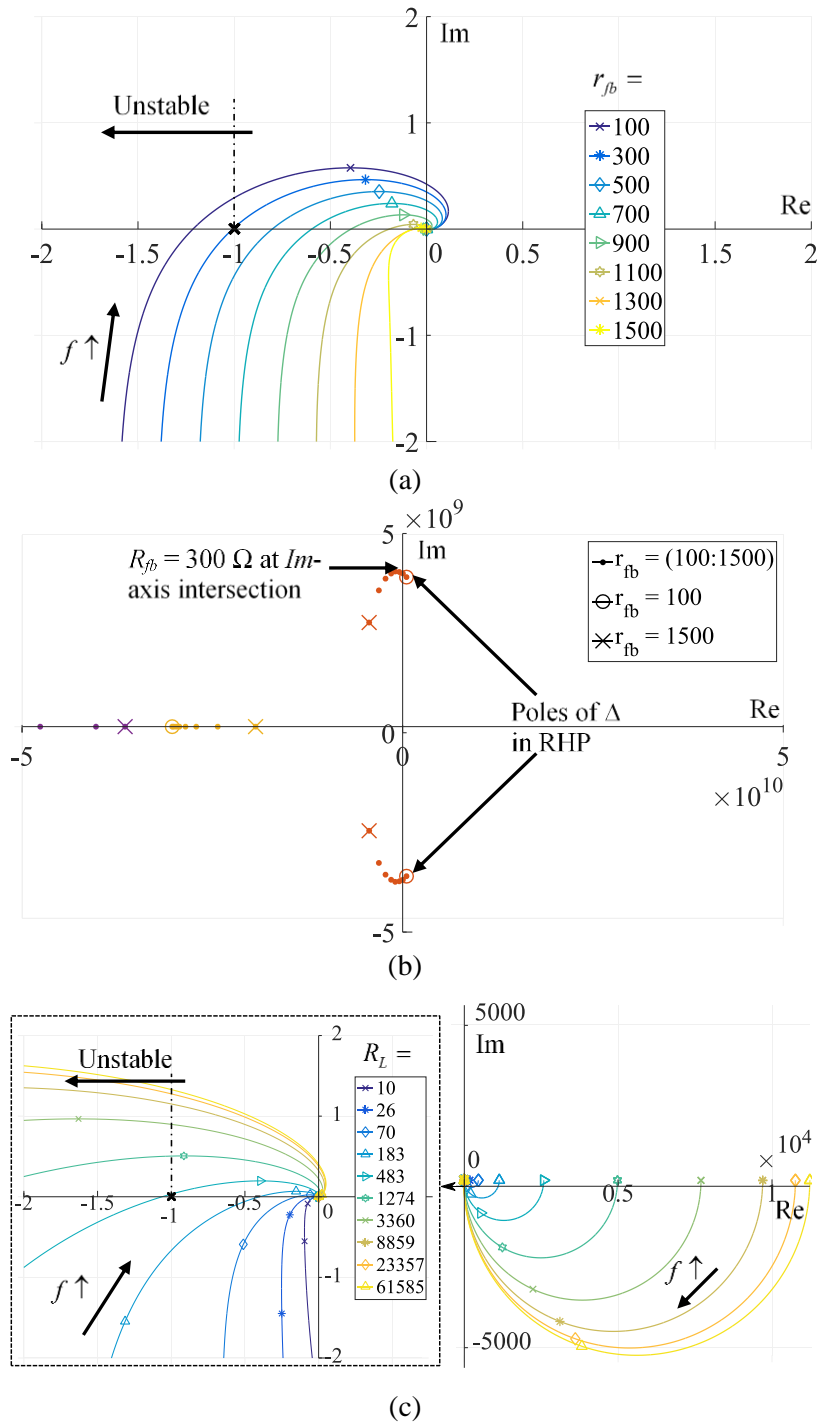
(a)

(b)

(c)

Figure 5.17. (a) Nyquist plot of the feedback loop indicated in Figure 5.14 for c_{fb} = 360 fF and R_L = 50 Ω for various values of r_{fb}. (b) Root-locus plot of the closed loop gain for varying r_{fb} and values set as in (a). (c) Nyquist plot of the feedback loop indicated in Figure 5.14 for c_{fb} = 360 fF, r_{fb} = 1.2 kΩ and varying R_L.
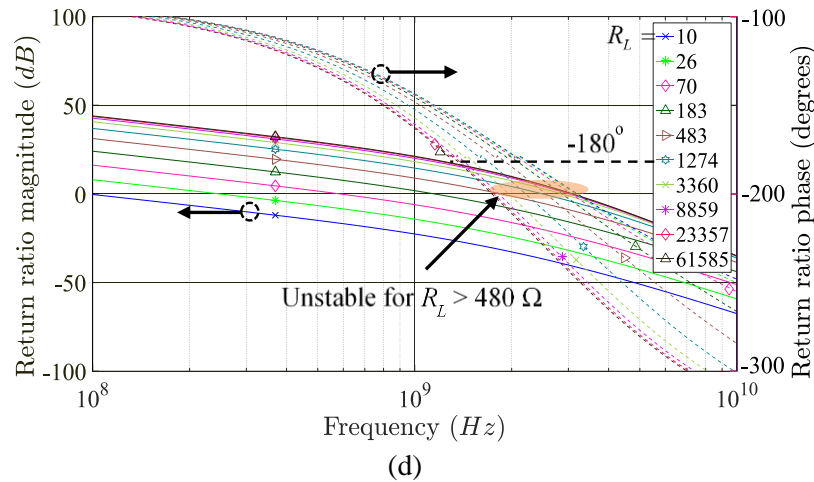
(d)

Figure 5.17 (cont.). (d) Bode diagram corresponding to (c).

Finally, it is also important to consider the effects of process tolerances on stability, as shown in Figure 5.18, where $R_L = 50 \ \Omega$.
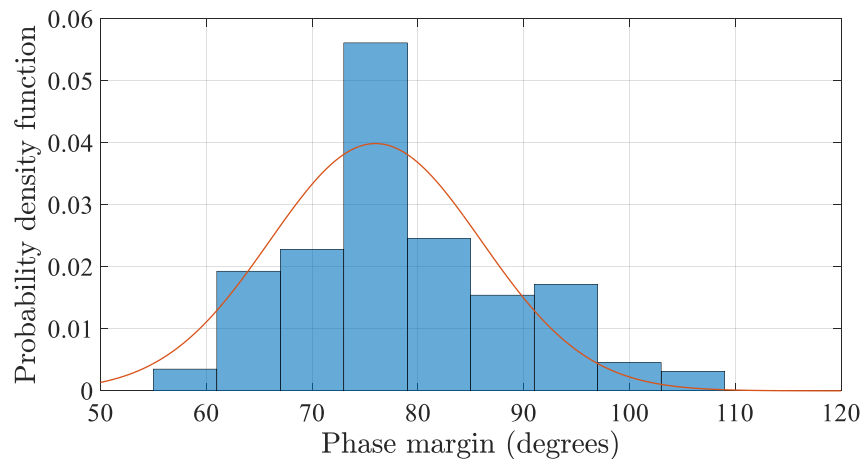

Figure 5.18. Probability density function of the phase margin for the CCII+ in Figure 5.14 across process design corners (Monte Carlo analysis).

Even though post-production tunability of the feedback gain is evidently not necessary in this case to ensure a stable design, it will be used to control peaking (thereby reducing passband ripple) as shown later. To this end, r_{fb} is implemented as an NMOS operating in the triode region with V_{B2} used to tune the effective channel resistance. The transistor aspect ratio is chosen such that, for $V_{B2} = V_{SS}$, the channel resistance is 300 Ω. By reducing V_{B2} the effective resistance can be tuned up to 2 kΩ.

5.3.3.2 Physical implementation and measured results

The proposed CCII+ is manufactured using the ams AG 0.35 µm CMOS process. A micrograph of the top-view is shown in Figure 5.19 (with the background removed for clarity). The feedback capacitor (c_{fb}) occupies a significant portion of the IC real estate. This illustrates the importance of using the Miller-effect in the proposed design, as the capacitor would be even larger without it. Power supply lines as well as bondpads are clearly visible.
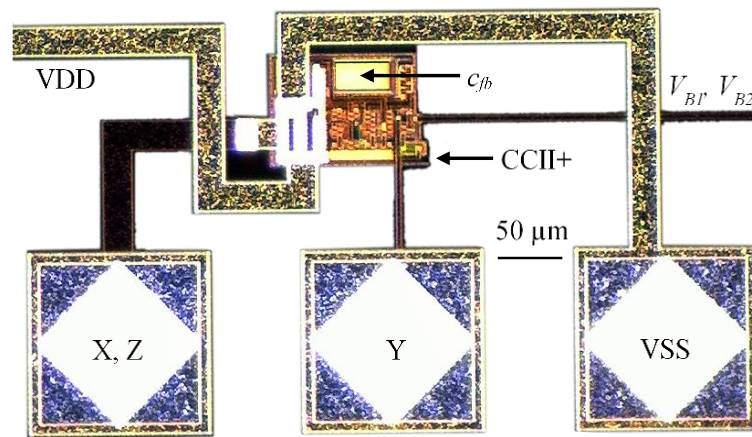

Figure 5.19. Micrograph showing the top view of the CCII+ of Figure 5.14.

A PCB is designed to house the IC and supply the necessary bias voltages and RF test signals, as shown in Figure 5.20. Bondwires are used to connect the IC to the PCB pads. Power supply and DC bias voltages are supplied by the IDC connector shown. The X, Z and Y terminals interface with SMA connectors which are connected to a VNA for measurements. As mentioned earlier in Section 5.3.3.1, the X and Z terminals of the CCII are shorted internally and therefore connected to only one SMA connector. To ensure wideband operation, decoupling capacitors are placed as close to the IC as possible (less than $\lambda/10$ at 1 GHz).

Measured results are shown in Figure 5.21. Including layout parasitics in the simulation (using an RC layout extraction) reduces the bandwidth from 1 GHz (in the nominal corner) to 850 MHz. Measured data indicates a further reduction of bandwidth to 500 MHz, which could indicate operation in the (*cmostmwn, restm*) process corner, as shown.
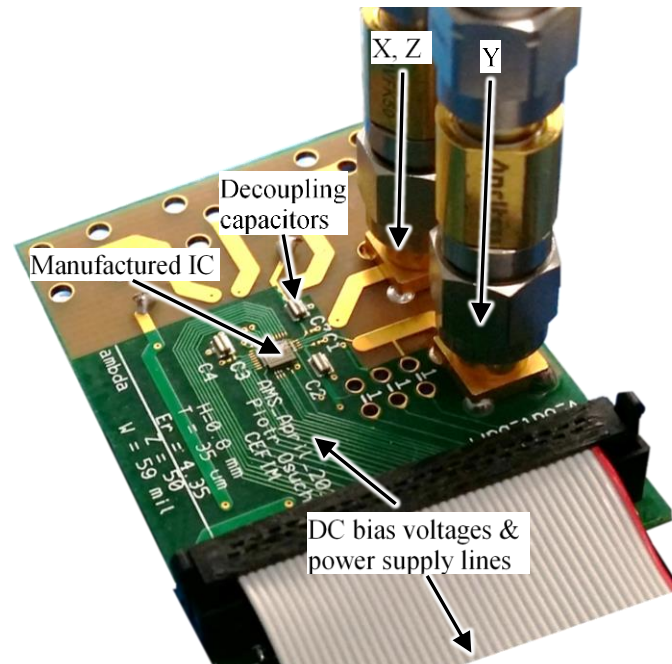
Figure 5.20. Test PCB for housing and testing the proposed CCII+.

As illustrated, including layout parasitics and considering various design corners in the circuit simulations is paramount to obtain accurate results. Moreover, since process tolerances alone lead to a bandwidth variation of more than 100%, worst-case corners should be considered for the intended application. The post-production tunability of r_{fb} (through variation of V_{B2}) is used to fine-tune the resulting voltage following curve as seen in Figure 5.21(a). A flat passband with 0.1 dB ripple (1.15% transfer error, though this is exacerbated by measurement noise) and a bandwidth of 500 MHz is obtained for $V_{B2} = 0.35$ V. A tradeoff between reducing peaking and reducing achievable bandwidth is also evident. Input versus output sinusoid responses are measured with a digital oscilloscope and further shown in Figure 5.21(b) for two different frequencies, with the calculated THD shown in Table 5.9. The DC voltage shift between the input and output ports is caused by the biasing conditions and aspect ratios of M14 – M18, which have been optimized to meet the design goals of Table 5.7. The condition of achieving exactly 0 V DC at the output was not enforced.
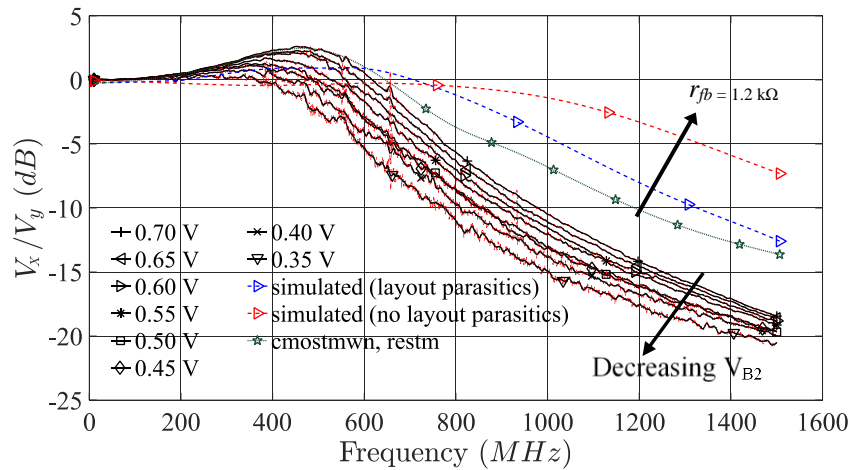
(a)
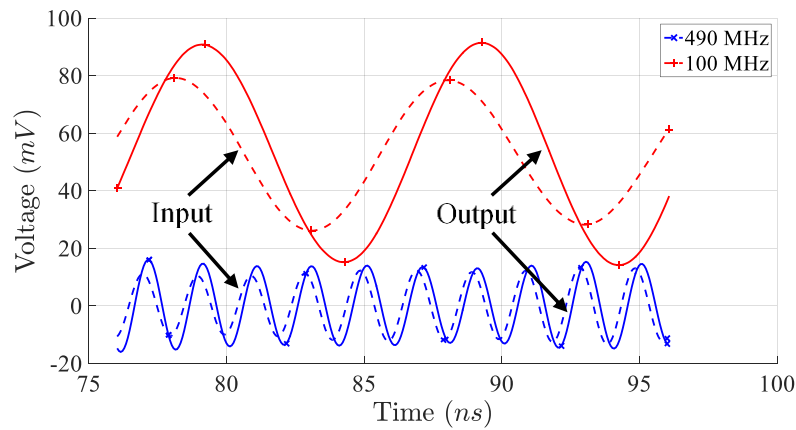
(b)

Figure 5.21. (a) Measured voltage transfer curve of the proposed CCII+. Peaking post-production tunability is accomplished by varying V_{B2}. (b) Measured input versus output sinusoid signals for two frequencies across the passband.

Table 5.9. Simulated and measured THD.

Frequency (MHz)	THD measured (dBc)	THD simulated (dBc)
100	-20	-19
500	-11	-6

Measured results are compared to the design goals stated earlier in Table 5.8, where possible, as shown in Table 5.10. Goals 1 and 2 are used to guide the optimizer and are therefore not measured. Goals 6 and 8 can be inferred partially from the measurements (and the assumption of a 50 Ω load) but cannot be measured directly due to restrictions imposed by the measurement setup already discussed. Other design goals were not measured due to

restrictions on the measurement setup however it is reasonable to expect that they too are met.

<div align="center">Table 5.10. Numerical optimizer goals compared to measured results.</div>

Goal	Quantity	Requirement	Weight	Measured
3	$\beta\big|_{f=0\text{ Hz}} = \overline{(V_X / V_Y)}_{f=0\text{ Hz}}$	$< 0.05\%$	1	0.04
4	$\beta_{BW} = BW_{-3dB}(V_X / V_Y)$	> 700 MHz	2	500 MHz
5	$\beta_{ripp} = \Delta_{\max}(|V_X / V_Y|)$	< 3 dB	1	0.1 dB
6	$\alpha\big|_{f=0\text{ Hz}} = \overline{(I_Z / I_X)}_{f=0\text{ Hz}}$	$< 0.1\%$	1	~0.04 (inferred)
8	$\alpha_{ripp} = \Delta_{\max}(|I_Z / I_X|)$	< 3 dB	1	~0.1 dB (inferred)

5.3.4 Multi-loop feedback analysis application – astable example

To further illustrate the importance of performing a stability analysis when designing high-precision CCIIs, the simulated CCII+ proposed in [94] is implemented in 0.35μm CMOS and manufactured without stability analysis, as shown in Figure 5.22.
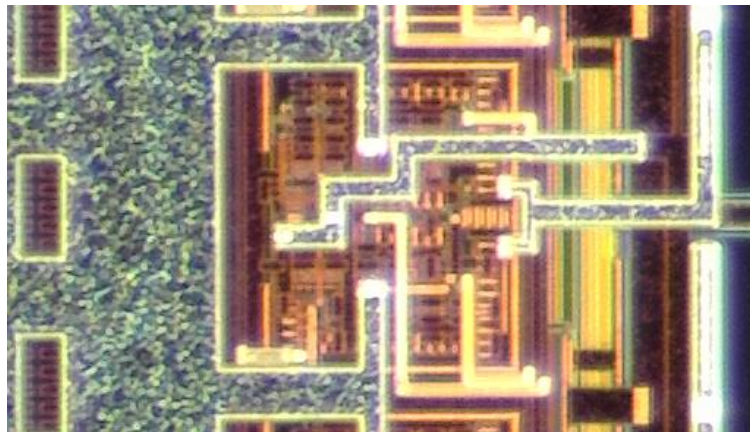

<div align="center">Figure 5.22. Micrograph showing the top view of the CCII+ of Figure 5.23.</div>

The circuit schematic of the resulting CCII+ is shown in Figure 5.23.
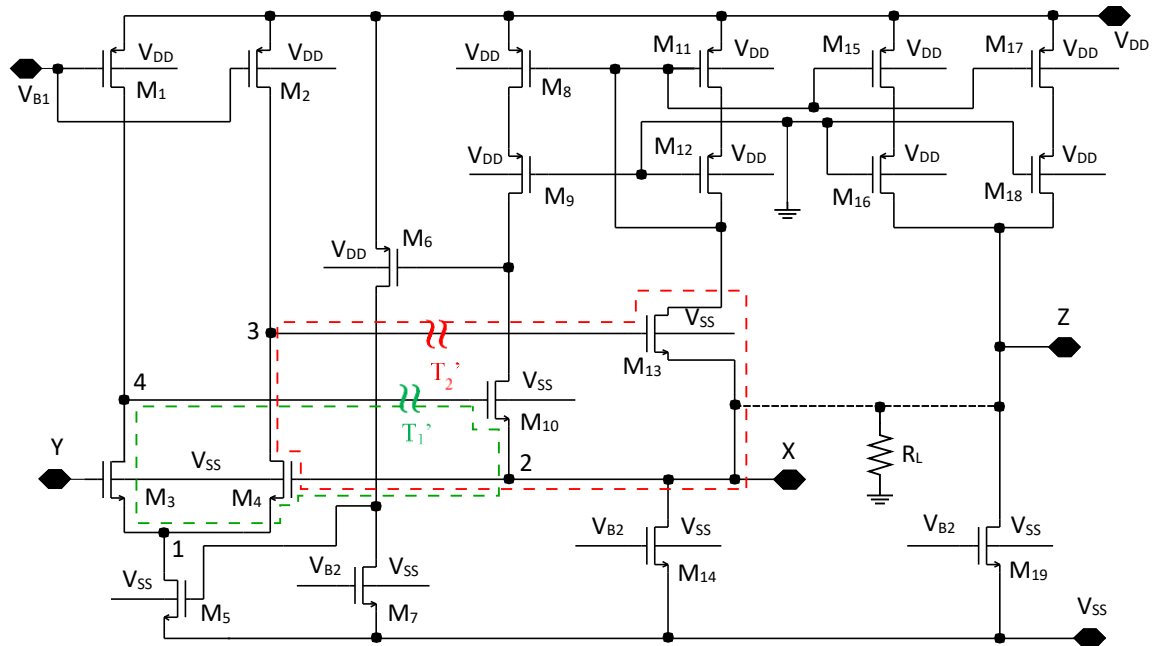
Figure 5.23. Circuit schematic of the high-precision CCII+ presented in [94].

Transistor aspect ratios are chosen as indicated in Table 5.11. Once again, for measurement purposes, terminals X and Z are shorted and terminated in $R_L = 50\ \Omega$, effectively employing the CCII as a unity gain voltage buffer.

Table 5.11. Initial transistor sizes chosen for devices in Figure 5.23.

Device	W/L (µm/µm)	Device	W/L (µm/µm)	Device	W/L (µm/µm)
M1	100/0.35	M7	3.5/0.35	M13	100/0.35
M2	100/0.35	M8	60/0.35	M14	35/0.35
M3	100/0.35	M9	60/0.35	M15	60/0.35
M4	100/0.35	M10	100/0.35	M16	60/0.35
M5	30/0.35	M11	60/0.35	M17	60/0.35
M6	8/0.35	M12	60/0.35	M18	60/0.35
M19	35/0.35				

Next, the multi-loop analysis from Section 5.3.2 is performed on the circuit. Two feedback loops are identified as shown in Figure 5.23 (with numbered circuit nodes corresponding to specific graph nodes in Figure 5.13). The return ratio of T_1' is found first (with loop T_2' left intact) after which T_2' is found with loop T_1' open-circuited. Using (5.15) and (5.17), the effective open-loop gain is computed as shown in Figure 5.24 (a) and (b) for various values of R_L. The validity of the proposed multi-loop feedback analysis is confirmed by

Figure 5.24(c) where the closed-loop poles are computed *independently* by calculating the roots of the transfer function (not using Δ from the injection-based cut approach). This is done by using the symbolic SLiCAP software package. It is clear that the CCII+ is unstable for all $R_L \in (10 : 1.7k)$ Ω. For $R_L = 50$ Ω, oscillation is expected at ~500 MHz. In contrast, the analysis considering only the return ratio $T_1{'}$ of the first loop is computed as shown in Figure 5.24(d), and is clearly incomplete when compared to Figure 5.24(c).
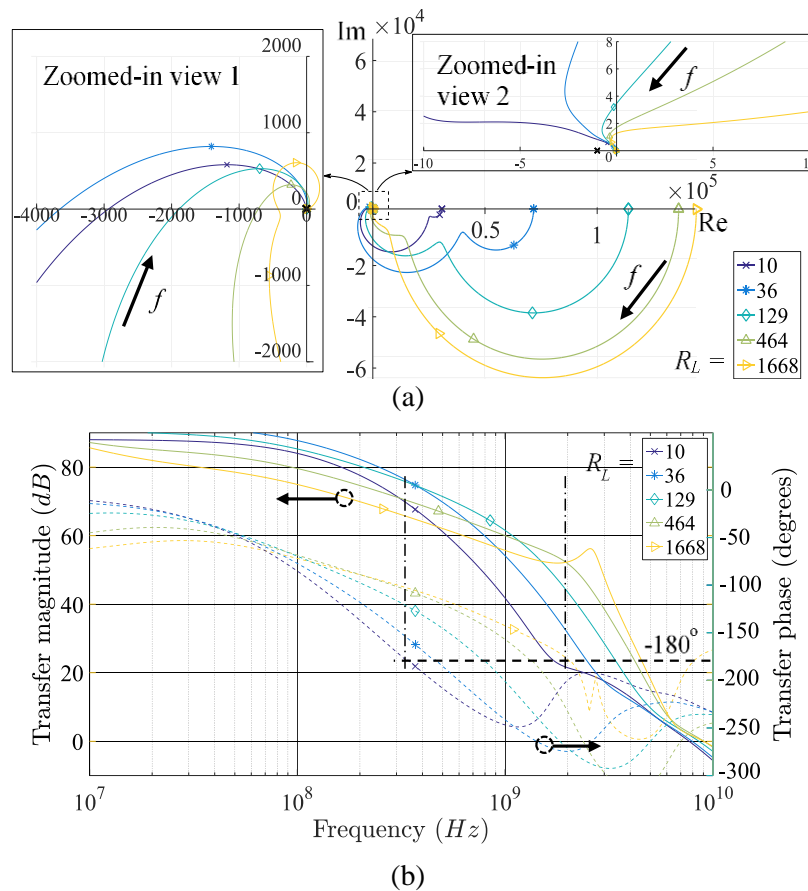

(a)

(b)

Figure 5.24. (a) Nyquist plot of the effective open-loop gain of the CCII+ in Figure 5.23 for varying R_L. (b) Corresponding Bode plot of the open loop gain in (a).
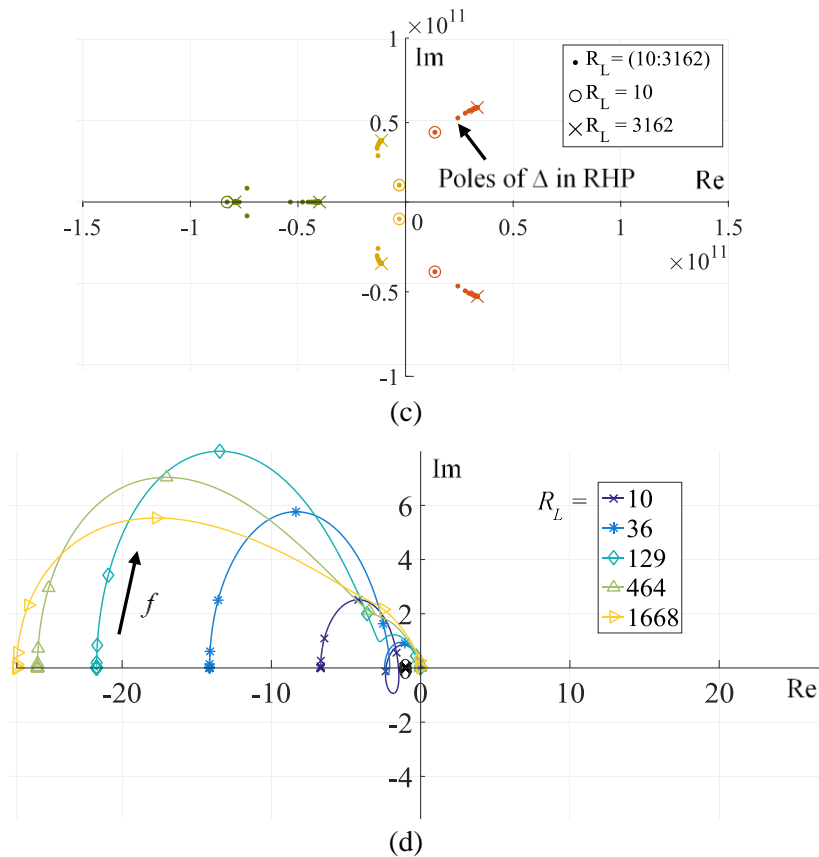
(c)

(d)

Figure 5.24 (cont.). (c) Root-locus plot of the closed-loop gain for varying R_L in the CCII+ in (a).
(d) Nyquist plot of T_1' for varying R_L.

Finally, to confirm the theoretical analysis, the manufactured CCII's output response is measured as shown in Figure 5.25(a – b) for $R_L = 50\ \Omega$. The output noise (with the CCII switched off) is also shown to demonstrate that the measured output, while the CCII is switched on, is indeed caused by internal oscillation. The oscillation frequency is measured as 480 MHz, which corresponds well to the theoretical prediction of ~500 MHz. This result further supports the validity of the presented multi-loop analysis methodology.
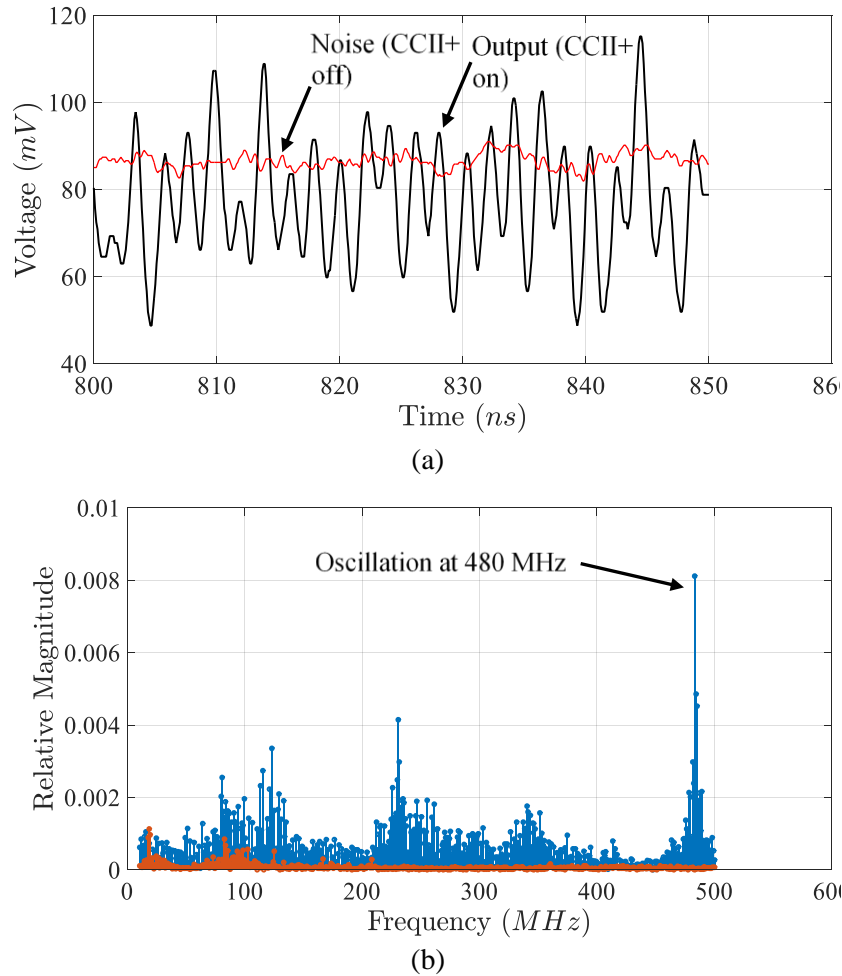
(a)

(b)

Figure 5.25. (a) Measured output response. (b) FFT of measured output response.

5.3.5   Conclusion

A high-precision, high bandwidth CMOS CCII+ with a post-production tunable phase margin and peaking compensation network is presented. A measured transfer error of roughly 1.15% with a bandwidth of 500 MHz is achieved and an $R_X < 5\ \Omega$ is obtained in simulation in 0.35 µm CMOS. A practical numerical optimization based design methodology is presented using accurate device models as well as layout parasitics, allowing good agreement with measured design goals to be obtained. It has been shown that process tolerances can result in more than 100% bandwidth variation, with layout parasitics contributing up to 20% in bandwidth reduction. Additionally, a rigorous multi-loop feedback analysis methodology has been applied to the design and analysis of CCIIs for the first time.

It is shown, using an example of a high-precision CCII+ from literature and measured results, that failure to perform a multi-loop analysis can lead to an unstable design.

5.4   BICMOS ALL-PASS NETWORK DESIGN AND RESULTS

Having presented an on-chip second-order all-pass network in CMOS a high-frequency mm-wave bandwidth design is proposed next in a BiCMOS technology node, as per the proposed research methodology. In contrast to the CMOS design, this all-pass network does not make use of CCIIs.

5.4.1   Introduction

A single transistor second-order all-pass network with inductor Q-factor enhancement has been proposed [56] which addresses many of the shortcomings mentioned in the literature review. However, at mm-wave frequencies, layout RCL parasitic components complicate this simple design, the operation of which relies on zero length connections and ideal components.

In this Section, an on-chip implementation of the single transistor second-order all-pass network is presented, designed for the GlobalFoundries US 8HP 130nm SiGe BiCMOS process. An optimization-based, layout-focused design procedure incorporating accurate device models, distributed elements as well as layout-specific RC parasitic extraction is presented as a unique contribution to the SOTA. The effects of component tolerances, as well as temperature stability on the group delay and magnitude responses, are further investigated.

5.4.2   Design

The single-transistor second-order all-pass network proposed in [56] is adapted for a BiCMOS implementation, as shown in Figure 5.26. An HBT NPN transistor (with emitter parameters as shown) is used. Standard PDK emitter length (L_e) and width (W_e) parameters are chosen which guarantee a maximum unity gain frequency (f_T). The multiplicity refers to

the number of parallel layout sections used, as justified later. Transmission lines T_1 and T_2, placed over a deep trench, act as inductors connected in series with capacitor C_1 to form a resonant LC tank. Transmission line effects are modelled by T_3 and T_4 while the inductive bondwire V_{DD} connection is modeled with a 5 nH inductor. Decoupling capacitors (which are crucial for wideband operation) are modeled by C_3. A gain-enhancement capacitor C_2 is added to extend the bandwidth of the network. Ground-signal-ground (GSG) probe pads are included in the model. Layout specific parasitics (RC) are further extracted and included in the final layout-ready simulation as demonstrated later in Section 5.4.2.2. Since device models supplied by the foundry PDK (pcells) typically include device parasitics, in this work the term "layout parasitics" will be used to refer to additional layout parasitics (eg. interconnect wiring). Furthermore, since T_3 and T_4 model parasitic inductances, they will be included in the term "layout parasitics".
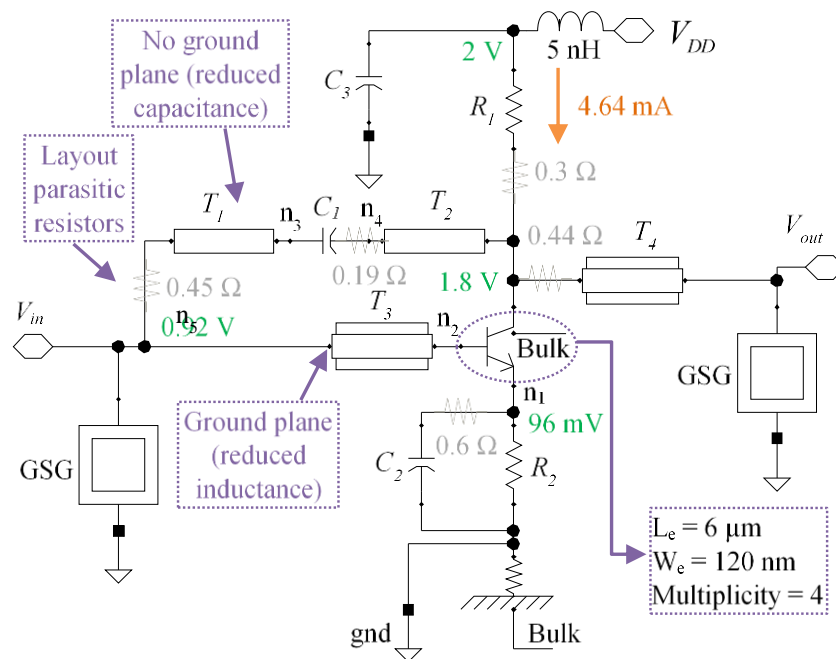

Figure 5.26. Schematic of the single-transistor second-order all-pass network.

The following *design methodology* is proposed incorporating a layout-focused design optimization approach which is a unique contribution to the SOTA. This is necessary to

obtain the desired group delay and at the same time a maximally flat magnitude response at mm-wave frequencies.

1. Simplified design equations are derived which exclude layout parasitics and set $C_3 \rightarrow \infty$, $C_2 \rightarrow 0$.

2. Initial component values are calculated from the equations in *(1)*.

3. An optimization is performed in simulation using the foundry PDK and initial values from the previous step. Additionally, C_3 is chosen sufficiently large to decouple the circuit for the shown bond-wire inductance to V_{DD}, whereas the value of C_2 is chosen to minimize high-frequency peaking.

4. Using the values obtained in *(3)*, a layout is constructed and layout specific parasitics extracted.

5. A second optimization is performed, this time including all parasitics (T_3, T_4 and extracted RC).

6. The layout is finally updated to reflect the optimized changes. Steps 5 and 6 are repeated until the design objectives are met.

5.4.2.1 Theoretical calculations

The *voltage* transfer function of the circuit in Figure 5.26. can be approximated by [56]:

$$T(s) = K \cdot \frac{s^2 \cdot LC_1 - s \dfrac{C_1}{G_2 - G} + 1}{s^2 \cdot LC_1 + s \dfrac{C_1}{G_2 - G} + 1},$$ (5.23)

where L is the combined inductance of T_1 and T_2, $G_2 = 1/R_2$, G is the combined parasitic conductance in parallel with T_1, T_2 and C_1 and K is a real constant.

Equating similar terms in (5.23) and (5.2) leads to:

$$\frac{1}{\omega_0^{\,2}} = LC_1 , \tag{5.24}$$

$$\frac{1}{Q_D \cdot \omega_0} = \frac{C_1}{G_2 - G} . \tag{5.25}$$

In order to ensure an all-pass response, the following condition must be met [56]:

$$G_1 = \frac{a \cdot G_2^{\,2} + b \cdot G_2 + c}{2G + G_2 \cdot \left(1/\beta - 1\right) + G_S} , \text{ where:} \tag{5.26}$$

$a = 1 + \dfrac{1}{\beta}$, $b = -2G + G_L + G_S - \dfrac{2G + G_L}{\beta}$, $c = -2G \cdot \left(G_L + G_S\right) - G_L G_S$, β is the NPN small

signal current gain and G_L and G_S are the load and source conductances. Finally, to avoid signal clipping, the following inequality applies:

$$\frac{G_2}{G_1} \le \frac{V_{DD} - V_{sat} - V_{swing}}{2 \cdot \left(V_B - V_{BE}\right)} , \tag{5.27}$$

where V_{sat} is the HBT saturation voltage, V_{swing} is the output p-p voltage, V_B is the input bias voltage and V_{BE} is the base-emitter voltage drop. The above equations are solved by first choosing G_2 and V_B such that $G_1 > 0$ in (5.26), (5.27) is satisfied, and an optimal collector-emitter bias current is obtained that maximizes the transistor f_t. From these values, (5.25) is solved for C_1 and, (5.24) is solved for L.

Applying step 2 of the design methodology produces $\Delta\tau = 38$ ps, $f_0 = 30$ GHz, $G_S = 0.02$ S, $G_L = 0.02$ S (chosen for 50 Ω system integration for measurement with a VNA), $G = 0.04$ S, G_2 is chosen as 0.086 S and $V_B = 0.92$ V resulting in $G_1 = 0.025$ S, $C_1 = 136$ fF and $L = 207$ pH. The chosen Q_D is therefore 1.8, found from $\Delta\tau$, f_0 and (2.2). Initial values for C_2 and C_3 are chosen as 60 fF and 75 pF respectively. The circuit schematic of Figure 5.26 with device parameters as calculated above, but excluding layout parasitics, is simulated using the foundry PDK. The results are shown in Figure 5.27 (blue trace). Practical

non-idealities (as included in the PDK models) contribute to the magnitude ripple of 7 dB and group delay response deviation from the ideal.
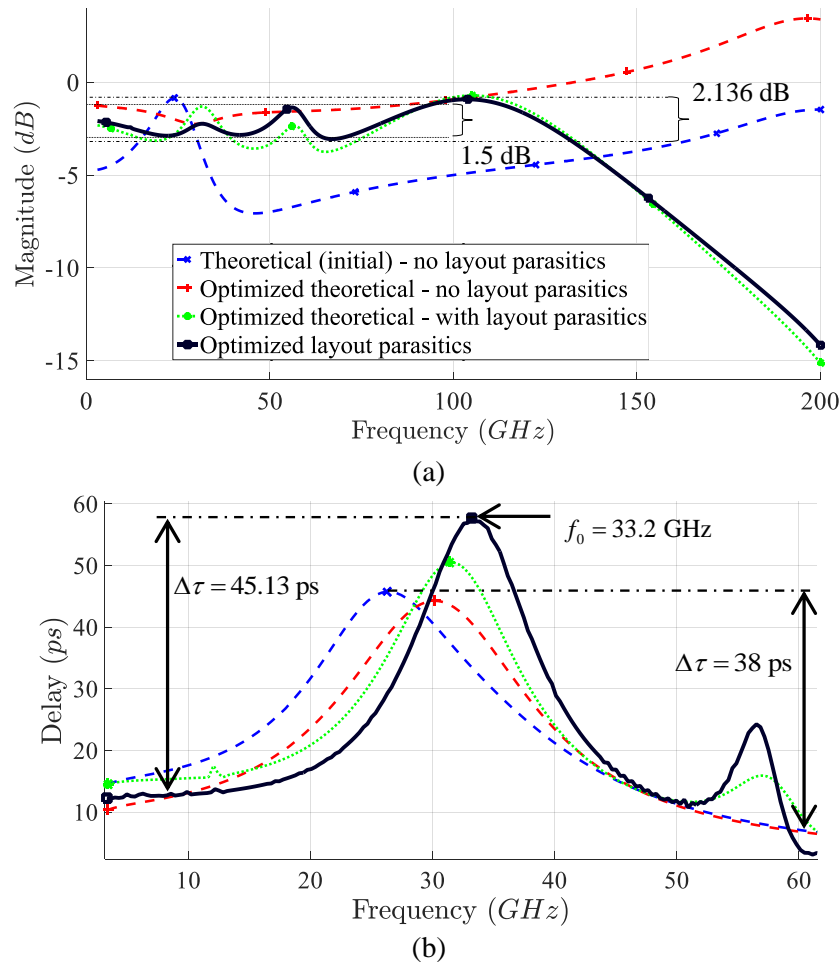

Figure 5.27. Transmission magnitude and delay responses (voltage transfer).

5.4.2.2 Application of layout-based optimization

The first optimization is now performed to incorporate these device non-idealities into the design (step *3* of the design methodology). To this end, the desired f_0, $\Delta\tau$ (chosen as 30 GHz, 38 ps respectively) and the magnitude variation in the passband (chosen to be lower than 1.5 dB) are set as gradient-based optimization goals with C_1, C_2, L, R_1 and R_2 chosen as variable parameters to be optimized. The choice in $\Delta\tau$ results in a Q_D of 1.8 > 1. The value of the objective function for each iteration is shown in Figure 5.28 (solid curve).
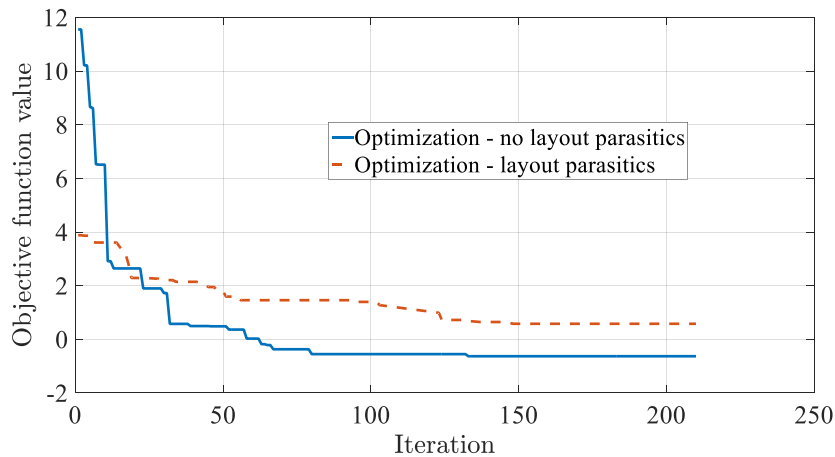
Figure 5.28. Objective value as a function of the optimizer iteration.

The resulting magnitude and group delay responses are shown in Figure 5.27 (red trace). The magnitude ripple is improved to 0.8 dB in the passband and the group delay curve has the desired f_0 and $\Delta\tau$ of 30 GHz and 38 ps respectively (Q_D of 1.8).

Next (step *(iv)* of the proposed design methodology), based on these component values, a layout is drawn and layout parasitics extracted as shown in Figure 5.29.
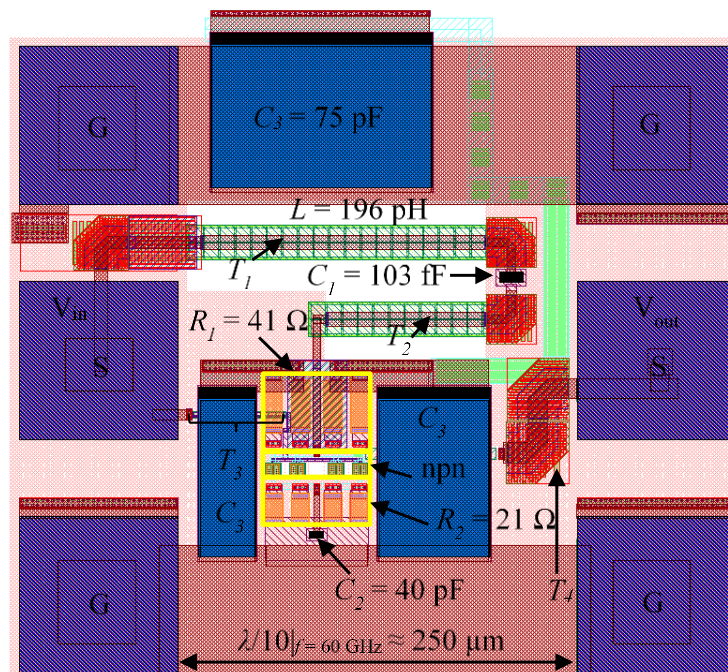

Figure 5.29. Camera-ready layout of the BiCMOS second-order all-pass network.

The design is smaller than $\lambda/10$ at the highest frequency of interest, therefore distributed effects are not considered in the model. The NPN transistor is implemented using 4 parallel sections in order to improve layout symmetry. A total of 146 parasitic capacitors and 35 parasitic resistors are extracted from the layout, with the dominant layout parasitic resistors indicated in Figure 5.26. The sum of layout parasitic capacitances at each node is: n_1: 56.88 fF, n_2: 48.74 fF, n_3: 1.2 fF, n_4: 3.61 fF, n_5: 71.95 fF, V_{DD}: 284.5 fF, V_{in}: 9.81 fF, V_{out}: 0.67 fF, bulk: 4.07 pF, and gnd: 4.6 pF. As these values are comparable to some component values (e.g. C_1) the importance of the RC extraction and the layout-based approach is evident.

After RC extraction, another optimization is performed, this time including layout parasitics (step *(v)*). The value of the objective function for each iteration is shown in Figure 5.28 (dashed curve). Optimized component values are shown in Figure 5.29.

The resulting magnitude and group delay responses are shown in Figure 5.27 (black trace). A final magnitude ripple of 1.5 dB (2.14 dB including peaking) and a second-order group delay curve with $f_0 = 33.2$ GHz and $\Delta\tau = 45.13$ ps (Q_D of 2.4) is obtained. Even though a 3 dB bandwidth of 130 GHz is observed, at 50 GHz the all-pass assumption is no longer valid resulting in a second delay peak at 56 GHz not accounted for in the theoretical synthesis. Therefore, the usable "group delay bandwidth" of this design is 50 GHz in 130 nm BiCMOS.

A noise simulation is also performed, as shown in Figure 5.30, with an average output noise of 0.69 nV/√Hz.
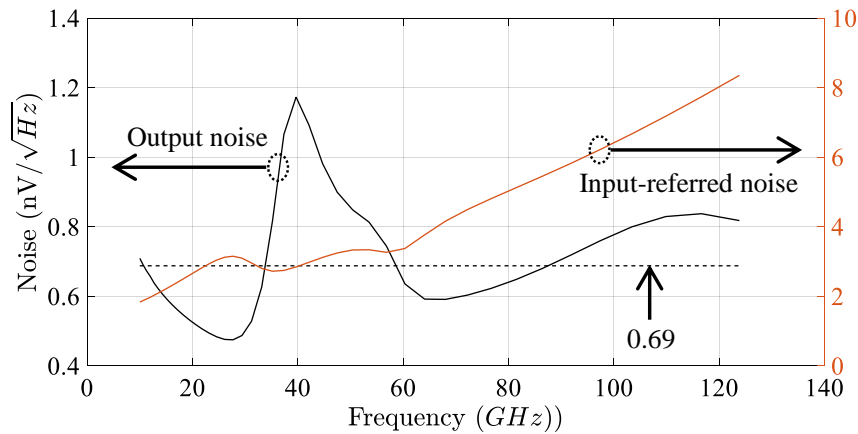
Figure 5.30. Noise simulation of the all-pass network.

5.4.3 Results

The layout in Figure 5.29 is manufactured using the GlobalFoundries US 8HP 130 nm SiGe BiCMOS process as shown in the micrograph in Figure 5.31. The measurement setup is also shown, where the Picoprobe Model 110H GSG wafer probes are connected to an Anritsu ME7828A VNA. A DC bias needle supplies the V_{DD} power supply voltage as shown.
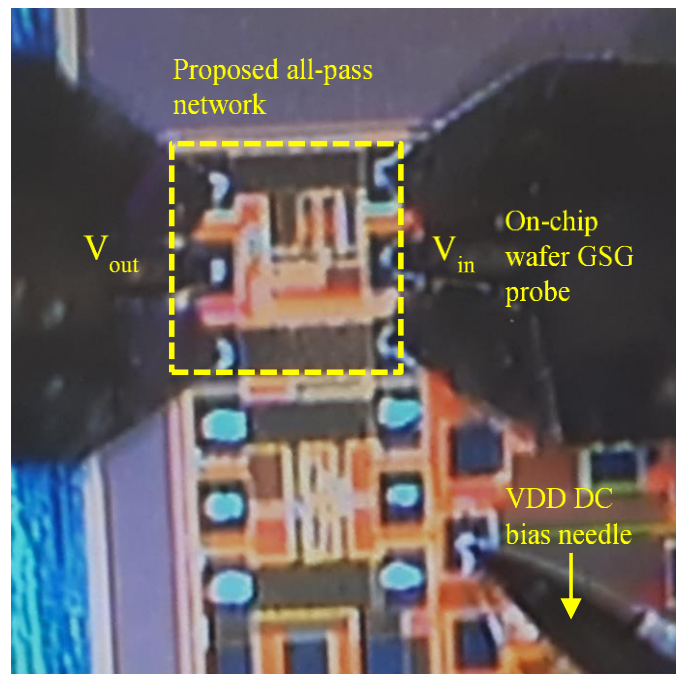

Figure 5.31. Micrograph of the second-order all-pass network.

The measured results of the proposed all-pass network are shown in Figure 5.32 and compared to the simulated response from Figure 5.27 (black trace). Four separate manufactured ICs are measured, as indicated by the separate magnitude response and group delay curves.

A magnitude response ripple of 1.4 dB, excluding peaking, is measured, which corresponds well to the simulated value of 1.5 dB. If peaking is however considered, a 3.44 dB ripple is measured, which is larger than the simulated value of 2.14 dB. This discrepancy, as well as the discrepancy between the measured and simulated magnitude response curve shapes can be explained by TaN resistor process variations (R_1, R_2) as well as additional parasitic elements, as will be shown later (Section 5.4.3.1) in the corner analysis.

The obtained $\Delta\tau$ of 62.4 ps (Q_D of 3.6) is larger than the simulated value of 45 ps. This discrepancy can again be attributed to TaN resistor variations, as shown next in Section 5.4.3.1. The center frequency of the group delay response is within 10% of the simulated value. The discrepancy is likely caused by additional parasitic RC components not considered in simulation (Figure 5.27 would indicate that incorporating parasitics increases the center frequency of the delay response).
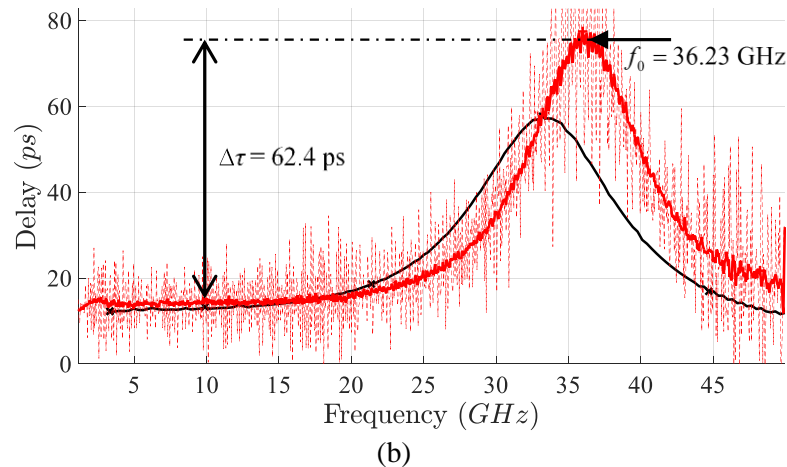

(a)

(b)
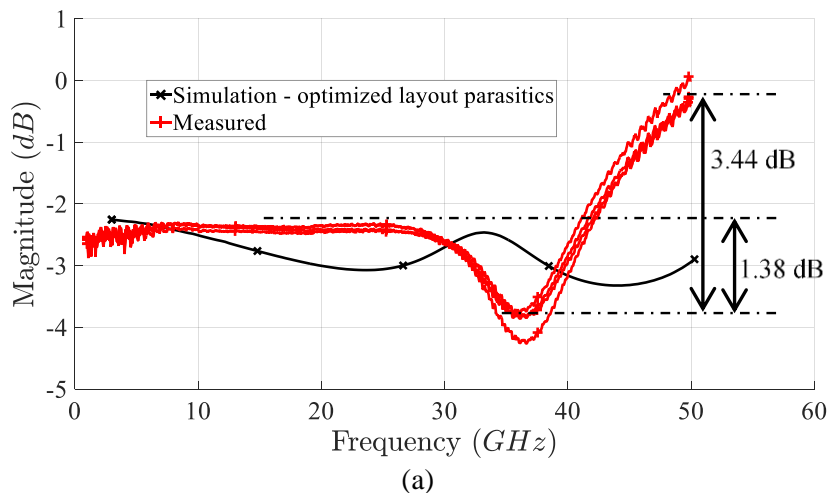
Figure 5.32. Measured magnitude and delay responses (voltage transfer).

A usable bandwidth of 40 GHz is obtained over which the magnitude variation is less than 1.4 dB. However, in applications where a larger magnitude variation is acceptable the usable bandwidth can be larger. Lastly, the compression plot is shown in Figure 5.33, indicating a 1 dB compression at -6.7 dBm.
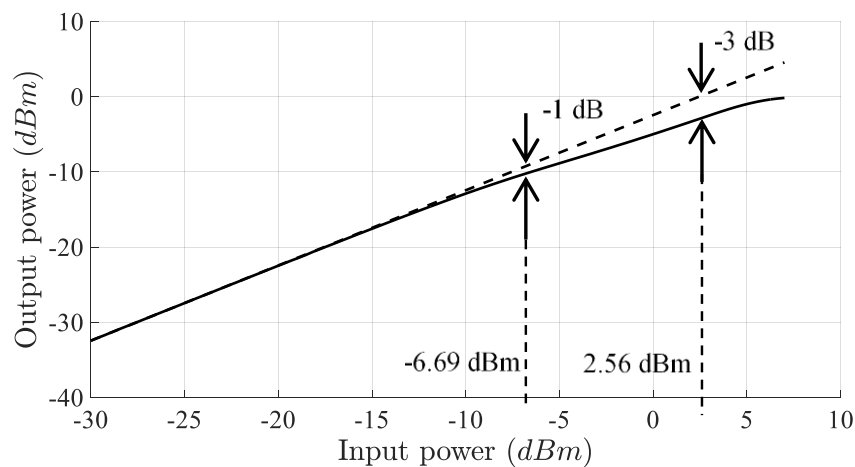
Figure 5.33. Compression plot of the proposed network at 1 GHz.

5.4.3.1   Sensitivity to process tolerances and operating conditions

To explain the aforementioned discrepancies between simulations and measured results a corner analysis is performed, using the PDK tolerance and temperature bounds, as shown in Figure 5.34. The poor absolute tolerances of on-chip resistors contribute mostly to deviation from the nominal corner and explain the magnitude ripple and $\Delta\tau$ discrepancies mentioned

above. The local minimum observed in the measured magnitude response at 36.23 GHz also appears for some TaN resistor process corners, as shown, explaining the abovementioned discrepancy in the magnitude response shape. However, neither process tolerances nor the temperature sweep, account for the discrepancy in f_0, which could suggest that additional unmodeled parasitic components are responsible for this. The temperature simulations further indicate that in applications where a wide-range of extreme temperatures are expected, post-production tunability of one or more components is likely required. It is important to note that in this work the temperature sweep does not serve to explain discrepancies. The temperature sweep should therefore be seen as additional information, which could be useful to the reader.
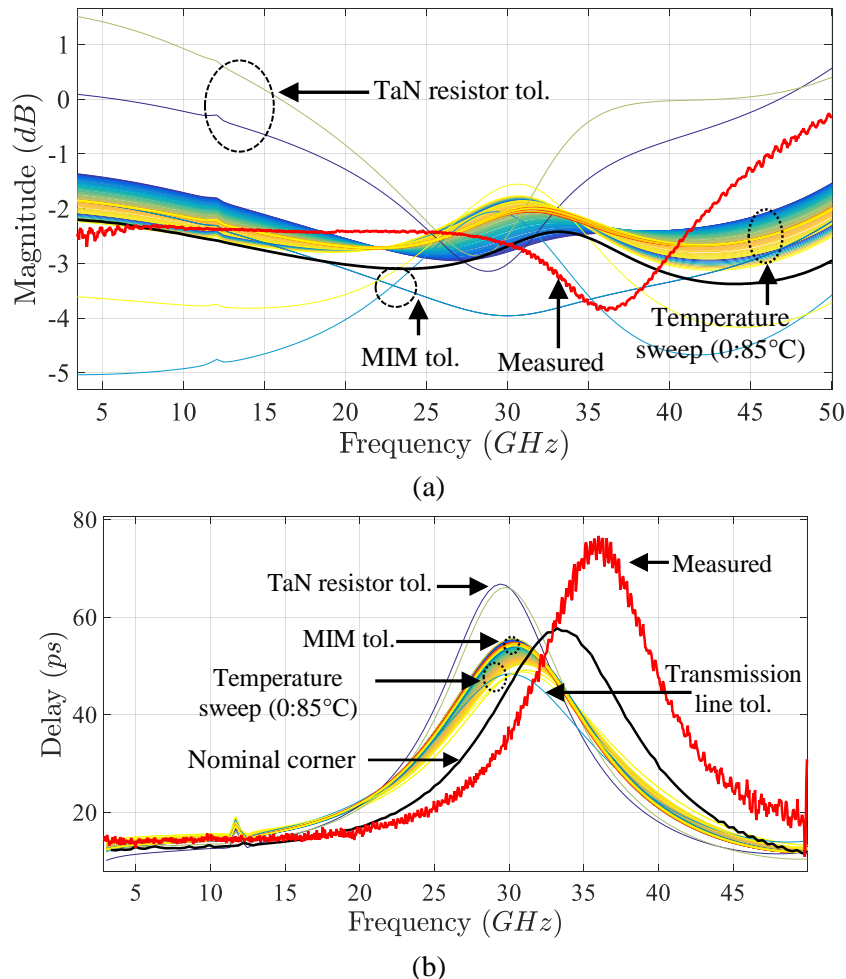

(a)

(b)

Figure 5.34. Component tolerance and temperature corner sweep of the transmission magnitude and delay responses (voltage transfer).

5.4.4   Conclusion

An optimization-based, layout-focused design procedure is proposed and used to design an active on-chip mm-wave second-order all-pass network in a 130 nm SiGe BiCMOS technology node with an effective usable bandwidth of 40 GHz, peak-to-nominal delay of 62 ps, magnitude ripple of 1.4 dB and an output noise of 0.69 nV/$\sqrt{Hz}$. Worst-case component tolerances are presented and used to explain discrepancies between simulated and measured results.

5.5   CONCLUSION

In this Chapter, practical implementations of on-chip second-order all-pass networks are presented for both CMOS and BiCMOS processes, achieving Q_D-values of 1.15 and 3.6 respectively. These are the highest in reported literature and, for the first time make an on-chip ASP viable.

First, a novel on-chip active second-order all-pass network is proposed in a 0.35 µm CMOS process, with post-production tunability to account for CCII non-idealities as well as process tolerances. A high-precision, high bandwidth CMOS CCII+ with a post-production tunable phase margin and peaking compensation network is further proposed as a building block used to construct this all-pass network. A measured transfer error of roughly 1.15% is achieved with a measured bandwidth of 500 MHz and a simulated $R_X < 5\ \Omega$ in 0.35 µm CMOS. Additionally, a rigorous multi-loop feedback analysis methodology has been applied to the design and analysis of CCIIs. An active on-chip mm-wave second-order all-pass network is further demonstrated in a 130 nm SiGe BiCMOS technology node with a bandwidth of 40 GHz, peak-to-nominal delay of 62 ps, Q_D-value of 3.6 and a magnitude ripple of 1.4 dB. This presents the first mm-wave all-pass network in published literature.

CHAPTER 6   CONCLUSION

6.1   INTRODUCTION

A summary of the research work undertaken is presented in this Chapter. The research questions are evaluated, challenges and limitations indicated and a scope for future work suggested.

6.2   CRITICAL EVALUATION OF RESEARCH QUESTIONS

The monolithic integration of all-pass networks in commercial CMOS and BiCMOS technology nodes has presented a number of yet unsolved challenges. Firstly, the state-of-the-art methods for synthesizing quasi-arbitrary group delay functions using all-pass elements lack a theoretical synthesis procedure that guarantees minimum-order networks. Secondly, second-order all-pass networks, the building blocks of ASPs, have traditionally been limited to off-chip media which are unsuited for volume production.

The following primary research question was therefore posed in this work and is repeated here for convenience.

How can a quasi-arbitrary group delay function be synthesized with a minimum-order network and implemented on-chip in commercial CMOS and BiCMOS technology nodes?

The primary research question and associated secondary research questions (as detailed in Section 1.2) are answered as follows.

1. A minimum-order group delay function representing any quasi-arbitrary delay function can be synthesized by cascaded second-order all-pass networks as shown in Chapter 4. A novel analytically-based quasi-arbitrary group delay function synthesis

theory using second-order all-pass elements is developed to do so. This is done by finding a novel approximation to a second-order all-pass function, thus leading to an approximate analytical solution to the synthesis problem in the form of initial second-order all-pass pole/zero solution intervals. The intervals are found in such a way that a gradient-based optimization always converges to the optimal equi-ripple solution. The resulting group delay function is, furthermore, an approximation of the required response to within any arbitrarily specified maximum delay variation across the passband. Therefore, any continuous delay function can be synthesized with cascaded second-order all-pass networks to within any arbitrary equi-ripple error. Moreover, the method is implementation abstracted, since the synthesized poles / zeros may be implemented in various ways, therefore providing the most general answer possible to the primary research question. The proposed method is demonstrated by synthesizing linear, quadratic, Gaussian and higher-order group delay functions and compared to SOTA methods. It is further shown that existing methods do not always converge to a global optimum (as the proposed method does).

2. Second-order all-pass networks have traditionally been limited to off-chip media due to the high losses of on-chip elements, making existing implementations unsuitable for ASP applications (Q_D-values smaller than 1). It is found that second-order all-pass networks with Q_D-values larger than 1 (the building blocks of ASPs) can be implemented on-chip in CMOS processes using active-enhanced RC networks. This eliminates the need for on-chip inductors, which suffer from low Q-factors and lead to otherwise low all-pass Q_D-values. In this work, a novel on-chip CMOS second-order all-pass network is proposed which employs CCIIs as active elements and achieves a bandwidth of 280 MHz, peak-to-nominal delay variation of 10 ns, Q_D-value of 1.15 and magnitude variation of 3.1 dB. It is shown that a tuning mechanism controlled post-production with a genetic local optimizer can compensate for CMOS process tolerances and parasitics, which make physical realization feasible.

It is further found that second-order all-pass networks with Q_D-values larger than 1 and mm-wave bandwidths can be implemented on-chip in BiCMOS technology

nodes using RLC networks with an actively enhanced inductor to compensate for on-chip losses. An optimization-based, layout-focused design approach is proposed leading to an on-chip mm-wave second-order all-pass network in a 130 nm SiGe BiCMOS technology node with a bandwidth of 40 GHz, peak-to-nominal delay of 62 ps, Q_D-value of 3.6 and a magnitude ripple of 1.4 dB.

3. Improvements are also made to the SOTA CCII literature, as is required by the CMOS all-pass design described above. In this work, the improved CCII is used as a buffer stage in the all-pass network, and an existing CCII from literature is used for the remaining CCII elements in order to demonstrate the ability of the network to compensate for CCII non-idealities. The proposed CCII can, however, be used throughout the all-pass network design to improve performance metrics as suggested for future work. A practical CCII design methodology that incorporates accurate foundry device models, process corners, layout parasitics, and explicit multi-loop feedback stability analysis into the design flow is presented for the first time. A post-production tunable feedback compensation network is further included to enable post-production stability and peaking control. The method is experimentally verified by producing a high-precision (1.15% transfer error – measured), high-bandwidth (> 500 MHz – measured) CCII+ in 0.35μm CMOS with $R_X < 5\ \Omega$ (simulated), along with an astable counter-example proposed in literature.

6.3   CHALLENGES AND LIMITATIONS

A significant challenge experienced in this work is the limited practical applicability of SOTA circuits to ASP on chip. Even though an initial literature review indicated that only a few improvements on the SOTA are necessary to implement a second-order all-pass network on-chip suitable for ASP applications and that the work's scope would lie in the synthesis and implementation of higher order delay networks with a certain ASP application in mind, this turned out more challenging than expected. Many results of on-chip all-pass networks presented in the literature are based on ideal theoretical simulations which do not take into account component non-idealities, process tolerances and parasitic elements. Where

measured results are presented, they are found to be unsuitable for ASP applications due to a low Q_D-value, as summarized in Table 5.5. Moreover, implementing CCIIs in CMOS (the active building block of the all-pass network in this work), turned out to require significant improvements on the SOTA. As was shown in Section 5.3.4, a detailed stability analysis had never been performed in literature, resulting in a first prototype which was unstable and subsequently used as a counter-example. The BiCMOS all-pass design also required improvements on the SOTA. Therefore, it is evident that simply implementing a second-order all-pass network in both CMOS and BiCMOS technology nodes, on its own, presents a wide research scope. This view is further justified by the contributions of this work to the field of Electronic Engineering, as detailed in Section 1.6.

A number of limitations can be identified in this work as listed below.

1. The proposed group delay synthesis theory experiences convergence problems if the left- and right-most peaks of the group delay are approximately equal, as mentioned in Section 4.2.2. The proposed method of raising/pre-distorting one band edge with an additional pole to circumvent this problem is general and no rigorous approach to achieving this pre-distortion is given.

2. A large number of optimization iterations are required to fine-tune the CMOS second-order all-pass network in post-production tuning as is evident in Section 5.2.4.

3. The proposed high-precision CCII is only used as a voltage buffer in the proposed second-order all-pass network (Section 5.2.3.2) and not as the active CCII element (an existing lower-precision design from literature is used instead) due to increased tuning complexity.

4. The proposed CCII achieves high A_i, A_v precision and a low R_X but at the cost of a low R_Z (Section 5.3.3.1).

5. Discrepancies between simulated and measured results of the mm-wave BiCMOS second-order all-pass network are not fully explained and only the most likely cause identified (Section 5.4.3.1). Relying entirely on accurate device models and parasitic RC extraction in the Ka-band is a limiting factor in the presented analysis.

6.4 SUGGESTED FUTURE WORK

This work has laid the foundations for the future synthesis and microelectronic implementation of all-pass networks with potential applications in ASPs, by presenting a synthesis theory for minimum-order all-pass networks and on-chip implementations of second-order all-pass networks, the building blocks of higher-order all-pass networks, in CMOS and BiCMOS technology nodes.

Future work should focus on addressing the limitations listed above, as follows. The numbering in this lists corresponds to the list of limitations in Section 6.3.

1. A formalized method of achieving the required predisposition to ensure convergence in the aforementioned case can be developed.

2. Methods of reducing the number of post-production tuning iterations can be investigated. This can possibly be achieved by combining accurate models of the design obtained from simulation with the numerical optimization algorithm, in order to reduce the number of unnecessary tuning operations. Furthermore, incorporating knowledge of the CMOS process corner into the simulation analysis and selecting initial values designed specifically for the relevant corner can further speedup the optimization.

3. Some or all of the CCII elements in the proposed CMOS all-pass network can be replaced with the proposed high-precision CCII to improve performance metrics (at the cost of a lower bandwidth).

4. Methods of improving the low R_Z, without appreciably deteriorating the remaining parameters, can be investigated, by for instance, including additional buffering.

5. A FEM electromagnetic simulation of the mm-wave BiCMOS all-pass network can be investigated in order to improve agreement between measured and simulated results. Post-production tunability can further be investigated to account for process tolerances.

Future work should also focus on implementing higher-order all-pass networks by cascading the proposed second-order delay networks. The proposed synthesis theory can then be used to synthesize minimum-order delay functions with certain ASP properties, as required by a specific application. Various potential applications of have been identified in Section 2.6.

REFERENCES

[1]   A. Ghosh *et al.*, "Millimeter-wave enhanced local area systems: a high-data-rate approach for future wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1152–1163, 2014.

[2]   R. A. Kertis *et al.*, "A 35 GS/s 5-bit SiGe BiCMOS flash ADC with offset corrected exclusive-or comparator," in *IEEE Bipolar/BiCMOS Circuits and Technology Meeting*, 2008, pp. 252–255.

[3]   J. Manley and F. Kapp, "The MeerKAT array and its digital signal processor," in *International Conference on Electromagnetics in Advanced Applications*, 2012, pp. 462–465.

[4]   E. Ifeachor and B. Jervis, "Analog I/O interface for real-time DSP systems," in *Digital Signal Processing: A practical approach, 2nd ed.*, Essex: Pearson Education Limited, 2002, pp. 43–44.

[5]   D. B. Davidson, "MeerKAT and SKA phase 1," in *IEEE 10th International Symposium on Antennas, Propagation and EM Theory*, 2012, pp. 1279–1282.

[6]   P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, "The Square Kilometre Array," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.

[7]   J. Brinkhoff, D. Luzon, A. Terry, Y. Moghe, and S. Jackson, "A wideband downconverter and analog baseband for radio astronomy applications," in *Microwave Conference Proceedings (APMC)*, 2011, pp. 1230–1233.

[8]   D. L. Jones *et al.*, "Big data challenges for large radio arrays," in *IEEE Aerospace Conference*, 2012, pp. 1–6.

[9]   R. R. Schaller, "Moore's law: past, present, and future," *IEEE Spectrum*, vol. 34, no. 6, pp. 53–59, 1997.

[10]   S. Gupta, A. Parsa, E. Perret, R. V. Snyder, R. J. Wenzel, and C. Caloz, "Group-delay

engineered noncommensurate transmission line all-pass network for analog signal processing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 9, pp. 2392–2407, 2010.

[11] C. Caloz, S. Gupta, Q. Zhang, and B. Nikfal, "Analog signal processing: a possible alternative or complement to dominantly digital radio schemes," *IEEE Microwave Magazine*, vol. 14, no. 6, pp. 87–103, 2013.

[12] L. Zou, S. Gupta, and C. Caloz, "Loss-gain equalized reconfigurable C-section analog signal processor," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 2, pp. 555–564, 2017.

[13] S. Gupta and C. Caloz, "Analog real-time fourier transformer using a group delay engineered C-section all-pass network," in *IEEE Antennas and Propagation Society International Symposium (APSURSI)*, 2010, pp. 1–4.

[14] S. Abielmona, S. Gupta, and C. Caloz, "Compressive receiver using a CRLH-based dispersive delay line for analog signal processing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 11, pp. 2617–2626, 2009.

[15] S. Gupta, S. Abielmona, and C. Caloz, "Microwave analog real-time spectrum analyzer (RTSA) based on the spectral-spatial decomposition property of leaky-wave structures," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 12, pp. 2989–2999, 2009.

[16] J. D. Schwartz, J. Azaña, and D. V. Plant, "Experimental demonstration of real-time spectrum analysis using dispersive microstrip," *IEEE Microwave and Wireless Components Letters*, vol. 16, no. 4, pp. 215–217, 2006.

[17] M. A. Laso *et al.*, "Real-time spectrum analysis in microstrip technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 3, pp. 705–717, 2003.

[18] H. V Nguyen and C. Caloz, "CRLH delay line pulse position modulation transmitter," *IEEE Microwave and Wireless Components Letters*, vol. 18, no. 8, pp. 527–529, 2008.

[19] B. Xiang, A. Kopa, Z. Fu, and A. B. Apsel, "Theoretical analysis and practical considerations for the integrated time-stretching system," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 11, pp. 3449–3457, 2012.

[20] J. D. Schwartz, J. Azaña, and D. V. Plant, "An electronic temporal imaging system for compression and reversal of arbitrary UWB waveforms," in *2008 IEEE Radio and Wireless Symposium (RWS)*, 2008, pp. 487–490.

[21] B. Nikfal, D. Badiere, M. Repeta, B. Deforge, S. Gupta, and C. Caloz, "Distortion-less real-time spectrum sniffing based on a stepped group-delay phaser," *IEEE Microwave and Wireless Components Letters*, vol. 22, no. 11, pp. 601–603, 2012.

[22] G. Zhang, Q. Zhang, F. Yang, Y. Chen, C. Caloz, and R. D. Murch, "Phaser-based feeding network for uniformly scanning antenna arrays," in *Antennas and Propagation & USNC/URSI National Radio Science Meeting*, 2015, pp. 236–237.

[23] G. Zhang *et al.*, "Dispersive feeding network for arbitrary frequency beam scanning in array antennas," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 6, pp. 3033–3040, 2017.

[24] M. K. Mandal, D. Deslandes, and K. Wu, "Complementary microstrip-slotline stub configuration for group delay engineering," *IEEE Microwave and Wireless Components Letters*, vol. 22, no. 8, pp. 388–390, 2012.

[25] N. Udayanga *et al.*, "All-pass filter based synthesis of multifunctional microwave active circuits all-pass filter based synthesis of multifunctional microwave active circuits," in *IEEE Vehicular Technology Conference*, 2017, vol. 1, pp. 1–4.

[26] P. J. Osuch and T. Stander, "A geometric approach to group delay network synthesis," *Radioengineering*, vol. 25, no. 2, pp. 351–364, 2016.

[27] S. Hellerstein, "Synthesis of all-pass delay equalizers," *IRE Transactions on Circuit Theory*, vol. 8, no. 3, pp. 215–222, 1961.

[28] Q. Zhang, J. W. Bandler, and C. Caloz, "Design of dispersive delay structures (DDSs) formed by coupled C-sections using predistortion with space mapping," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4040–4051, 2013.

[29] B. Nikfal, S. Gupta, and C. Caloz, "Increased group delay slope loop system for enhanced-resolution analog signal processing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 6, pp. 1622–1628, 2010.

[30] R. Kumar, P. Keerthan, and K. J. Vinoy, "Design of wideband semi-lumped all-pass

network for linear group delay response," in *IEEE Proceedings of the Asia-Pacific Microwave Conference (APMC)*, 2016, pp. 1–4.

[31] S. Gupta, D. Sounas, Q. Zhang, and C. Caloz, "All-pass dispersion synthesis using microwave C-sections," *International Journal of Circuit Theory and Applications*, vol. 42, no. 12, pp. 1228–1245, 2014.

[32] Q. Zhang, S. Gupta, and C. Caloz, "Synthesis of broadband dispersive delay structures formed by commensurate C-and D-sections," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 24, no. 3, pp. 322–331, 2014.

[33] B. Al-Hashimi, F. Dudek, and M. Moniri, "Current-mode group-delay equalisation using pole-zero mirroring technique," *IEE Proceedings - Circuits, Devices and Systems*, vol. 147, no. 4, pp. 257–263, 2000.

[34] G. Fredenall, "Delay equalization in color television," *Proceedings of the IRE*, vol. 42, no. 1, pp. 258–262, 1954.

[35] S. Scanlan and J. Rhodes, "Microwave allpass networks-part I," *IEEE Transactions on Microwave Theory and Techniques*, vol. 16, no. 2, pp. 62–72, 1968.

[36] R. Crane, "All-pass network synthesis," *IEEE Transactions on Circuit Theory*, vol. 15, no. 4, pp. 474–478, 1968.

[37] G. Vergelli and D. Chakraborty, "Computer-aided group delay equalization for broad-band satellite transponder applications," *IEEE Transactions on Communications*, vol. 21, no. 10, pp. 1152–1156, 1973.

[38] R. Saal, "On the design of filters by synthesis," *IRE Transactions on Circuit Theory*, vol. 5, no. 4, pp. 284–327, 1958.

[39] Y. P. Tsividis, "Integrated continuous-time filter design - an overview," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 3, pp. 166–176, 1994.

[40] P. J. Osuch and T. Stander, "An on-chip post-production tunable group delay equaliser," in *International Conference on Actual Problems of Electron Devices Engineering, APEDE 2014*, 2014, pp. 177–184.

[41] C. Wilson and B. Wardorp, "Strip-Line Group Delay Equalisers," in *1st European Microwave Conference*, 1969, pp. 319–322.

[42] P. Aronhime, D. Nelson, J. Zurada, and C. Adams, "Realization of current-mode

complex pole all-pass networks using a single current conveyor," in *IEEE International Symposium on Circuits and Systems*, 1990, pp. 3193–3196.

[43] P. Aronhime, "Realizations of complex pole all-pass networks," *IEEE Transactions on Circuits and Systems*, vol. 4, no. 22, pp. 324–328, 1975.

[44] G. Gupta, S. V. Singh, and S. V. Bhooshan, "VDTA based electronically tunable voltage-mode and trans-admittance biquad filter," *Circuits and Systems*, vol. 93, no. 3, pp. 93–102, 2015.

[45] P. Aronhime, "Transfer-function synthesis using a current conveyor," *IEEE Transactions on Circuits and Systems*, vol. 21, no. 2, pp. 312–313, 1974.

[46] A. Toker, M. Discigil, O. Cicekoglu, and H. H. Kuntman, "Direct synthesis approach for voltage mode transfer functions using current conveyors," in *IEEE Asia-Pacific Conference on Circuits and Systems*, 1998, pp. 255–258.

[47] C. M. Chang, "Multifunction biquadratic filters using current conveyors," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 44, no. 11, pp. 956–958, 1997.

[48] J. W. Horng, "Voltage-mode universal biquadratic filters using CCIIs," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E87–A, no. 2, pp. 406–409, 2004.

[49] M. C. Chang, "Universal voltage-mode filter with four inputs and one output using two CCII's," *International Journal of Electronics*, vol. 86, no. 3, pp. 305–309, 1999.

[50] S. I. Liu and J. L. Lee, "Voltage-mode universal filters using two current conveyors," *International Journal of Electronics*, vol. 82, no. 2, pp. 145–150, 1997.

[51] J. W. Horng, M. H. Lee, H. C. Cheng, and C. W. Chang, "New CCII-based voltage-mode universal biquadratic filter," *International Journal of Electronics*, vol. 82, no. 2, pp. 151–156, 1997.

[52] J. W. Horng, "Novel universal voltage-mode biquad filter with three inputs and one output using only two current conveyors," *International Journal of Electronics*, vol. 80, no. 4, pp. 543–546, 1996.

[53] S. Özoğuz and E. O. Güneş, "Universal filter with three inputs using CCII+," *Electronics Letters*, vol. 32, no. 23, p. 2134, 1996.

[54] C. Chang and M. S. Lee, "Universal voltage-mode filter with three inputs and one output using three current conveyors and one voltage follower," *Electronics Letters*, vol. 30, no. 25, pp. 2112–2113, 1994.

[55] C. M. Chang and M. S. Lee, "Comment: universal voltage-mode filter with three inputs and one output using three current conveyors and one voltage follower," *Electronics Letters*, vol. 31, no. 5, p. 353, 1995.

[56] H. Rubin and R. Even, "Single-transistor all-pass networks," *IEEE Transactions on Circuit Theory*, vol. 20, no. 1, pp. 24–30, 1973.

[57] T. Yanagisawa, "RC active networks using current inversion type negative impedance converters," *IRE Transactions on Circuit Theory*, vol. 4, no. 3, pp. 140–144, 1957.

[58] P. Ziskka and M. Laipert, "Band-pass filter group delay equalization," in *IEEE Proceeding of the Thirty-Eighth Southeastern Symposium on System Theory*, 2006, pp. 289–293.

[59] M. Tian, V. Visvanathan, J. Hantgan, and K. Kundert, "Striving for small-signal stability," *IEEE Circuits and Devices Magazine*, vol. 17, no. 1, pp. 31–41, 2001.

[60] S. J. Mason, "Feedback theory - some properties of signal flow graphs," *Proceedings of the IRE*, vol. 41, no. 9, pp. 1144–1156, 1953.

[61] S. J. Mason, "Feedback theory: further properties of signal flow graphs," in *Proceedings of the IRE*, 1955, vol. 44, no. 7, pp. 920–926.

[62] R. Levy, "Realization of practical lumped element all-pass networks for delay equalization of RF and microwave filters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, no. 12, pp. 3307–3311, 2011.

[63] P. Ahmadi, B. Maundy, A. S. Elwakil, L. Belostotski, and A. Madanayake, "A new second-order all-pass filter in 130-nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 63, no. 3, pp. 249–253, 2016.

[64] N. Rajesh and S. Pavan, "Design of lumped-component programmable delay elements for ultra-wideband beamforming," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 8, pp. 1800–1814, 2014.

[65] C. Wijenayake, A. Madanayake, Y. Xu, L. Belostotski, and L. T. Bruton, "A Steerable DC-1 GHz all-pass filter-Sum RF space-time 2-D beam filter in 65 nm CMOS," in

IEEE International Symposium on Circuits and Systems, 2013, vol. 1, pp. 1276–1279.

[66] M. F. Quélhas and A. Petraglia, "Optimum design of group delay equalizers," *Digital Signal Processing: A Review Journal*, vol. 21, no. 1, pp. 1–12, 2011.

[67] P. Okoniewski and J. Piskorowski, "An efficient method of group delay equalization for digital IIR filters," *Metrology and Measurement Systems*, vol. 20, no. 3, pp. 395–406, 2013.

[68] E. G. Cristal, "Theory and design of transmission line all-pass equalizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 17, no. 1, pp. 28–38, Jan. 1969.

[69] H. T. Hsu, H. W. Yao, K. A. Zaki, and A. E. Atia, "Synthesis of coupled-resonators group-delay equalizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 8, pp. 1960–1968, 2002.

[70] H. Voorman and H. Veenstra, "Tunable high-frequency Gm-C filters," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 8, pp. 1097–1108, Aug. 2000.

[71] I. C. Hunter, L. Billonet, B. Jarry, and P. Guillon, "Microwave filters - applications and technology," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 794–805, 2002.

[72] D. Li and Y. Tsividis, "Design techniques for automatically tuned integrated gigahertz-range active LC filters," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 8, pp. 967–977, Aug. 2002.

[73] B. Georgescu, S. Member, H. Pekau, J. Haslett, and J. Mcrory, "Tunable coupled inductor Q-enhancement for parallel resonant LC tanks," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 50, no. 10, pp. 705–713, 2003.

[74] J. Kulyk and J. Haslett, "A monolithic CMOS 2368+30 MHz transformer based Q-enhanced series-C coupled resonator bandpass filter," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 2, pp. 362–374, 2006.

[75] T. Soorapanth and S. S. Wong, "A 0-dB IL 2140+30 MHz bandpass filter utilizing Q-enhanced spiral inductors in standard CMOS," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 5, pp. 579–586, 2002.

[76] B. Georgescu, I. G. Finvers, and F. Ghannouchi, "2 GHz Q-enhanced active filter with low passband distortion and high dynamic range," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 9, pp. 2029–2039, 2006.

[77] J. K. Nakaska and J. W. Haslett, "2 GHz automatically tuned Q-enhanced CMOS bandpass filter," in *IEEE International Microwave Symposium (MTTS)*, 2007, pp. 1599–1602.

[78] S. Gupta, D. L. Sounas, H. Van Nguyen, Q. Zhang, and C. Caloz, "CRLH–CRLH C-section dispersive delay structures with enhanced group-delay swing for higher analog signal processing resolution," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 12, pp. 3939–3949, 2012.

[79] X. Lin, J. Liu, H. Lee, and H. Liu, "A 2.5- to 3.5-Gb/s adaptive FIR equalizer with continuous-time wide-bandwidth delay line in 0.25-μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 8, pp. 1908–1918, 2006.

[80] M. Maeng *et al.*, "0.18-um CMOS equalization techniques for 10-Gb/s fiber optical communication links," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 11, pp. 3509–3519, 2005.

[81] S. K. Garakoui, E. A. Klumperink, B. Nauta, and F. E. van Vliet, "Compact cascadable gm-C all-pass true time delay cell with reduced delay variation over frequency," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 3, pp. 693–703, 2015.

[82] Y. W. Chang, T. C. Yan, and C. N. Kuo, "Wideband time-delay circuit," in *European Conference In Microwave Integrated Circuits*, 2011, pp. 454–457.

[83] I. Mondal and N. Krishnapura, "A 2-GHz bandwidth, 0.25-1.7 ns true-time-delay element using a variable-order all-pass filter architecture in 0.13 μm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 8, pp. 2180–2193, 2017.

[84] A. Soliman, "Inductorless realization of an all-pass transfer function using the current conveyor," *IEEE Transactions on Circuit Theory*, vol. 20, no. 1, pp. 80–81, 1973.

[85] C. Acar and S. Özoğuz, "High-order voltage transfer function synthesis using CCII+ based unity gain current amplifiers," *Electronics Letters*, vol. 32, no. 22, p. 2030, 1996.

[86] R. I. Salawu, "Realization of an all-pass transfer function using the second generation

current conveyor," *Proceedings of the IEEE*, vol. 68, no. 1, pp. 183–184, 1980.

[87] P. J. Osuch and T. Stander, "Wideband high-precision CMOS CCII+ with stability and peaking control," *IEEE Journal of Solid-State Circuits*, 2017.

[88] A. Fayed and M. Ismail, "On-chip resistors and capacitors," in *Adaptive Techniques for Mixed Signal System on Chip*, 2006, pp. 67–94.

[89] A. S. Sedra and K. C. Smith, "A second-generation current conveyor and its applications," *IEEE Transactions on Circuit Theory*, vol. 17, no. 1, pp. 132–134, 1970.

[90] R. Senani, D. R. Bhaskar, and A. K. Singh, *Current conveyors: variants, applications and hardware implementations*. Springer, 2014.

[91] F. Seguin, B. Godara, F. Alicalapa, and A. Fabre, "2.2 GHz all-n-p-n second-generation controlled conveyor in pseudoclass AB using 0.8-μm BiCMOS technology," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 51, no. 7, pp. 369–373, Jul. 2004.

[92] A. S. Sedra, G. W. Roberts, and F. Gohh, "The current conveyor: history, progress and new results," *Circuits, Devices and Systems, IEE Proceedings*, vol. 2, no. 137, pp. 78–87, 1990.

[93] W. Surakampontorn, V. Riewruja, K. Kumwachara, and K. Dejhan, "Accurate CMOS-based current conveyors," *IEEE Transactions on Instrumentation and Measurement*, vol. 40, no. 4, pp. 699–702, 1991.

[94] U. Yodprasit, "High-precision CMOS current conveyor," *Electronics Letters*, vol. 36, no. 7, 2000.

[95] A. Ben Saied, S. Ben Salem, and D. S. Masmoudi, "Low voltage high frequency CCII based multifunction filters and chaos generator," *Asian Journal of Applied Sciences*, vol. 4, no. 6, pp. 603–617, Jun. 2011.

[96] G. Palmisano and G. Palumbo, "A simple CMOS CCII+," *International Journal of Circuit Theory and Applications*, vol. 23, no. 6, pp. 599–603, 1995.

[97] K. Watanabe and H. W. Cha, "Wideband CMOS current conveyor," *Electronics Letters*, vol. 32, no. 14, pp. 1245–1246, 1996.

[98] W. S. Hassanein, I. A. Awad, and A. M. Soliman, "New high accuracy CMOS current

conveyors," *International Journal of Electronics and Communications (AEU)*, vol. 59, no. 7, pp. 384–391, 2005.

[99] S. Ben Salem, M. Fakhfakh, D. S. Masmoudi, M. Loulou, P. Loumeau, and N. Masmoudi, "A high performances CMOS CCII and high frequency applications," *Analog Integrated Circuits and Signal Processing*, vol. 49, no. 1, pp. 71–78, 2006.

[100] K. Moustakas and S. Siskos, "Improved low-voltage low-power class AB CMOS current conveyors based on the flipped voltage follower," in *Proceedings of the IEEE International Conference on Industrial Technology*, 2013, pp. 961–965.

[101] H. Mostafa and A. M. Soliman, "Novel low-power accurate wide-band CMOS negative-second-generation-current-conveyor realizations based on floating-current-source building blocks," in *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, 2009, pp. 720–725.

[102] A. Awad and A. M. Soliman, "New CMOS realization of the CCII-," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 4, pp. 460–463, 1999.

[103] E. Ibaragi, A. Hyogo, and K. Sekine, "A phase compensation technique without capacitors for the CMOS circuit with a very low impedance terminal," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 83, no. 2, pp. 236–242, 2000.

[104] F. Sequin and A. Fabre, "New second generation current conveyor with reduced parasitic resistance and bandpass filter application," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 6, pp. 781–785, 2001.

[105] M. Alami, "Second generation current conveyors with enhanced input resistance," *International Journal of Electronics*, vol. 86, no. 4, pp. 405–412, 1999.

[106] M. A. Yakout, A. I. Abdelfattah, and A. S. Elbazz, "BiCMOS current conveyor: design and application," in *IEEE International Symposium on Circuits and Systems*, 2000, vol. 3, pp. 463–466.

[107] A. Fabre, "Wideband translinear current convertor," *Electronics letters*, vol. 20, no. 6, pp. 241–242, 1984.

[108] F. Seguin and A. Fabre, "2 GHz controlled current conveyor in standard 0.8 μm BICMOS technology," *Electronics Letters*, vol. 37, no. 6, p. 329, 2001.

[109] M. Alami, A. Fabre, A. Jaadoud, and A. Touhami, "Simplified BiCMOS current controlled conveyor," in *14th IEEE International Conference on Electronics, Circuits, and Systems*, 2007, pp. 383–386.

[110] W. Surakampontorn and P. Thitimajshima, "Integrable electronically tunable current conveyors," *IEE Proceedings G (Electronic Circuits and Systems)*, vol. 135, no. 2, pp. 71–77, 1988.

[111] A. Fabre, "Translinear current conveyors implementation," *International Journal of Electronics Theoretical and Experimental*, vol. 59, no. 5, pp. 619–623, 1985.

[112] H. W. Bode, "Network analysis and feedback amplifier design," *Bell Telephone Laboratories series*. p. 551, 1945.

[113] P. J. Hurst, "A comparison of two approaches to feedback circuit analysis," *IEEE Transactions on Education*, vol. 35, no. 3. pp. 253–261, 1992.

[114] L. Tasny-Tschiassny, "The return-difference matrix in linear networks," *Proceedings of the IEE-Part IV: Institution Monographs*, vol. 100, no. 5, pp. 39–46, 1953.

[115] I. W. Sandberg, "On the theory of linear multi-loop feedback systems," *Bell System Technical Journal*, vol. 42, no. 2, pp. 355–382, 1963.

[116] L. Codecasa, "Novel feedback theory of electric circuits part II: loop invariants," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 59, no. 7, pp. 1505–1518, 2012.

[117] W. K. Chen, "The hybrid matrix in linear multiple-loop feedback networks," *IEEE Transactions on Circuits and Systems*, vol. 24, no. 9, pp. 469–474, 1977.

[118] I. Gresham *et al.*, "A compact manufacturable 76-77-GHz radar module for commercial ACC applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 1, pp. 44–58, 2001.

[119] S. Gupta, L. Zou, M. A. Salem, and C. Caloz, "Bit-error-rate (BER) performance in dispersion code multiple access (DCMA)," in *IEEE Antennas and Propagation Society, AP-S International Symposium*, 2015, pp. 1015–1016.

[120] S. Sinha, "A multi-dimensional spread spectrum transceiver," University of Pretoria,

2008.

[121] G. Avenier *et al.*, "0.13 um SiGe BiCMOS technology fully dedicated to mm-wave applications," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2312–2321, 2009.

[122] J. S. Dunn *et al.*, "Foundation of rf CMOS and SiGe BiCMOS technologies," *IBM Journal of Research and Development*, vol. 47, no. 2.3, pp. 101–138, 2003.

[123] R. Dalpatadu and K. Mouthaan, "Active transversal S-band filter using lumped elements in standard 0.13 μm CMOS," in *IEEE German Microwave Conference (GeMiC)*, 2012, pp. 1–4.

[124] Y. L. Lai and C. Y. Zheng, "A novel folded active inductor for wireless communication SoC," in *International SoC Design Conference*, 2013, pp. 306–309.

[125] H. Ramiah and T. Z. Zulkifli, "Design and optimization of integrated MOS varactors for high-tunability RF circuits," *IETE Journal of Research*, vol. 57, no. 4, pp. 346–350, 2011.

[126] P. Ahmadi, M. H. Taghavi, L. Belostotski, and A. Madanayake, "10–GHz current-mode 1st and 2nd order allpass filters on 130nm CMOS," in *IEEE 56th International Midwest Symposium on Circuits and Systems*, 2013, pp. 1–4.

[127] M. Hamouda, G. Fischer, R. Weigel, and T. Ussmueller, "A compact analog active time delay line using SiGe BiCMOS technology," in *2013 IEEE International Symposium on Circuits and Systems*, 2013, pp. 1055–1058.

[128] A. Ç. Ulusoy, B. Schleicher, and H. Schumacher, "A tunable differential all-pass filter for UWB true time delay and phase shift applications," *IEEE Microwave and Wireless Components Letters*, vol. 21, no. 9, pp. 462–464, 2011.

[129] G. Alpaydın, S. Balkır, and G. Dündar, "An evolutionary approach to automatic synthesis of high-performance analog integrated circuits," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 3, pp. 240–252, 2003.

ADDENDUM A CIRCUIT LAYOUTS

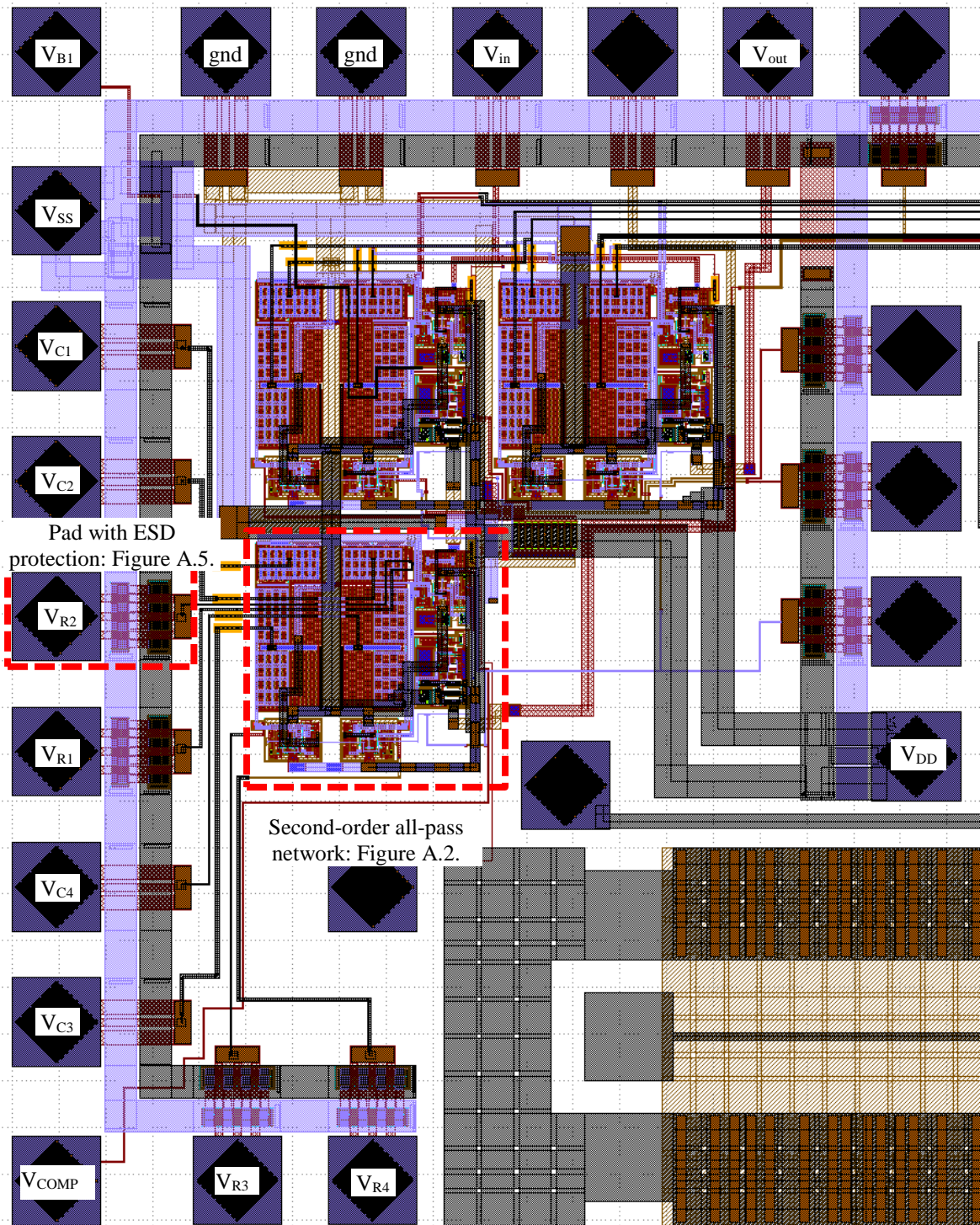## A.1 CAMERA-READY CIRCUIT LAYOUTS

Figure A.1. Camera-ready layout of a second-order all-pass network for the 0.35 µm CMOS technology node from ams AG – overall view including pads.
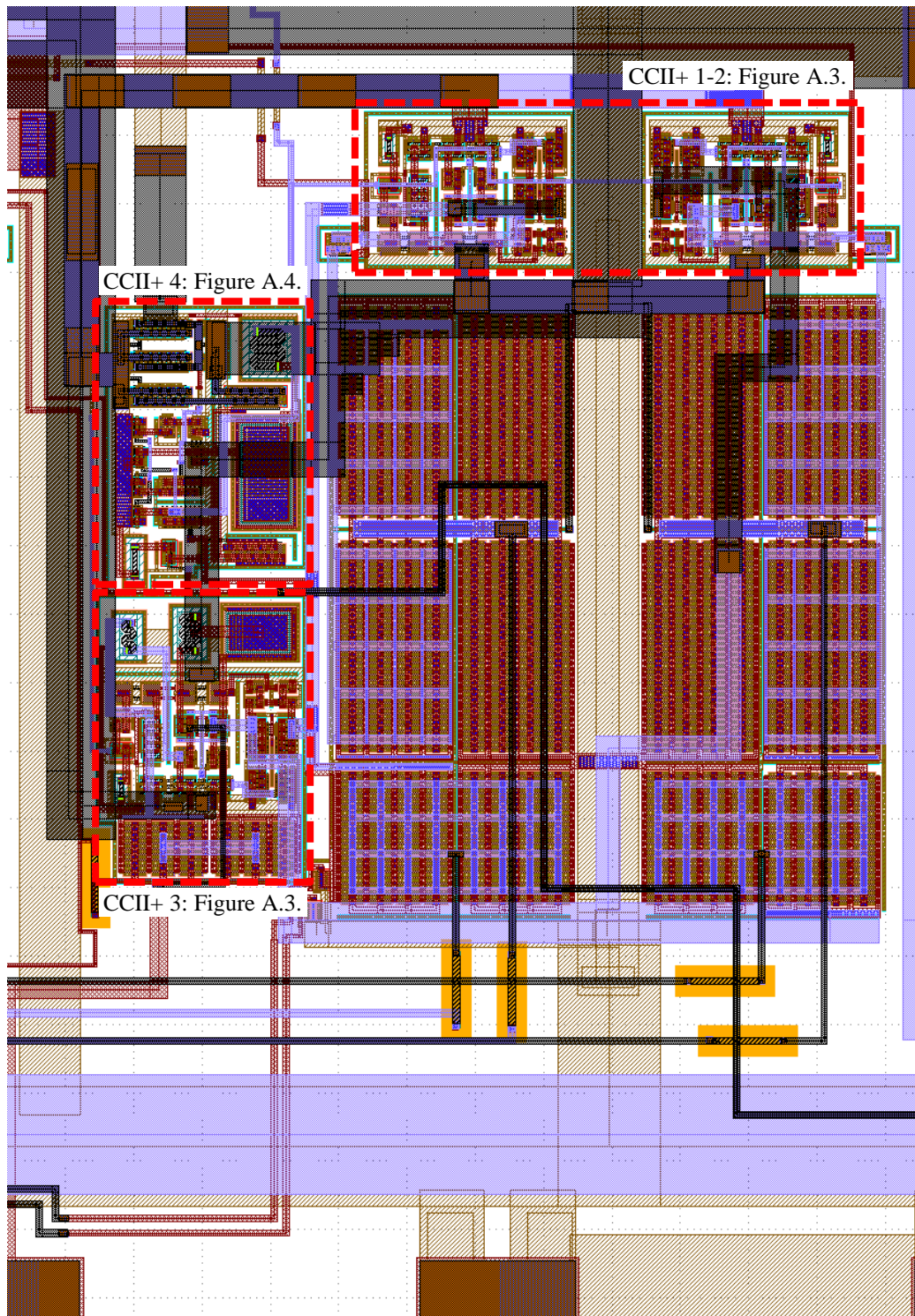
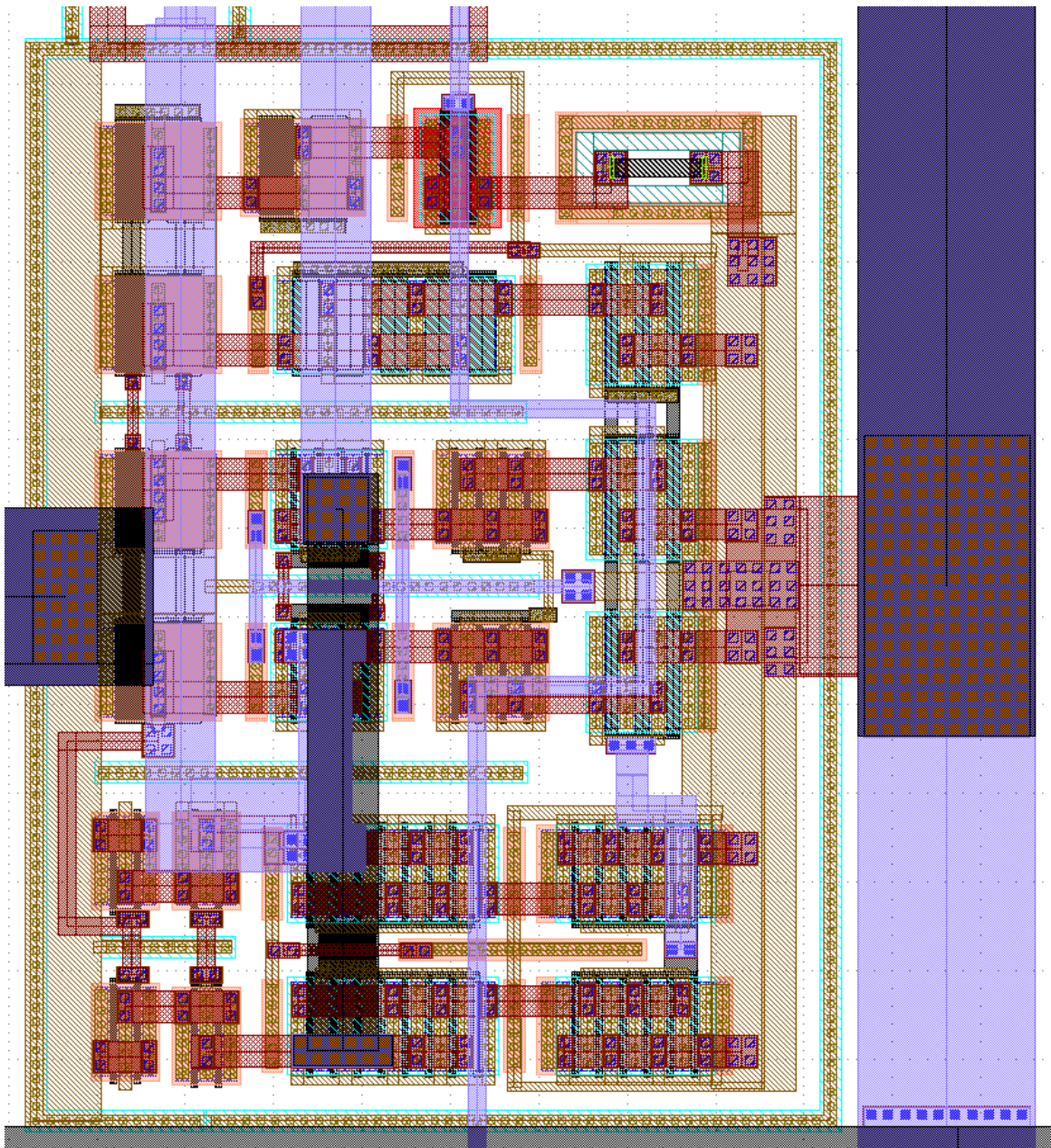**Figure A.2.** Camera-ready layout of a second-order all-pass network for the 0.35 μm CMOS technology node from ams AG.

Figure A.3. Camera-ready layout of the CCII+ used for the CMOS second-order all-pass network as per the circuit schematic of Figure 5.4.
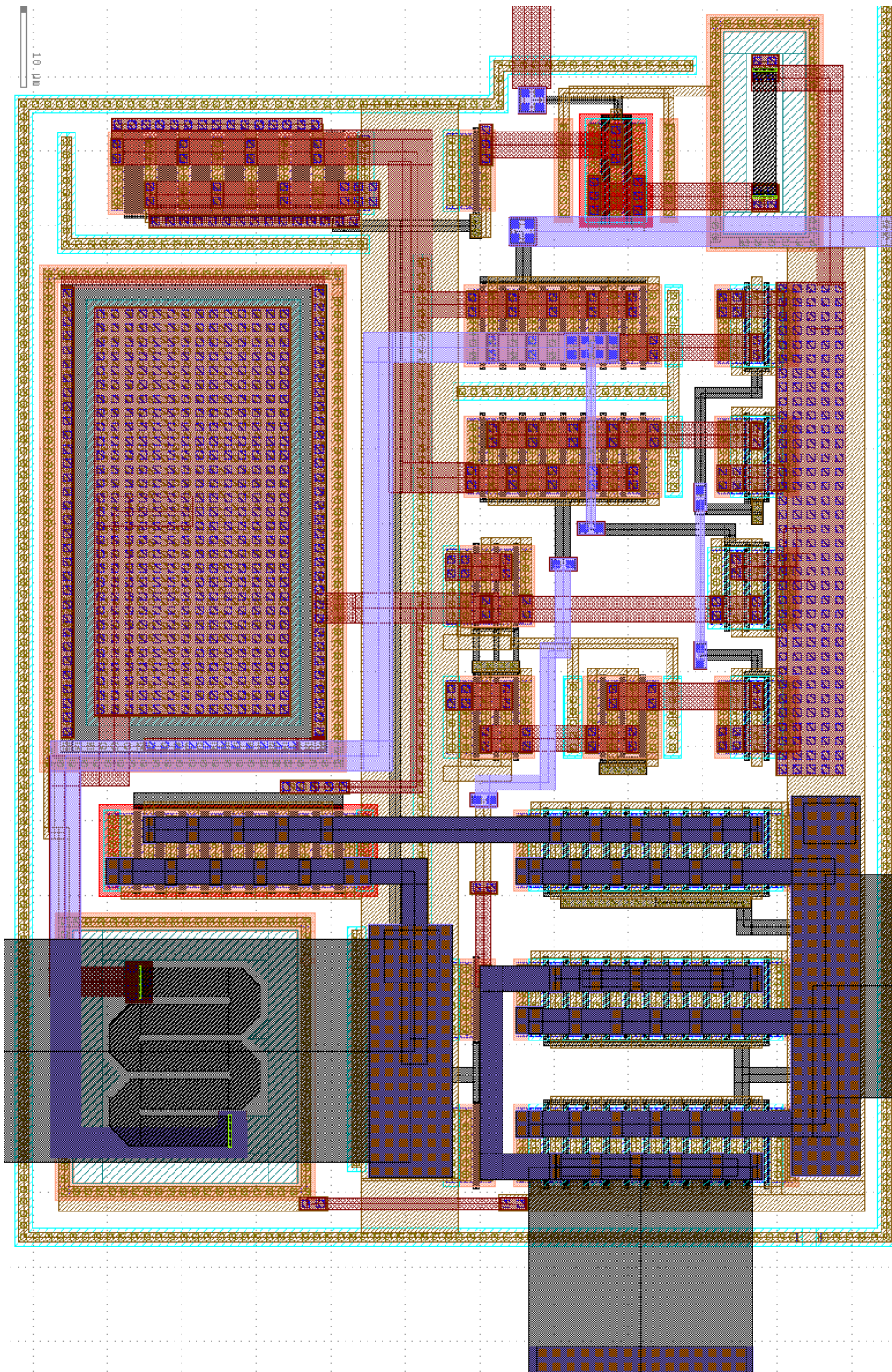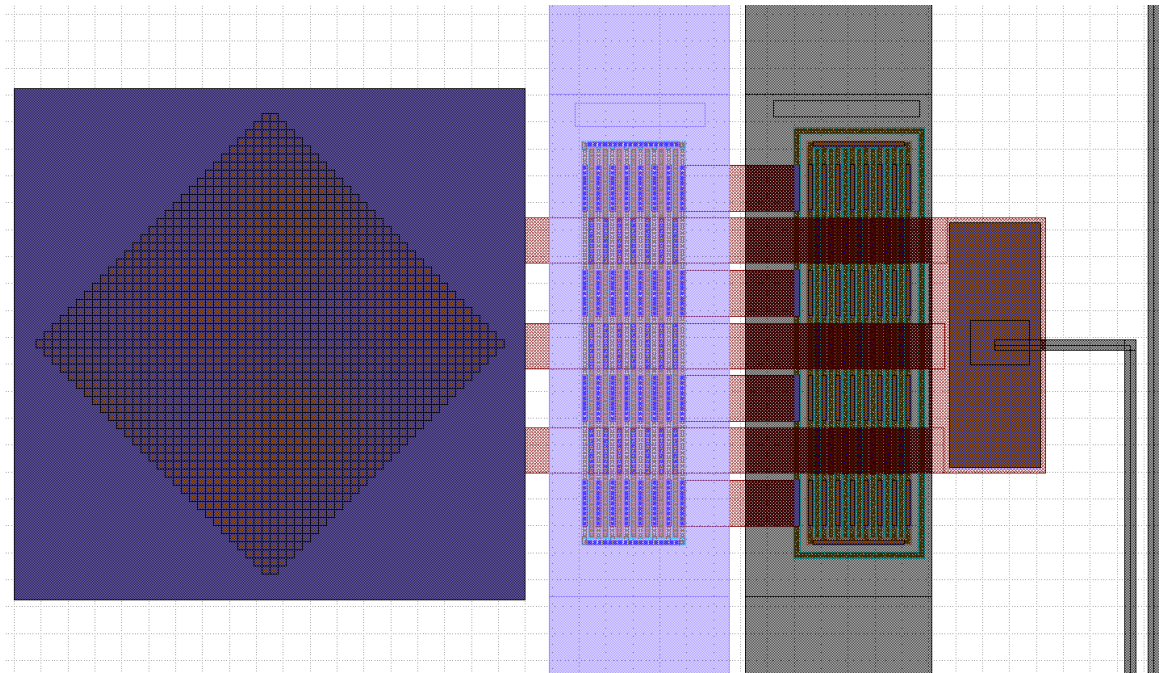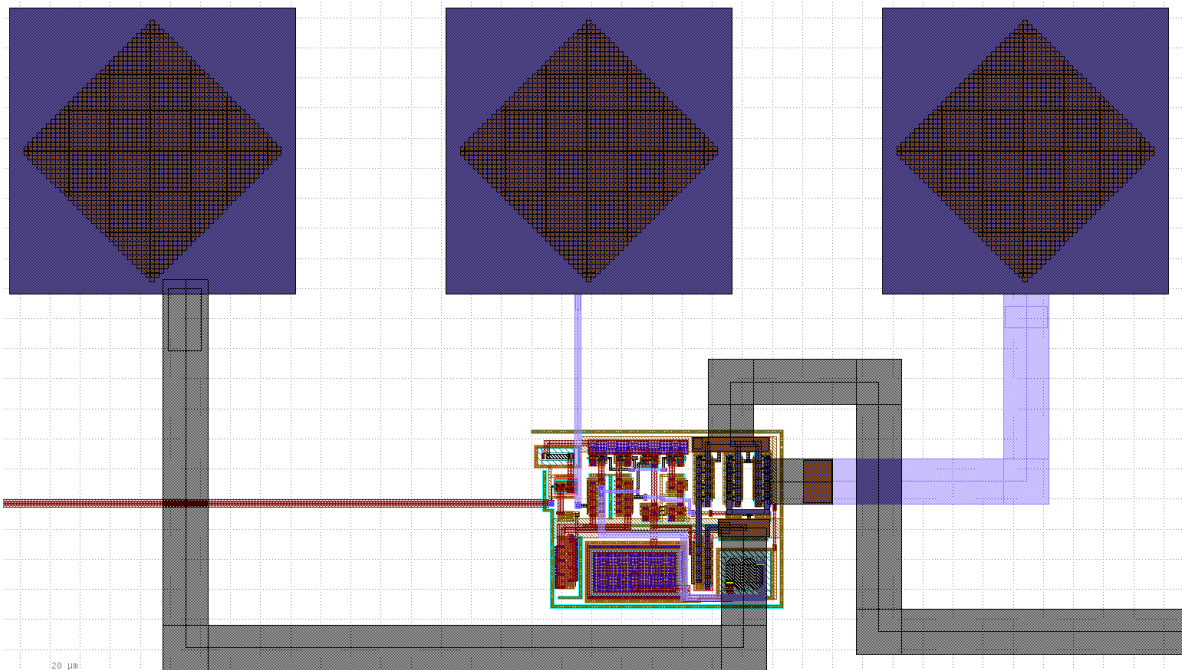
Figure A.4. Camera-ready layout of the high-precision CCII+ used for the CMOS second-order all-pass network as per the circuit schematic of Figure 5.14.

Figure A.5. Pad used for DC biasing with ESD diode protection.

Figure A.6. Camera-ready layout of the high-precision CCII+ used for the CMOS second-order all-pass network as per the circuit schematic of Figure 5.14 and corresponding to the micrograph in Figure 5.19.
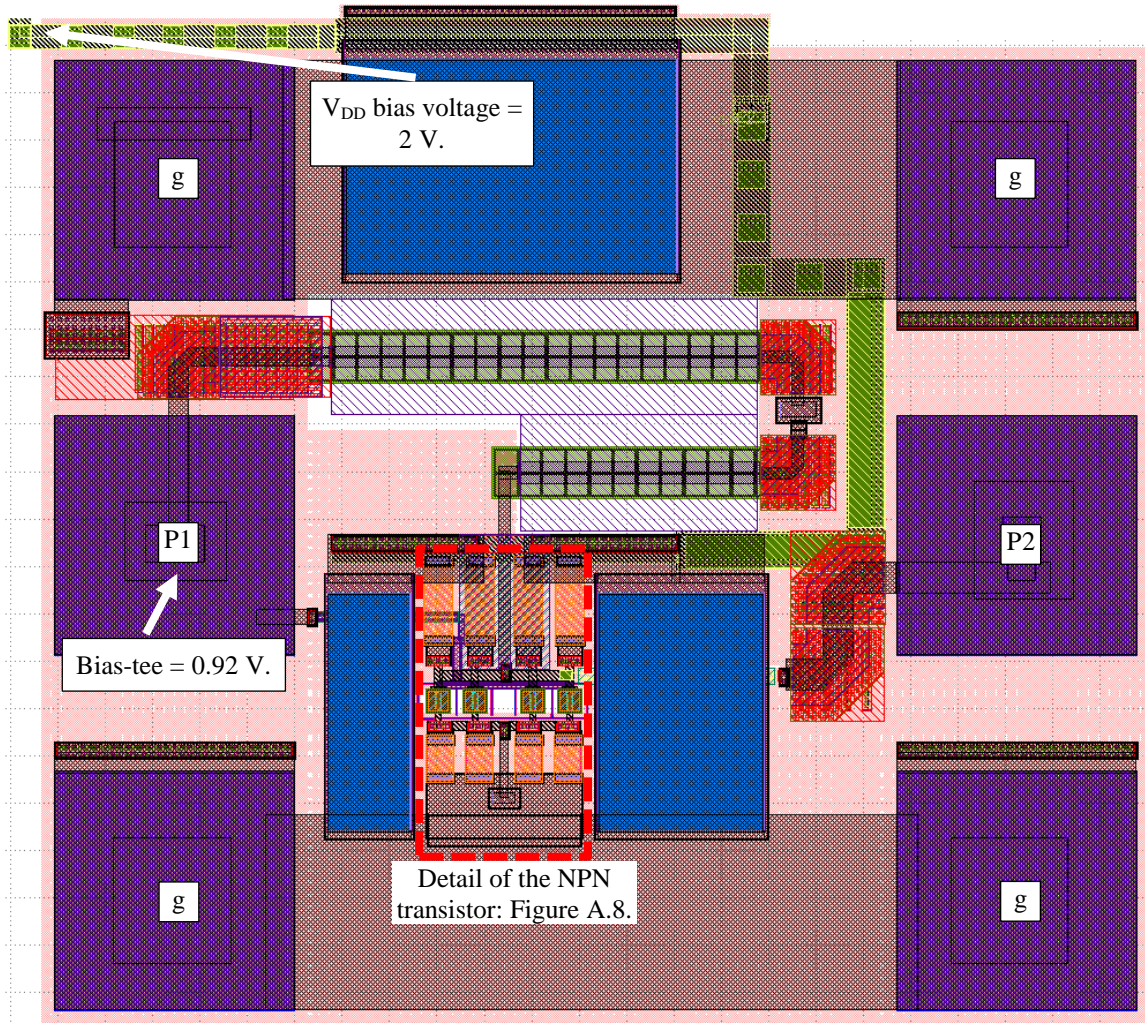
Figure A.7. Camera-ready layout of the BiCMOS second-order all-pass network as per the circuit schematic of Figure 5.26 and micrograph in Figure 5.31.
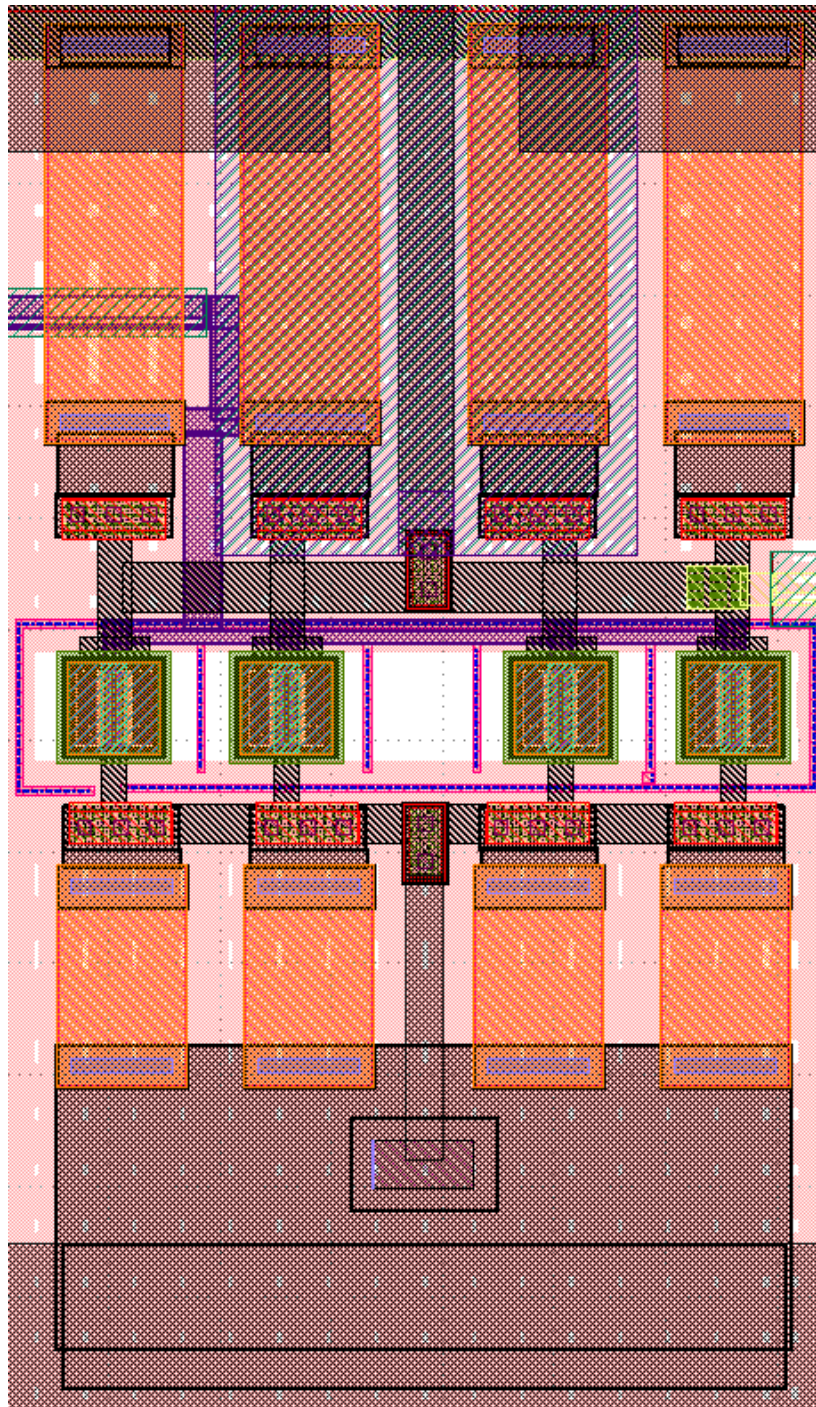
Figure A.8. Detail of the NPN transistor used in the design of Figure A.7.