Ontwerp van gebeurtenisgebaseerde automatische versterkingsregeling
en hogesnelheidsdatapad voor optische ontvangers met meerdere kanalen

Design of Event-Driven Automatic Gain Control
and High-Speed Data Path for Multichannel Optical Receiver Arrays

Jochen Verbrugghe

Promotoren: prof. dr. ir. X. Yin, prof. dr. ir. J. Bauwelinck
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Elektrotechniek

Vakgroep Informatietechnologie
Voorzitter: prof. dr. ir. D. De Zutter
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2014 - 2015

UNIVERSITEIT
GENT

ISBN 978-90-8578-787-7
NUR 959
Wettelijk depot: D/2015/10.500/30

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur
Vakgroep Informatietechnologie

Promotoren: prof. dr. ir. Xin Yin
prof. dr. ir. Johan Bauwelinck

Universiteit Gent
Faculteit Ingenieurswetenschappen en Architectuur

Vakgroep Informatietechnologie
Sint-Pietersnieuwstraat 41, B-9000 Gent, België

Tel.: +32 (9) 264 33 46
Fax.: +32 (9) 264 35 93

Dit werk werd financieel gesteund door een specialisatiebeurs van het IWT-Vlaanderen (agentschap voor Innovatie door Wetenschap en Technologie).

Proefschrift tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Elektrotechniek
Academiejaar 2014-2015

Voorwoord

Dit proefschrift kwam tot stand binnen de vakgroep Informatietechnologie van de Universiteit Gent en is het resultaat van een aantal jaren studie en onderzoek als medewerker van het Intec Design laboratorium. Ik wens dan ook uitdrukkelijk de vakgroepvoorzitter prof. dr. ir. Daniël De Zutter te bedanken voor de aangeboden infrastructuur en omkadering.

Ik wens zowel het oude als het nieuwe hoofd van het labo, prof. dr. ir. Jan Vandewege en prof. dr. ir. Johan Bauwelinck, laatstgenoemde tevens ook copromotor van dit werk, te bedanken voor de geboden kansen en het gestelde vertrouwen. Een woord van dank gaat verder uit naar prof. dr. ir. Xin Yin, promotor van dit werk, en dr. ir. Guy Torfs voor de leerrijke discussies en het nauwgezet nalezen. Ik ben prof. dr. ir. Xing-Zhi Qiu erkentelijk voor het onderhouden van contacten met de technologische industrie alsook het agentschap voor Innovatie door Wetenschap en Technologie (IWT) voor de toegekende beurs.

In het bijzonder ben ik ir. Bart Moeneclaey dankbaar voor zijn bijdrage in de ontwikkelde chips en de gemeenschappelijke interesse in snelle analytische technieken, waarbij de vele discussies leidden tot diepere inzichten of zelfs ontspoorden tot heuse zijprojecten.

Verder wens ik mijn ex-collega's en collega's van de Intec Design groep te bedanken voor de samenwerking tijdens deze jaren. Bovenal was ing. Jan Gillis onmisbaar voor de technische ondersteuning, PCB's, layout of gewoon zomaar. Ook denk ik met name aan dr. ir. Christophe Van Praet, dr. ir. Guy Torfs, dr. ir. Jasmien Put en ir. Wouter Soenen voor de uitgebreide discussies, aangename koffiepauzes en het gezelschap.

Een speciaal woord van dank gaat ook uit naar dr. ir. Renato Vaernewyck en ir. Xiao Li voor hun bijdrage in het PCB ontwerp en naar dr. ir Bart Baekelandt voor de hulp bij de beursaanvraag.

Mijn ouders verdienen ook zeker een plaats in dit dankwoord. Niet in het minst voor het sponseren van een Elektuurabonnement tijdens mijn jeugd en om mij de kans te geven om me verder te verdiepen in elektronica.

Tenslotte bedank ik mijn vriendin, die mijn voortdurend groeiende verzameling antieke meetapparatuur moet tolereren.

Gent, December 2014
Jochen Verbrugghe

Contents

Glossary

ADC	analog to digital converter
AGC	automatic gain control
AGM	asymptotic gain model
AOC	active optical cable
APD	avalanche photodetector
AWG	arrayed waveguide grating
BER	bit-error ratio
BNG	broadband network gateway
CDN	content delivery network
CDR	clock-and-data recovery
CID	consecutive identical digits
CML	current-mode logic
CT	chain theorem
CW	continuous-wave
CWDM	coarse wavelength division multiplexing
D-OA	design-oriented analysis
DAC	digital to analog converter
DIDF	dual-input describing function
DNL	differential non-linearity
dnti	double-null triple-injection
DSP	digital signal processing
DT	dissection theorem
DUT	device under test
DWDM	dense wavelength division multiplexing
EDFA	erbium doped fiber amplifier
EET	extra element theorem
ER	extinction ratio
FP7	Seventh Framework Programme
GBW	gain-bandwidth product
GFT	general feedback theorem

GNT	general network theorem
HDL	hardware description language
IIDF	incremental-input describing function
INL	integral non-linearity
ISI	inter-symbol interference
ISP	internet service provider
LCA	lightwave component analyzer
LHP	left hand-side plane
LTI	linear time-invariant
LUT	look-up table
MA	main amplifier
MIM	metal-insulator-metal
ndi	null double-injection
NRZ	non-return-to-zero
ODB	optical duobinary
OOK	on-off keying
PAM	pulse-amplitude modulation
PON	passive optical network
PP	power penalty
PRBS	pseudo-random bit sequence
PSD	power spectral density
pss	periodic steady-state
PVT	process, supply and temperature
RHP	right hand-side plane
S2D	single-ended to differential
SI	strong inversion
si	single-injection
SNR	signal-to-noise ratio
SOI	silicon-on-insulator
SPI	serial peripheral interface
TIA	transimpedance amplifier
VGA	variable gain amplifier
WDM	wavelength division multiplexing
WI	weak inversion

Nederlandstalige Samenvatting

Het internet heeft zonder enige twijfel ons leven ingrijpend veranderd. Het stijgend aanbod aan nieuwe online toepassingen, binnen sectoren zoals amusement, handel, industrie en gezondheidszorg vraagt steeds meer bandbreedte en stelt steeds grotere eisen aan de kwaliteit van de netwerk- en ICT-infrastructuur. Vooral de explosieve groei van online video en sociale media vereist grotere datasnelheden. Het net levert inhoud voor consumptie aan, in plaats van louter in connectiviteit tussen machines te voorzien. Het efficiënt schalen van netwerken en datacentra met beperkte kost vormt hierbij een grote uitdaging. Daarbij mag het vermogenverbruik maar beperkt stijgen met de bandbreedte, om extra kosten te vermijden en om de uitstoot van broeikasgassen te beperken. Hierbij moet er wel voor gezorgd worden dat de kwaliteit van de aangeboden diensten niet afneemt.

In de huidige architectuur van het internet zijn eindgebruikers verbonden met het openbare netwerk via het toegangsnetwerk van de lokale internetaanbieder. Traditioneel zijn deze toegangsnetwerken gebaseerd op de bestaande koper- of coaxiale verbindingen, maar tegenwoordig worden er nieuwe passieve optische netwerken (PONs) ontplooid. Deze PONs maken gebruik van optische vezel waardoor ze veel hogere datasnelheden kunnen aanbieden, voor een fractie van het vermogenverbruik. Het dataverkeer van de toegangsnetwerken wordt via Ethernetswitches en breedbandnetwerkgateways doorgesluist naar het ruggengraatnetwerk. Het geheel van deze regionale verbindingen is het metronetwerk. Datacentra zijn via een eigen router verbonden met het ruggengraatnetwerk.

Er zijn verschillende initiatieven in het leven geroepen om een oplossing voor het probleem van de vraag naar stijgende datasnelheden te vinden, met aandacht voor de ecologische en economische impact. Het werk beschreven in deze thesis werd uitgevoerd in de context van twee projecten in het kader van het Europese FP7 programma. Het doel van beide projecten is het ontwikkelen van hooggeïntegreerde opto-elektronische componenten die hogere datasnelheden ondersteunen met laag vermogenverbruik, maar elk concentreert zich op een ander aspect van de netwerkarchitectuur. Mirage focust zich op datacentra, terwijl C3PO zich toelegt op middellange-afstandsnetwerken, zoals het metronetwerk. In het bijzonder besteedt dit werk aandacht aan twee aspecten van de optische ontvangers: ten eerste het vergroten van het dynamisch bereik van een lineaire ontvanger voor modulatieformaten met meerdere niveaus; ten tweede de integratie van meerdere kanalen op eenzelfde chip met beperkte oppervlakte.

Datacentra bieden de eindgebruiker opslagruimte, rekenkracht en software aan op afstand. Ze bestaan uit gateway routers, een lokaal netwerk, servers en opslag, waarbij tot 100 000 apparaten geïntegreerd worden in standaardrekken. Ook hier is schaalbaarheid een groot probleem. Er zijn ingrijpende veranderingen nodig op het vlak van interconnectie en vermogenverbruik om tot een duurzame oplossing te komen. In het bijzonder

leidt een lager vermogenverbruik ook tot een reductie in koelkosten. De verschillende rekken zijn onderling verbonden met actieve optische kabels (AOCs), die op dit moment snelheden tot 25 Gb/s met aan-uitmodulatie (NRZ) ondersteunen. Om hogere datasnelheden over langere kabels te ondersteunen, kunnen een aantal verbeteringen doorgevoerd worden. Daarbij is vooral het gebruik van modulatieformaten met meerdere niveaus interessant. Zo'n modulatieformaten gebruiken de beschikbare bandbreedte efficiënter dan NRZ, maar vereisen wel een lineaire voorversterker in de ontvanger. Het grootste deel van deze dissertatie beschrijft het ontwerp en implementatie van een gebeurtenisgebaseerde automatische versterkingsregeling (AGC) voor het datapad van een lineaire transimpedantieversterker voor hoge snelheden.

Binnen de topologie van het internet bevindt het metronetwerk zich tussen de toegangsnetwerken en het ruggengraatnetwerk. Het strekt zich uit over afstanden tot 500 km. Studies geven aan dat het metro-dataverkeer in de nabije toekomst sterk zal verhogen. Dit is vooral te wijten aan het toenemende verkeer, vooral videostreaming, van content delivery netwerken en bijhorende datacentra die rechtstreeks op de lokale metronetwerken worden aangesloten. Deze evolutie gaat gepaard met een vraag naar hogere datasnelheden. De meest voor de hand liggende oplossing is het gebruik van golflengtemultiplexering met hoge dichtheid (DWDM), waardoor een enkele glasvezel veel hogere datasnelheden kan ondersteunen. Aan de ontvangerkant wordt elke golflengte van het gedemultiplexeerde invallende licht gekoppeld in een fotodiode van een fotodioderooster, die op zijn beurt verbonden is met een kanaal van een meerkanaals ontvanger. Om veel kanalen te ondersteunen moet de fysieke oppervlakte van de kanalen klein zijn. Bovendien dient het vermogenverbruik per kanaal beperkt te zijn om de temperatuur laag genoeg te houden zonder koeling. In het tweede deel van dit werk wordt dan ook dieper ingegaan op de implementatie van een vierkanaals ontvanger, voor een datasnelheid van 4×25 Gb/s, met laag vermogenverbruik en beperkte oppervlakte. De afstand tussen de kanalen is gelijk aan die van een standaard fotodioderooster, 250 µm.

Hoofdstuk 1 beschrijft de impact van de stijgende vraag naar bandbreedte in combinatie met lager vermogenverbruik op het ontwerp van de optische ontvanger, zowel in de context van middellange verbindingen (metronetwerken) als van korte verbindingen (datacentra). Daarnaast wordt een korte beschrijving gegeven van een aantal aspecten van geïntegreerd analoog ontwerp: de ontwerpsmethodiek, de modellen voor handberekeningen van de transistoren en een hulpprogramma. Ook wordt een overzicht van de gebruikte technologie gegeven.

Hoofdstuk 2 definieert een aantal essentiële concepten in verband met optische ontvangers. Die zijn nodig voor een goed begrip van de verdere tekst.

De basisprincipes van tegengekoppelde AGC systemen worden aangebracht in het eerste deel van Hoofdstuk 3. Een basismodel in continuë tijd is beschreven, waarbij de instelbare versterker (VGA) het datapad van een lineaire optische ontvanger is. Om modulatieformaten met meerdere niveaus goed te kunnen ontvangen, moet de VGA een goed gecontroleerde frequentiekarakteristiek hebben. In het bijzonder moet doorschot in het tijdsdomein beperkt worden, en dit voor het volledige instelbereik. Er wordt betoogd dat dit moeilijk volledig analoog te implementeren valt. Daarom wordt een gebeurtenisgebaseerde uitbreiding van het continuëtijdsmodel voorgesteld, waarbij zowel de structurele aspecten als het dynamisch gedrag besproken worden. Het resultaat is een systeemmodel van een gekwantiseerde AGC lus, dat als basis dient voor het ontwerp op systeemniveau in Hoofdstuk 4. De gedetailleerde implementatie op circuitniveau wordt verder uitgewerkt in Hoofdstuk 5, waarbij experimentele resultaten de haalbaarheid van de voorgestelde

structuur bevestigen.

Hoofdstuk 6 beschrijft het ontwerp en de implementatie van een 4×25 Gb/s optische ontvanger met meerdere kanalen, voor NRZ modulatie met kleine chipoppervlakte. Er wordt vooral aandacht besteed aan de ingangstrap, waarbij technieken worden beschreven die de bandbreedte en het dynamisch bereik vergroten. Meetresultaten voor NRZ en optische duobinary modulatie worden voorgelegd, alsook de invloed van overspraak op de prestaties.

Het laatste hoofdstuk 7 geeft een overzicht van de belangrijkste conclusies en suggereert een aantal onderwerpen voor verder onderzoek.

Op het einde van het werk zijn twee appendices opgenomen. Appendix A geeft een overzicht van het general network theorem, een theorema dat vaak gebruikt wordt in dit werk en ook numeriek werd geïmplementeerd. De resultaten van Appendix B, een analyse van een tweetrapsopamp, gecompenseerd met capaciteitsvermenigvuldiger, zijn gebruikt bij het ontwerp van een bouwblok in het AGC systeem.

English Summary

The internet has become the ubiquitous tool that has transformed the lives of all of us. New broadband applications in the field of entertainment, commerce, industry, healthcare and social interactions demand increasingly higher data rates and quality of the networks and ICT infrastructure. In addition, high definition video streaming and cloud services will continue to push the demand for bandwidth. These applications are reshaping the internet into a content-centric network. The challenge is to transform the telecom optical networks and data centers such that they can be scaled efficiently, at low cost. Furthermore, from both an environmental and economic perspective, this scaling should go hand in hand with reduced power consumption. This stems from the desire to reduce CO_2 emission and to reduce network operating costs while offering the same service level as today.

In the current architecture of the internet, end-users connect to the public network using the access network of an internet service provider (ISP). Today, this access network either reuses the legacy copper or coaxial network or uses passive optical network (PON) technologies, among which the PON is the most energy efficient and provides the highest data rates. Traffic from the access network is aggregated with Ethernet switches and routed to the core network through the provider edge routers, with broadband network gateways (BNGs) to regulate access and usage. These regional links are collectively called the metro network. Data centers connect to the core network using their own dedicated gateway router.

The problem of increasing data rates, while reducing the economic and environmental impact, has attracted considerable attention. The research described in this work has been performed in the context of two projects part of the European Union Seventh Framework Programme (FP7), which both aim for higher data rates and tight integration while keeping power consumption low. Mirage targets data center applications while C3PO focuses on medium-reach networks, such as the metro network. Specifically, this research considers two aspects of the high-speed optical receivers used in the communication networks: increasing dynamic range of a linear receiver for multilevel modulation through automatic gain control (AGC) and integration of multiple channels on a single chip with a small area footprint.

The data centers of today are high-density computing facilities that provide storage, processing and software as a service to the end-user. They are comprised of gateway routers, a local area network, servers and storage. All of this is organized in racks. The largest units contain over 100 000 servers. The major challenges regarding data centers are scalability and keeping up with increasing amounts of traffic while reducing power consumption (of the devices as well as the associated cooling) and keeping cost minimal. Presently, racks are primarily interconnected with active optical cables (AOCs) which

employ signal rates up to 25 Gb/s per lane with non-return-to-zero (NRZ) modulation. A number of technological developments can be employed in AOCs of the future to provide terabit-capacity optical interconnects over longer distances. One such innovation is the use of multilevel modulation formats, which are more bandwidth-efficient than traditional NRZ modulation. Multilevel modulation requires a linear amplifier as front-end of the optical receiver. The greater part of this dissertation discusses the design and implementation of an AGC system for the data path of a linear transimpedance amplifier (TIA).

The metro network is the intermediate regional network between the access and core network of the internet architecture, with link lengths up to 500 km. It is estimated that in the near future metro-traffic will increase massively. This growth is attributed mainly to increasing traffic from content delivery networks (CDNs) and data centers, which bypass the core network and directly connect to the metro network. Internet video growth is the major reason for traffic increase. This evolution demands increasingly higher data rates. Today, dense wavelength division multiplexing (DWDM) is widely recognized as being necessary to provide data capacity scalability for future optical networks, as it allows for much higher combined data rates over a single fiber. At the receiver, each wavelength of the demultiplexed incoming light is coupled to a photo diode in a photo diode array which is connected to a dedicated lane of a multichannel receiver. The high number of channels requires small physical channel spacing and tight integration of the diode array with the receiver. In addition, active cooling should be avoided, such that power consumption per receiver lane must be kept low in order not to exceed thermal operation limits. The second component of this work presents the development of an integrated four-channel receiver, targeting 4×25 Gb/s data rate, with low power consumption and small footprint to support tight integration with a p-i-n photo diode array with a 250 µm channel pitch.

Chapter 1 discusses the impact of increasing data rates and the desire to reduce power consumption on the design of the optical receiver component, in wide metropolitan area networks as well as in short-reach point-to-point links in data centers. In addition, some aspects of integrated analog circuit design are highlighted: the design flow, transistor hand models, a software design tool. Also, an overview of the process technology is given.

Chapter 2 provides essential optical receiver concepts, which are required to understand the remainder of the work.

Fundamentals of feedback AGC systems are discussed in the first part of Chapter 3. A basic system model is presented in the continuous-time domain, in which the variable gain amplifier (VGA) constitutes the multistage datapath of a linear optical receiver. To enable reliable reception of multilevel modulation formats, the VGA requires controlled frequency response and in particular limited time-domain overshoot across the gain range. It is argued that this control is hard to achieve with fully analog building blocks. Therefore, an event-driven approach is proposed as an extension of the continuous-time system. Both the structural and behavioral aspects are discussed. The result is a system model of a quantized AGC loop, upon which the system-level design, presented in Chapter 4, is based. In turn, Chapter 5 discusses the detailed implementation of the various building blocks on the circuit level and presents experimental results that confirm the feasibility of the proposed approach.

Chapter 6 discusses the design and implementation of a 4×25 Gb/s optical receiver array for NRZ modulation with a small area footprint. The focus lies on the input stages and techniques to extend bandwidth and dynamic range are presented. Measurement results for NRZ and optical duobinary (ODB) modulation are presented, as well as the

influence of crosstalk on the performance.

Finally, Chapter 7 provides an overview of the foremost conclusions of the presented research and includes suggestions for future research.

Two appendices are included. Appendix A gives an overview of the general network theorem (GNT), which is used throughout this work and which has been implemented numerically. The results from Appendix B, the analysis of a two-stage opamp compensated with capacitance multipliers, were used to design a building block for the AGC system.

Chapter 1

Introduction

This introductory chapter provides the background against which the research described in this work is performed. An overview of today's network topology and its challenges are given. In particular, in Section 1.1, the impact of increasing data rates and the desire to reduce power consumption on the design of the optical receiver component, in wide metropolitan area networks, as well as in short-reach point-to-point links in data centers, is discussed. In Section 1.2, some aspects of the analog design flow used throughout this work are discussed, while the used process technology is shortly described in Section 1.3. Finally, an outline of this dissertation is given in Section 1.4.

1.1 Background

The internet has become the ubiquitous tool that has transformed the lives of all of us. New broadband applications in the field of entertainment, commerce, industry, healthcare and social interactions demand increasingly higher data rates and quality of the networks and ICT infrastructure. In addition, high definition video streaming and cloud services will continue to push the demand for bandwidth [1]. These applications are reshaping the internet into a content-centric network. The challenge is to transform the telecom optical networks and data centers such that they can be scaled efficiently, at low cost. Furthermore, from both an environmental and an economic perspective, this scaling should go hand in hand with reduced power consumption. This stems from the desire to reduce CO_2 emission and to reduce network operating costs while offering the same service level as today [2,3].

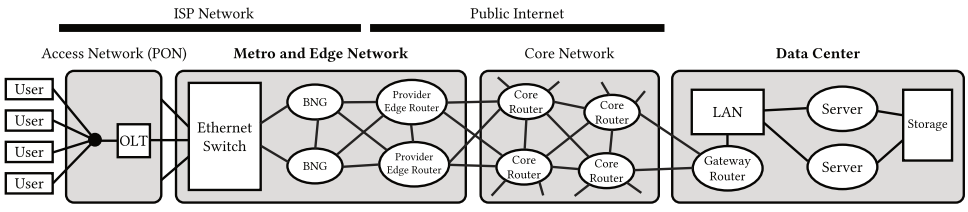


Figure 1.1: Architecture of the internet as it is today [4].

In the current architecture of the internet (Fig. 1.1), end-users connect to the public

network using the access network of an internet service provider (ISP) [4]. Today this access network either reuses the legacy copper or coaxial network or uses passive optical network (PON) technologies, among which the PON is the most energy efficient and provides the highest data rates [3, 5]. Traffic from the access network is aggregated with Ethernet switches and routed to the core network through the provider edge routers, with broadband network gateways (BNGs) to regulate access and usage. These regional links are collectively called the metro network. The core network comprises high-capacity transport networks and core routers, mainly using optical wavelength division multiplexing (WDM) fiber links. Data centers connect to the core network using their own dedicated gateway router. They are comprised of server and networking infrastructure: servers, storage, local area network and gateway routers. The services offered to the end-user include remote storage, remote software and remote processing [6].

In 2011, the internet was estimated to consume between 170 GW and 307 GW. Although this is a small fraction of the global power consumption, 1.1 % to 1.9 %, its relative importance will only increase as the internet will continue to provide substitutes for other functions of society that use much more energy [7]. An estimated 2 % to 10 % of the human carbon emission comes from ICT, of which 37 % is due the production of the energy required to operate the telecom infrastructure and devices. The metro/access networks are responsible for 60 % of the total power consumption of the communication networks [8].

In 2013, data centers in the U.S. consumed an estimated 91×10^9 kW h of energy, equivalent to an average power consumption of 10.5 GW and is expected to increase to roughly 16 GW in 2020 [9]. The SMARTer 2020 report, a widely recognized study, projects the global data center emissions until 2020 and expects these will grow 7 % year-on-year [10].

The problem of increasing data rates, while reducing the economic and environmental impact, has attracted considerable attention. The research described in this work has been performed in the context of two projects part of the European Union Seventh Framework Programme (FP7), which both aim for higher data rates and tight integration while keeping power consumption low. Mirage targets data center applications while C3PO focuses on medium-reach networks, such as the metro network.

1.1.1 Data Centers

In the content-centric internet of today, millions of users demand instant access to vast amounts of data. Data centers are high-density computing facilities that provide storage, processing and software as a service to the end-user [4]. In recent years, they have received significant attention due to the migration to cloud computing. They serve as hosting sites for multi-billion services such as video content distribution, social networking and large-scale computations [11].

Data centers are comprised of gateway routers, a local area network, servers and storage. The devices are organized in racks, spanning up to 2 km distance (short-reach), and are interconnected with fast optical links. Today's largest data centers contain over 100 000 servers [12]. Cisco predicts that global data center traffic will increase by 2.8 times by 2018, reaching 715 exabytes per month [13]. By then, 78 % of all workloads will be processed in the cloud, nearly double compared to 2013. This is mainly attributed to increased use of cloud services. Figures 1.2 and 1.3 show the projected adoption of residential and business services and growth of online services worldwide in 2018. These figures indicate that online video streaming has a major impact on IP traffic.

Data centers consume enormous amounts of power, with associated cost and envi-

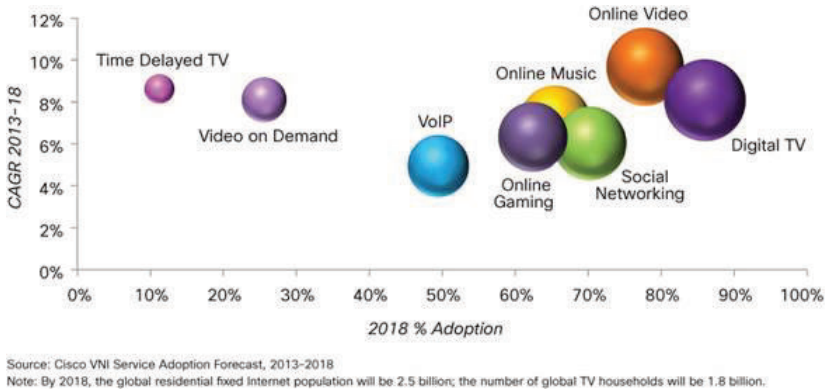


Figure 1.2: Residential services adoption and growth [14]

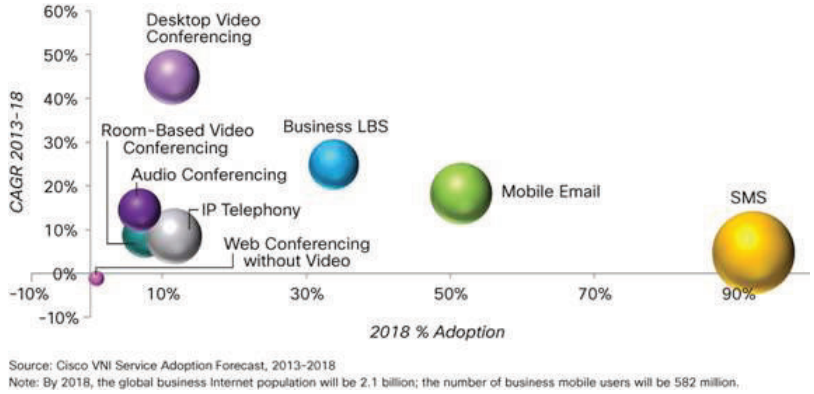


Figure 1.3: Business services adoption and growth [14]

ronmental impact. In 2005, data centers worldwide consumed 0.8 % of the total electricity consumption, amounting to $7.2 billion per year [15]. Typically, 50 % to 60 % of the energy is consumed in cooling and other overhead [4,16]. In 2010, it was estimated that data center electricity usage likely accounted for between 1.1 % and 1.5 % of the total electricity use, with associated emissions equivalent to 70 to 90 500 MW coal-fired power plants [17]. While a reduction in the pace of growth from previous estimates can be observed, the continued expansion of the industry means that the energy consumption and emission of greenhouse gasses will continue to grow [9].

The major challenges regarding data centers are scalability and keeping up with increasing amounts of traffic while reducing power consumption (of the devices as well as associated cooling) and keeping cost minimal. In order to offer a fast, reliable service, data centers are often geographically distributed with high-capacity optical links between the locations. As cloud services become more widespread and data rates increase, the energy consumption of the cloud devices and network will grow.

In today's data centers, racks are primarily interconnected with active optical cables (AOCs), used to carry hundreds of gigabits through hundreds of meters. Current standards (100G Ethernet, InfiniBand EDR, 32G Fibre Channel, PSM4) employ signal rates up to 25 Gb/s per lane with non-return-to-zero (NRZ) modulation. A number of technological

developments can be employed in AOCs of the future to provide terabit-capacity optical interconnects over longer distances:

Long wavelengths Wavelengths of 1310 nm or 1550 nm suffer less attenuation than the traditionally used 850 nm, enabling longer-reach networks [18].

Single-mode fiber Multi-mode fiber suffers more from limited bandwidth-distance product than single-mode fiber [5].

Wavelength division multiplexing (WDM) and space-multiplexing Combining parallel strands of single-mode fiber with WDM allows for much higher combined data rates [19].

Space-multiplexing in multi-core fibers As high-quality glass fiber are becoming broadly available, prices will drop and application in AOCs becomes feasible. This can lead to cost savings as the transceivers require less separate interfaces. Furthermore, in comparison with fiber bundles, fiber management and ventilation issues are mitigated in large data centers.

Multilevel modulation Instead of using traditional NRZ modulation, the effective bit rate can be increased by using bandwidth-efficient multilevel modulation. When lower-order modulation formats such as 4-pulse-amplitude modulation (PAM) are used, the increase of the complexity of the electronics is limited [20].

Mirage

The project Mirage (Multi-coRe, multilevel, WDM-enAbled embedded optical enGine for Terabit board-to-board and rack-to-rack parallel optics) aims to raise the bar of optical interconnect technology currently used in data centers. The ambition of the project is to increase the optical interconnect speed, which currently tops at around 140 Gb/s per link, to terabit per second numbers. The project introduces several techniques to increase data center scalability and introduce new degrees of parallelism into the optical interconnects of AOCs: data transmission in multicore single-mode fiber and development of the required chip-to-fiber interfaces; introduction of multilevel modulation schemes (4-PAM); introduction of wavelength multiplexing in AOCs.

In this context, the author was involved in the realization of a multichannel high-dynamic range linear transimpedance amplifier (TIA) for multilevel modulation formats. In particular, the author was responsible for the automatic gain control (AGC) system, which was implemented as an event-driven control loop around the linear high-speed datapath, targeting 20 GBd 4-PAM or 26 Gb/s NRZ data.

1.1.2 The Metro Network

The metro network is the intermediate regional network between the access and core network of the internet architecture, with link lengths up to 500 km.

Cisco estimates that metro-only traffic will surpass long-haul traffic (traffic in the core network) in 2015 and will account for 62 % of the total IP traffic in 2018, 2.6 times higher than long-haul (Fig. 1.4). This growth is attributed mainly to increasing traffic from content delivery networks (CDNs) and data centers, which bypass the core network and directly connect to the metro network. These CDNs will carry 55 % of the total internet traffic by 2018. As long-haul traffic is routed through the metro networks, total metro traffic already exceeds long-haul traffic. Internet video growth, including video-on-demand,

videoconferencing and video file sharing, is the major reason for traffic increase and will account for 79 % of the total traffic in 2018 [14].

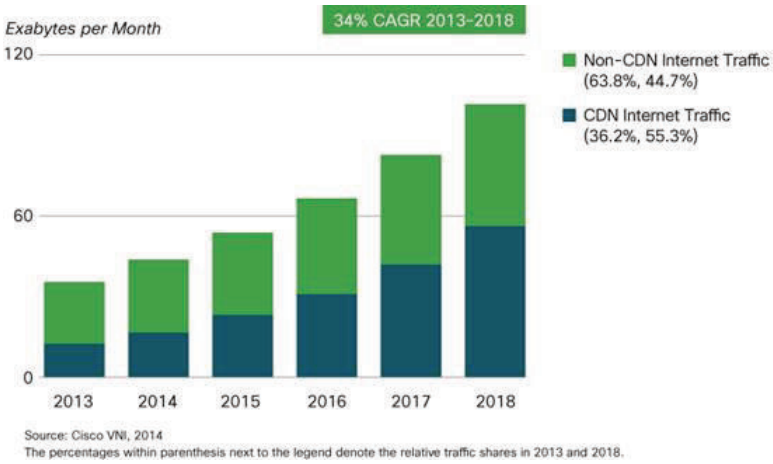


Figure 1.4: Estimated increase of CDN traffic by 2018 [14].

Metro networks consist mainly of electronic BNGs and edge routers, which are the gateway to the access network and core network, respectively. Traditionally, metropolitan area networks were based on switched synchronous optical network/synchronous digital hierarchy (SONET/SDH) architectures. To increase flexibility, scalability and cost-effectiveness, ISPs shifted to Ethernet-based networks. However, this still required optical-electronic-optical conversion at each intermediate node, with associated high cost and power consumption [21]. These redundant conversions were eliminated by assigning a dedicated wavelength channel (colors of light) to each connection between an intermediate node and a hub node. The latter connects the metro network to the core or access network. Furthermore, this technique enables higher data rates on a single fiber due to simultaneously transmitting multiple wavelengths of light. In addition exploiting spatial diversity using parallel fibers can further increase the data rates. Although optical technology is energy efficient, the aggregation of access network traffic can only be performed electronically. Therefore, proper optimization of both the optical and electronic devices is important [8].

Two types of WDM exist. Coarse wavelength division multiplexing (CWDM) is a WDM technology, where the available bandwidth is divided in a coarse grid of up to 20 wavelengths. The relatively high distance between the wavelengths or channels makes optical component performance, such as wavelength stability of the laser, less critical and hence enables lower cost. However, total capacity is limited which implies scalability issues. Still, the small number of channels makes equipment compact.

Today, using dense wavelength division multiplexing (DWDM) across core, metro and access network is widely recognized as being necessary to provide data capacity scalability for future optical networks. The denser wavelength grid allows more frequency diversity. The channel spacing in today's DWDM systems is 100 GHz, 50 GHz or 25 GHz (around 0.8 nm, 0.4 nm or 0.25 nm), which allows up to 160 channels in one fiber [22]. The dense packing of the wavelengths imposes additional constraints on the equipment: high-selectivity optical (de)multiplexers (arrayed waveguide gratings (AWGs)) are required to

aggregate the wavelengths at the transmitting end and to split the combined signal into separate channels at the receiving end. Moreover, high-precision, temperature-stable expensive lasers must keep the channels exactly on target as almost no frequency drift is allowed. Currently, this means keeping the laser at a constant temperature which implies additional cooling costs. At the receiving end, each wavelength of the demultiplexed incoming light is coupled to a photo diode in a photo diode array which is connected to a dedicated lane of a multichannel receiver. The high number of channels requires small physical channel spacing and tight integration of the diode array with the receiver. In addition, as active cooling should be avoided, power consumption per receiver lane must be kept low in order not to exceed thermal operation limits.

C3PO

C3PO (Colorless and Coolerless Components for low Power Optical Networks) focuses on reducing power consumption on multiple levels of the physical network architecture, while enabling bandwidth increase and constraining cost. The project develops colorless (non-wavelength specific) devices for use in DWDM systems. This allows to use the technology in various contexts, from low-cost, reconfigurable router interface for metro networks to the optical modem of the end-user in access networks. Coolerless components operate without active cooling, hence reducing operating costs and allowing tighter integration of photonic and electronic components. This reduction of thermal management requires lower power consumption of the active components, such as laser, modulator, driver and receiver. The project targets four lanes/wavelengths of 25 Gb/s each, resulting in 100 Gb/s total line rate. Since the envisioned system has to support distances up to 500 km, the components have to support optical duobinary (ODB) modulation, next to conventional NRZ, for its greater dispersion tolerance. The ODB signals are directly decoded by the same direct detection receiver.

The author was responsible for the development of the integrated four-channel receiver, targeting 4×25 Gb/s data rate, with low power consumption and small footprint to support tight integration with a p-i-n photo diode array with a 250 µm channel pitch.

1.2 Aspects of Analog Circuit Design

This section shortly presents the design flow, active device models for hand design, analytical tools and a software design tool used during the course of this work.

1.2.1 Design Flow

Although it is impossible to devise a general recipe for the design of analog circuits, it is useful to describe, from a high-level point of view, the design flow that is used for most blocks in this works. Even though presented in an idealized linear fashion, inevitably some (hopefully limited) iterations or even a full restart are required. Also, it may be possible to skip some steps. Sometimes, a specific circuit requires a completely different approach.

The starting point is the main block specifications, which may be derived using any combination of system level modeling, rules of thumb, good design practices or experience. The flow is based on the 3-sigma flow described in [23].

Topology selection A candidate topology is selected that could satisfy the block specifications.

Design equations A circuit model is set up and analytical expressions that express the main block specifications, as a function of device parameters and the topology, are derived. Care must be taken that the model is not overly complicated or simplified. The expressions should be useful for design (see also Appendix A) [24]. A simulator can be used to develop and/or validate the model.

Initial sizing Based on the design equations, knowledge of device operation and its trade-offs, rules of thumb, good practices and experience, an initial set of device parameters is chosen. Note that the set of design equations is invariably underdetermined which explains the aspect of 'choice'. This initial sizing is verified with a simulator.

(Partial) 3-sigma corner extraction To cope with process, supply and temperature (PVT) variations, operating corners and local mismatch, partial 3-sigma corners are extracted for each block specification, based on a Monte Carlo simulation run with limited number of samples (e.g. 100). The block specifications need to be formalized to allow numerical validation. Their probability density function is estimated. Software tool support is a must. In this work Cadence's Virtuoso ADE XL is fit for the purpose.

Resizing If necessary, the initial sizing is adapted such that all the specifications are satisfied across the extracted partial 3-sigma corners.

Monte Carlo run with auto stop A Monte Carlo simulation run is executed to verify the circuit under PVT, operating corners and mismatch variations. The simulator automatically stops the run when the yield estimates for the specifications are met with a certain confidence for a given target yield. This ensures an optimum number of samples. It can be shown that, on average, at least 1400 samples are needed to predict 3-sigma yield with 95 % statistical confidence.

Layout and post-layout verification The circuit is converted to a physical layout. Parasitics and layout dependent effects (e.g. well proximity effect, electromigration effects) are extracted and back-annotated. Another Monte Carlo run with auto stop verifies the post-layout circuit behavior.

For parasitic- and layout-sensitive circuits, specialized tools such as Cadence's Electrically-Aware Design suite can reduce the iteration time by allowing to include parasitics and layout dependent effects early on in the flow. They can be estimated or even extracted from a partial layout. In addition, tools such as worst-case corner extraction, high-sigma yield, global and local optimization, sensitivity and reliability analyses... help the designer in the sizing and verification process and shorten design time.

1.2.2 Transistor Hand Calculation Models

This section concisely introduces the transistor models used for hand design. The active devices offered in the process technology used for the designs in this work (Section 1.3) are enhancement mode MOSFETs and bipolar junction transistors.

1.2.2.1 MOSFETs

A MOS transistor can be operated anywhere in the continuum between weak inversion (WI), strong inversion (SI) and SI with velocity saturation. The inversion level can, a.o., be expressed as inversion coefficient IC, transconductance efficiency g_m/I_D or effective gate-source voltage V_{EFF}. For hand calculations and as a general MOSFET design approach, this work follows [25,26], which is based on the EKV model [27]. The inversion coefficient is a primary design parameter, along with channel length and drain current. Drain current is given by:

$$I_D = I_o \frac{W}{L} IC \tag{1.1}$$

in which $I_o = 2n_0\mu_0 C'_{ox} U_t^2$ is the technology current, with n_0, μ_0, C'_{ox} and U_t the substrate factor, mobility, gate-oxide capacitance and thermal voltage, respectively. The selection of inversion coefficient IC, drain current I_D and channel length L is governed by the circuit design equations (e.g. which impose a certain g_m) and permits different tradeoffs in bandwidth, gain, matching, noise, etc. Channel width W follows from Eq. (1.1). Velocity saturation can be included by replacing IC by a modified value. For example, in the g_m/I_D curve (see Fig. 1.6(a)), replace IC by $IC(1 + IC/IC_{CRIT})$, in which IC_{CRIT} denotes the transition point between the SI g_m/I_D without velocity saturation and the velocity saturated value. While I_o is relatively independent of the channel length (and certainly fulfills the needs for a hand calculation model), IC_{CRIT} is dependent on the electrical field strength and should be extracted for different channel lengths. The reader is referred to [26] for further parameters, expressions and relations.

This design methodology based on inversion level allows to fully exploit the potential of MOS transistors and can also be used with modern process technologies (45 nm and below). The partnership between the BSIM6 and EKV modeling group, an effort to replace the older BSIM3, BSIM4 and PSP compact models for bulk CMOS processes [28], and emerging publications on methodologies in processes with very small channel lengths [29, 30], indicate the increased importance of both accurate modeling and hand design in all regions of operations.

1.2.2.2 Bipolar Junction Transistors

The collector current of a bipolar transistor is exponentially dependent on the base-emitter voltage, over decades of collector current. A simple version of the Gummel-Poon model, which has the Ebers-Moll equations at its core, augmented with junction capacitances, is sufficient as a hand calculation model [31]. Effects such as emitter crowding, Kirk-effect and self-heating are assessed graphically (Section 1.2.3) and with the simulator.

1.2.3 Tools

During the course of this work, various analytical techniques, collectively referred to as methods of design-oriented analysis [24, 32, 33] were acquired and applied consciously. These methods make it easier to analyze a circuit, and aim for resulting expressions that are useful to achieve the design objectives. These techniques range from simple, such as the notion of inverted poles and zeros and doing algebra on the graph, to more advanced. An example of the latter is the general network theorem (GNT) and its descendants, the general feedback theorem (GFT), extra element theorem (EET) and chain theorem (CT),

that allow to decompose a transfer function such that its constituent parts have a specific meaning to the circuit. In order to be able to apply these theorems numerically to real designs, they were integrated as a new analysis in Cadence Virtuoso during this work. Appendix A presents a summary.

A software tool was developed to aid the designer in visualizing and exploring the various trade-offs inherent to a single device, i.e. a bipolar transistor, MOSFET or even a passive. This tool allows to plot single parameters or expressions governing device behavior as a function of operating point, temperature and physical parameters. It helps the designer in sizing a device and choosing quiescent currents, based on the graphical representation. This graphical design method was used extensively in the vacuum tube and early transistor era [34, 35] and is complementary to equation-based design. A lot of principal design equations that describe the behavior of a circuit are a function of operating point parameters of active devices: inversion coefficient, transconductance, output conductance…A major design task involves translating these parameters to bias current, emitter area, gate area (see Section 1.2.1). The effect of a given choice on the design equations as well as on other, perhaps overlooked parameters, is quickly made clear by the various plots. This leads to less iterations and more efficient design.

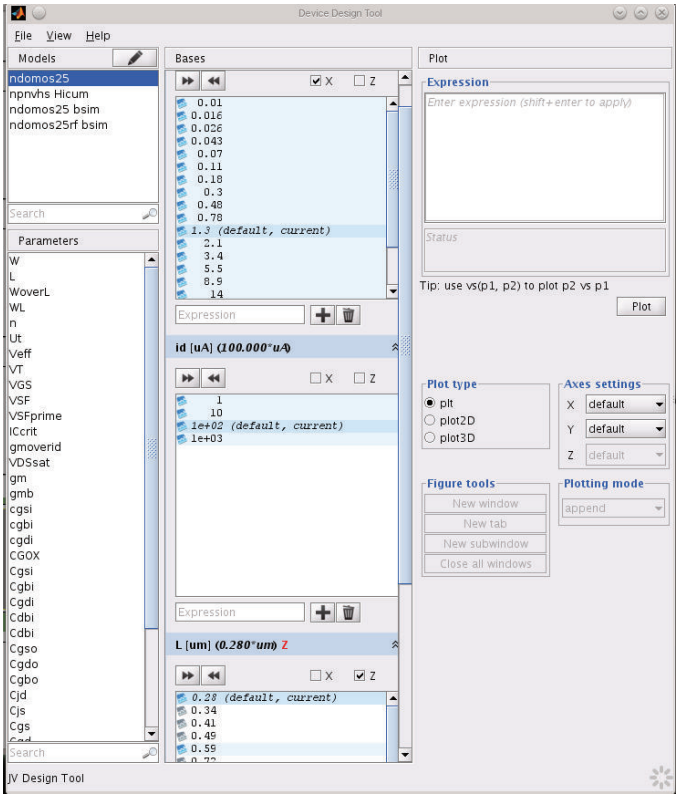
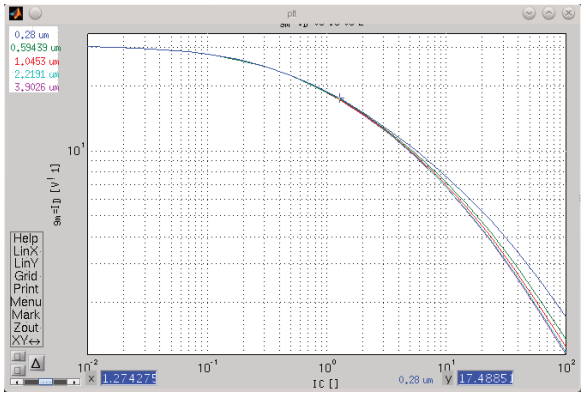


Figure 1.5: Main interface of the design tool.

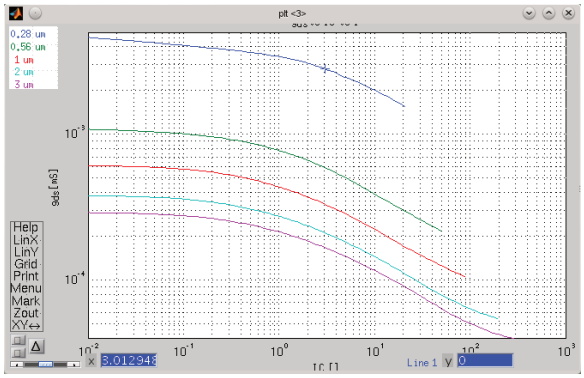
It is the author's experience that the different trade-offs and optimal operating region for a certain parameter are identified with increasingly less effort as the shape of a graph (including some key numbers) is 'automatically' memorized over time. Then, dependence

on the software gradually reduces. This may be of higher interest than trying to remember a bunch of potentially complex equations. Case in point is the MOS transistor, whose properties change widely over inversion level, drain current and gate length (already ignoring all the higher-order effects).

Figure 1.5 depicts the main interface of the software tool. Example plots for the technology used in this work (Section 1.3) are shown in Figs. 1.6 and 1.7: the transconductance efficiency g_m/I_D curve of a typical NMOS transistor (Fig. 1.6(a)), for various channel lengths; the output conductance g_{ds} of a NMOS transistor, for various channel lengths, at a fixed drain current (Fig. 1.6(b)); the transition frequency f_T of a bipolar transistor, for various emitter widths (Fig. 1.7(a)); the forward current gain β of an NPN bipolar transistor, for various emitter widths (Fig. 1.7(b)).

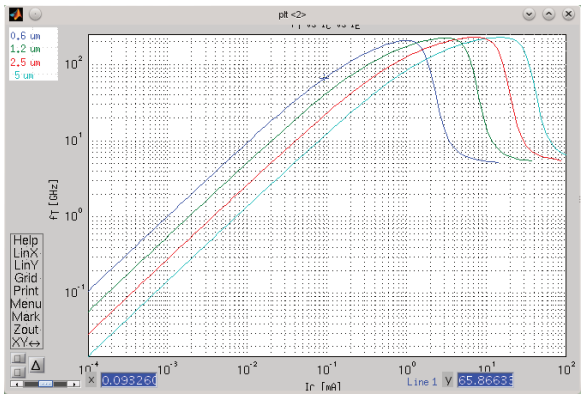


(a) Transconductance efficiency g_m/I_D of a MOS transistor, for various channel lengths, versus inversion coefficient IC.

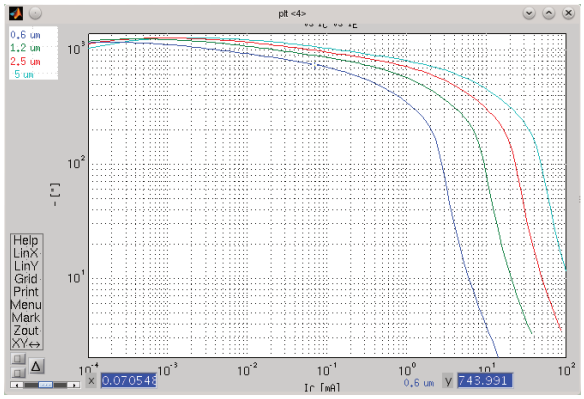


(b) Output conductance g_{ds} of a MOS transistor, for various channel lengths, at fixed drain current, versus inversion coefficient IC.

Figure 1.6: Illustration of plots generated by the design tool for a double-oxide NMOS transistor.



(a) Transition frequency f_T of a bipolar transistor, for various emitter widths, versus collector current I_C.



(b) Forward current gain β of a bipolar transistor, for various emitter widths, versus collector current I_C.

Figure 1.7: Illustration of plots generated by the design tool for a high-speed bipolar NPN transistor.

1.3 Technology

Both the optical multichannel NRZ receiver and the optical linear receiver with AGC system were implemented in a 0.13 µm SiGe BiCMOS process [36]. A BiCMOS technology allows integration of high performance analog blocks with low-density logic. It combines the high-speed, high current capability and low noise of bipolar devices with the low-power MOS transistor (for logic) [37]. For the implementation of the high-speed datapaths, the bipolar transistor is a better choice as it is faster, has higher current capability, has better transconductance than a MOS transistor at equal current, offers better input-referred voltage matching and is optimized for low noise. The MOS transistors are used in CMOS logic circuits and whenever e.g. their infinite input resistance is advantageous, a complementary topology is useful, smaller input voltages are needed (in WI)…

For a given BiCMOS node, a comparable CMOS technology would be scaled by two generations. Hence, 65 nm CMOS offers similar speed as 0.13 µm BiCMOS (bipolar tran-

sistors). However, analog circuits show less benefit from CMOS scaling than their digital counterparts as the passive components do not scale as fast (if at all) and the supply voltage decreases, which limits dynamic range and headroom in general. In addition, the cost per unit of area for a design in a BiCMOS process in a multiproject wafer run is lower than in a comparable CMOS process. Also, compared to bulk CMOS, the BiCMOS process has a high-ohmic P- substrate, which is beneficial to reduce crosstalk (Section 2.5.4).

In the front-end, the technology offers high-speed vertical NPN bipolar transistors with transition frequency f_T up to 220 GHz, as shown in Fig. 1.7(a). The (open base) collector-emitter breakdown voltage is 1.2 V. Figure 1.7(b) indicates that the forward current gain β easily exceeds 500, at typical bias currents. Therefore, in all small-signal equivalent circuit, β will be assumed infinite if not otherwise noted. Also available are lower speed (f_T = 120 GHz) NPN transistors with higher breakdown voltage (1.6 V). Low-speed, low-β lateral PNP devices are available for band gap references.

Single-oxide (1.2 V, 0.13 µm) and dual-oxide (2.5 V, 0.28 µm) NMOS and PMOS transistors are offered. Each type is available in a natural-, low- or high-V_T version. The low-V_T devices can be used as self-cascode devices. The relevant technology currents I_o (Section 1.2.2) are listed in Table 1.1. As expected, due to lower carrier mobility, PMOS devices have a 3 to 5 times lower I_o than NMOS devices.

	Device	L_{min} [µm]	I_o [µA]
double-oxide			
	NMOS	0.28	0.34
	PMOS	0.28	0.07
single-oxide			
	NMOS	0.13	0.55
	PMOS	0.13	0.13

Table 1.1: Technology currents of the MOS transistors in the given 0.13 µm BiCMOS technology (at 300 K).

As for passive devices, the technology offers capacitors, resistors and inductors. The capacitors are either of the poly-well (MOS-capacitors) or the metal-insulator-metal (MIM) type. The latter offers much greater linearity and stability over PVT variations, but has less capacitance per unit area. The resistors are of the diffusion-, metal- or poly-type, the latter with various doping levels. A high-sheet resistance poly resistor is also available.

In the back-end, 6 metal layers plus one aluminium top layer are available for routing. The upper two metal layers are much ticker than the lower metals.

1.4 Outline of the Dissertation

This chapter has presented the background against which the research in this work is performed. Furthermore, some aspects of integrated analog circuit design have been highlighted: the design flow, transistor hand models and a software tool. In addition, the technology in which the integrated circuits described in this work have been implemented, has been presented.

The remainder of this dissertation is composed of two major parts: the design and implementation of an event-driven AGC system for a high-speed linear optical receiver and the development of a limiting 4×25 Gb/s receiver array. It should be mentioned that, although the AGC system is presented first, the real chronology was in reverse order. Hence, the wealth of experience gathered from developing the 4×25 Gb/s receiver was applied to the design of the datapath in the linear receiver.

In Chapter 2, basic optical receiver concepts, necessary to understand the material in the subsequent chapters are reviewed, in particular the requirements and their implications on the design of the receiver array and the datapath of the linear receiver.

Chapters 3 to 5 constitute the bulk of this work: the event-driven AGC system embedded in a high-speed linear optical receiver. In the first part of Chapter 3, the fundamentals of feedback AGC systems are introduced. In the second part, an event-driven approach is proposed and developed. Basic limitations and design relations are derived and further built upon in Chapter 4, in which the AGC system is designed on the system level. Chapter 5 presents the detailed design of the building blocks on the circuit level, along with experimental results, which underline the successful implementation of the proposed system.

Chapter 6 presents the design of the 4×25 Gb/s receiver array, focusing on the TIA input stage and discussing techniques to improve bandwidth and dynamic range. Measurement results are presented.

Finally, Chapter 7 gives general conclusive remarks and suggests some topics for further research.

Two appendices are included. Appendix A gives a short introduction to the general network theorem (GNT) and its derived theorems. These theorems provide fast analytical techniques to analyze linear systems and are used throughout this book. Appendix B presents an analysis of a two-stage opamp compensated with capacitance multipliers. This topology was used for a building block in the AGC system.

Chapter 2

Optical Receiver Concepts

This chapter introduces essential optical receiver concepts which are relevant for the remainder of this work.

The basic structure of an optical receiver is presented in Section 2.1. Transmission modes and data formats are described in Sections 2.2 and 2.3, respectively. The dynamic range, determined by the sensitivity and overload limit, is discussed in Section 2.4. The concept of power penalty and the impact of several non-idealities, in particular the problem of crosstalk in a multichannel receivers, is presented in Section 2.5. Finally, an overview of the p-i-n photo diode is given in Section 2.6. Note that this work is only concerned with direct detection of the optical input signal. As such, coherent receivers are beyond the scope of this text.

2.1 Overview

An optical receiver for baseband modulation is generally composed of four building blocks (Fig. 2.1): a detector, transimpedance amplifier (TIA), main amplifier (MA) and clock-and-data recovery (CDR) block [18]. Information is transmitted on an optical carrier in a certain modulation format. The function of the receiver is to recover the information embedded in the received signal.

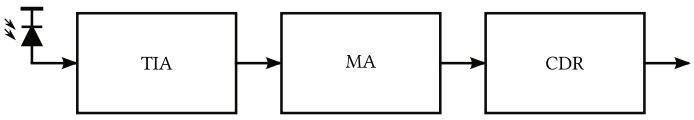


Figure 2.1: Conceptual block diagram of a optical receiver.

The detector linearly converts the incident optical power to a current, which is amplified and converted to a voltage by the TIA. The MA provides further amplification and can be limiting or linear, depending on the modulation format. The MA and CDR jointly recover amplitude and timing information and act as a decision block. These functional building blocks can be combined in the physical implementation.

In this work, the focus lies on the TIA and MA, which will be integrated on one chip. Although no CDR is included, the system is still denoted as a 'receiver'. Chapter 6 de-

scribes a multichannel limiting receiver, while Chapter 5 presents the design of an automatic gain control (AGC) system embedded in a linear receiver.

2.2 Transmission Mode

In continuous-wave (CW) transmission, data packets are transmitted in a continuous, uninterrupted stream of bits or symbols. The incident optical power is relatively constant once reception has started. In burst-mode transmission, data arrives in bursts with strongly varying power levels. This requires a different approach for the design of a burst-mode receiver [38, 39]. This work only considers CW transmission, in which data arrives in a continuous fashion and not in individual bursts.

2.3 Data Format

The data format of an optical information stream is given by the modulation format and type of line coding. The data format typically used with limiting receivers (Chapter 6) is non-return-to-zero (NRZ) or optical duobinary (ODB), while the linear receiver (Chapter 5) is made for multilevel modulation such as pulse-amplitude modulation (PAM).

2.3.1 Modulation

Information can be conveyed in a single-mode optical fiber by modulation of three physical attributes of the optical field: intensity, phase and polarization. Among the multitude of existing optical modulation formats, the simplest one is NRZ or on-off keying (OOK), in which information appears as two intensity levels: the optical signal is on for a one bit and off for a zero bit. This format puts the least constraints on the receiver linearity. Given a bit rate of R_b bits/s, it can be shown that most of the power is confined in a bandwidth of R_b Hz [40].

Spectral efficiency can be increased by adding more intensity levels, as in PAM. This also alleviates inter-symbol interference (ISI) induced by chromatic dispersion as, for a given bit-rate, the required bandwidth is lower compared to NRZ. However, it makes the design of the receiver more difficult as a linear datapath is required with controlled overshoot. In addition, a higher signal-to-noise ratio is needed for reliable detection.

A compromise between the chromatic dispersion tolerance of multilevel schemes and simplicity and low-cost of direct detection receivers is found in ODB modulation. The electrical three-level output of a precoder is mapped into three optical states: low and high optical intensity, of which the phase of the optical field can be inverting or non-inverting [41–43]. The precoded ODB signal can be demodulated into a NRZ-like signal by a conventional direct detection receiver.

Many more complex modulation schemes also embed information in the phase and polarization of the optical field, a discussion which is beyond the scope of this work [44].

2.3.2 Line Coding

Line coding is usually applied to a data stream to provide the following properties: dc balance, short run lengths and high transition density. The time-average of a dc balanced signal is centered between the extreme values and carries no information. This allows

for ac coupling or functionally equivalent operations, such as offset compensation. However, in order to limit the power penalty (Section 2.5.3) due to baseline wander or droop, the low-frequency content of the data should be limited. Hence, a maximum number of consecutive identical bits or symbols (run-length) is usually imposed. A high transition density is desired for the clock recovery. Well-known types of line coding include pseudo-random bit sequence (PRBS) scrambling and block coding (e.g. 8B/10B, 64B/66B) [45].

2.4 Dynamic Range

The dynamic range of an optical receiver is defined as the input (optical power or electrical current) range where a specified signal quality is achieved. Dynamic range is lower bounded by the sensitivity limit and upper bounded by the overload limit. Signal quality is usually expressed as a bit-error ratio (BER), although sometimes signal-to-noise ratio (SNR) or SNR per bit (E_b/N_0) is used. The BER associated with 'error-free' reception depends on the context. In this work, the target is 10^{-12}. Visually, received signal quality is sometimes evaluated on the eye diagram. Note that error-correcting codes can be used to obtain a lower BER at the expense of some overhead.

2.4.1 Sensitivity

The sensitivity limit equals the smallest input signal that can still be received with a certain quality. For NRZ modulation, free of ISI and only Gaussian signal-independent noise, the *average* optical sensitivity can be expressed as:

$$\overline{P}_{\text{sens}} = Q \frac{i_n^{\text{rms}}}{\mathcal{R}} \tag{2.1}$$

with i_n^{rms} the total rms input-referred noise current of the receiver and $\mathcal{R}$ the photo diode responsivity (Section 2.6). The Personick-Q Q expresses how much larger the optical signal needs to be, compared to the noise, to obtain a given BER. Formally, the threshold of the decision circuit should be at least Q standard deviations (of the noise) above or below the mean high or low level to ensure a desired error rate [46]. For a BER of 10^{-12}, $Q = 7.035$.

Equation (2.1) can also be expressed in the electrical domain as follows, assuming infinite extinction ratio (ER):

$$i_{\text{sens}}^{\text{pp}} = 2\,Q\,i_n^{\text{rms}} \tag{2.2}$$

in which $i_{\text{sens}}^{\text{pp}}$ is the minimum *peak-to-peak* input current.

In this work, this simple model will be used as p-i-n photo diode noise is largely signal-independent.

2.4.2 Overload Limit

As the input signal increases beyond the sensitivity limit, the BER continues to decrease, typically until a BER-floor is observed. At a given input level, signal quality degrades again as large-signal effects such as pulse-width modulation, jitter and asymmetric clipping distort the output eye. The overload limit, $\overline{P}_{\text{ovl}}$ or $i_{\text{ovl}}^{\text{pp}}$, equals the maximum input level

where a specified signal quality is still obtained. No general formula exists as the overload limit is highly topology-dependent.

The design described in Chapter 5 includes AGC to increase the overload limit and hence increase the dynamic range.

2.5 Power Penalties and Non-idealities

Several non-idealities in the receiver degrade the performance or quality of the received signal. A concept that allows to quantify these impairments is power penalty (PP). Formally, a PP related to an impairment is the increase in average optical transmit power necessary to achieve the same BER as in the absence of the impairment. The relevant PP's in the context of this work are described concisely in this section. Several other power penalties are covered in literature [18].

2.5.1 Extinction Ratio

When an optical transmitter transmits a zero, the light is not completely shut off. The ratio of the (time-average) power of a logical '1' to the (time-average) power of logical '0' is called ER:

$$ER = \frac{P_1}{P_0} \tag{2.3}$$

Ideally ER is infinite. It can be shown that, for unamplified p-i-n receivers, the PP due to finite ER is [18]:

$$PP = \frac{ER + 1}{ER - 1} \tag{2.4}$$

For $ER = 10\,\text{dB}$, $PP = 0.87\,\text{dB}$.

2.5.2 Bandwidth

Limited bandwidth of a receiver introduces ISI and limits the output swing of certain bit sequences (those with higher frequency content). As a result, the vertical eye opening is reduced from V_h to V_h'. In order to restore the vertical eye opening and original BER, the output swing needs to increase to $PP \cdot V_h$, such that:

$$PP = \frac{V_h}{V_h'} \tag{2.5}$$

2.5.3 High-pass Behavior

Certain circuit topologies, such as balancing, feedback offset compensation or ac coupling, introduce a high-pass pole in the small-signal gain of the datapath. This introduces drift, droop or baseline wander in the output signal when the input signal remains at the same level for a long period of time, e.g. when receiving a long string of consecutive identical digits (CID). This causes a degradation of signal quality (BER).

It can be shown that the PP, for NRZ modulation, associated with a single high-pass pole f_{pL} is [18]:

$$PP = 1 + \frac{2\pi f_{pL} r}{R_b} \tag{2.6}$$

in which R_b is the bit rate and r the maximum number of CIDs. Low-frequency content is usually limited by line coding (Section 2.3.2).

2.5.4 Crosstalk in Multichannel Receivers

Multichannel receivers integrate more than one receiver lane on a single chip, sharing control wires and possibly supply rails. This allows tighter integration with higher total data rates, e.g. one receiver channel for each wavelength in a dense wavelength division multiplexing (DWDM) system. However, multichannel receivers can incur an additional penalty due to crosstalk between neighboring channels.

Crosstalk can be defined as the undesired energy appearing in one signal path as a result of coupling from other (adjacent) signal paths [47]. In a system such as an optical receiver, the total crosstalk is hard to quantify analytically or even numerically. Hence, a general formula for the PP is not available. However, the PP can be measured (Section 6.7.3).

Crosstalk can be limited by technological and topological measures as discussed next.

Mechanisms and Mitigation of Crosstalk

Qualitatively, one can discern several crosstalk mechanisms, which can be coarsely classified as electromagnetic coupling via the bondwires and coupling via the substrate or supply rails.

Part of the input current of a given channel will originate from the other channels due to coupling between bondwires. At lower frequencies, bondwires are electrically short compared to the wavelength and the effect can be treated by inductive and capacitive coupling. At higher frequencies the electromagnetic field coupling dominates as each bondwire becomes a transmitting and receiving antenna and wave propagation effects must be taken into account. As both bondwires of a channel are terminated (at least at one side) by a relatively low impedance (a decoupled supply at the cathode bond pad and the low input impedance of the TIA at the anode bond pad), inductive coupling with other channels, expressed through the mutual inductance L_m, will be dominant (not unlike current transformer action). One way to reduce L_m is to increase the channel separation or reduce the spacing between both bondwires of a single channel (as current flows in opposite direction). However, this might not be possible when channel and bondwire spacing is fixed and dictated by the photo diode array. For the same reason, shielding might not be usable.

Other sources of crosstalk originate in the receiver chip itself [48, 49]. A first mechanism is coupling through the substrate. Assuming the substrate is tied to ground, substrate currents flow whenever the intended ground return path is not low impedance enough, such that a current divider is formed between the substrate and intended ground wire path. A voltage drop will develop across the non-zero impedance of the substrate. This can provoke local ground-bounce, emitter or source debiasing, etc. [49]. Quantification of this effect requires a decent substrate model and optimized tools to handle to vast amount

of resulting nodes. Several techniques are available to isolate building blocks and minimize inter-channel crosstalk:

Lightly doped substrate Use of a P- substrate makes the substrate higher impedant. In a SiGe process, this is usually the case. However, to avoid latch-up, the substrate must be firmly tied to ground. Silicon-on-insulator (SOI) technologies take this approach to the extreme.

Guard rings Single or multiple guards rings around devices collect return current from the substrate to the metal wires.

Triple wells Isolating wells combined with guard rings enhance the effect of guard rings alone.

Deep trench isolation In a SiGe BiCMOS process, the sidewalls of the collector tubs of the bipolar transistors are created using deep trenches. These extend deeper into the substrate than wells. When used as barriers between blocks, these trenches can provide superior isolation.

A second important mechanism is coupling through the supply rails. Currents in the supply lines develop a voltage drop across the rail impedance, which can influence connected circuits. Remedies include:

Supply decoupling and filtering On-chip supply decoupling provides a local return path for higher-frequency signals and minimizes current loop area. However, it might be still required to include a small resistance to damp potential ringing. A common technique to power a multistage amplifier is to apply the supply at the output stage and provide progressive supply filtering between the stages. This avoids that large currents in the supply rails flow past the sensitive input stages.

Separate power domains Separate blocks are supplied with dedicated supply rails. Crossing the domain boundaries should be avoided. However, this leaves less area for on-chip decoupling and requires more pins.

Differential circuits In a differential circuit, supply noise manifests itself approximately as common-mode interference, which is rejected to a large extent. Differential topologies, however, consume more power and require more silicon area than their single-ended counterparts and have implications on noise (generated by the circuit itself) performance.

Many of these techniques have been used in the design of the multichannel receiver in Chapter 6.

2.6 Detector: P-I-N Photo Diode

The photodetector converts incident light into a photocurrent. Three major types are commonly used: the p-i-n diode, the avalanche photodetector (APD) and the optically preamplified detector. In this work p-i-n diodes are used, which consist of three layers of semiconductor material: an intrinsic layer is sandwiched between a p-n junction. Reverse bias must be applied in normal operation. Their relevant characteristics are concisely presented next, without going further into the physical details [18, 50].

Multiple photo diodes can be integrated in an array, either with or without common nodes. A typical standard pitch is 250 µm [51].

2.6.1 Responsivity

The responsivity $\mathcal{R}$ [A/W] of a p-i-n diode expresses the amount the photocurrent generated per unit of incident optical power:

$$I_{\mathrm{PD}} = \mathcal{R}P \tag{2.7}$$

This is a linear relation, which is somewhat remarkable, as usually power increases with the square of the current. The responsivity depends on the wavelength of the incident light and the quantum efficiency of the photo diode and typically ranges between 0.2 A/W and 0.9 A/W.

2.6.2 Linear Model

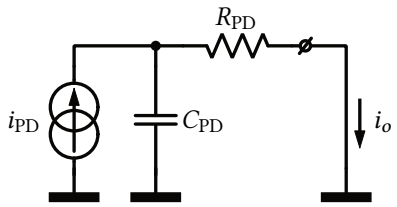


Figure 2.2: Linear model of the p-i-n diode.

A simple linear model of an unpackaged photodiode (Fig. 2.2) is a first order circuit. The output current, sensed with an ideal current sink, is given by:

$$i_o = \frac{i_{\mathrm{PD}}}{1 + \frac{s}{\omega_{\mathrm{PD}}}} \tag{2.8}$$

with the bandwidth determined by the ohmic contact resistance R_{PD}, junction depletion capacitance C_{PD} and the transit time τ_{PD} through the intrinsic layer:

$$\omega_{\mathrm{PD}} = \frac{1}{R_{\mathrm{PD}}C_{\mathrm{PD}} + \tau_{\mathrm{PD}}} \tag{2.9}$$

The transit time is optimally small at its saturation value when the diode is reverse biased at an appropriate device-dependent voltage.

Chapter 3

Event-driven AGC Concepts

In the context of the Mirage project (Section 1.1.1), a linear optical receiver for multi-level modulation formats is required. In order to obtain sufficient dynamic range while remaining adequately linear, some form of automatic gain control (AGC) is required. Conceptually, such a system adapts the gain of the high-speed datapath such that the output amplitude remains constant.

Depending on whether the system acts on the signal detected at the input or the output of the datapath, the architecture is called feedforward or feedback AGC, respectively [52]. In feedback systems the dynamic range at the input of the detector is limited, while in feedforward systems the detector is exposed to the entire input signal range and requires expanded linear range. The feedback approach is inherently more precise, but stability is a concern and should be carefully asserted. In the context of a high-speed optical receiver, it is desirable to avoid extra circuitry at the sensitive input stage to avoid extra noise. For these reasons, this work focuses on a feedback architecture. It should be noted however, that with some effort the event-driven concept could also be applied to feedforward architectures, in case of a more suitable application context.

In Section 3.1, the fundamentals of feedback AGC systems are discussed. Based on a generic large-signal model, a small-signal equivalent linearized model is derived. It is shown that in order to obtain a settling time independent of the input amplitude step, the variable gain amplifier (VGA) should be exponentially dependent on its control signal and a logarithmic amplifier should be included in the loop. A bounded settling time is desirable in optical receivers as often transient response must have vanished during a fixed-length data preamble. In Section 3.2, the loop is quantized and a digital look-up table (LUT) is included. This allows independent control of the datapath's gain, bandwidth and peaking behavior, not easily obtainable in an analog fashion. Indeed, controlled overshoot across the signal input range is important to retain signal quality when multilevel modulation formats are used. In addition, the system introduces flexibility, as the datapath can be reprogrammed for different scenarios, even on the fly. However, quantization introduces limit-cycling. This phenomenon is analyzed using a large-signal linearization technique known as describing functions. A system is introduced to avoid these undesired oscillations. It is shown that the settling time is upper-bounded by its continuous-time value. Furthermore, the salient parameters of the loop gain, as a function of building block parameters, are discussed, as well as a number of secondary aspects.

3.1 Fundamentals of Feedback AGC Systems

Feedback AGC systems have been studied for decades in classic literature and their behavior is well understood [52–56]. The following section provides a basic introduction. Starting from a general system description with only a few basic assumptions, a linear equivalent system is derived. The results are consistent with previous work, as is to be expected.

3.1.1 Continuous-time Feedback AGC

Consider the general continuous-time feedback AGC system depicted in Fig. 3.1. An input signal $A_i(t)p(t)$ with instantaneous amplitude $A_i(t)$ is applied to a VGA G. The frequency response of the VGA is much wider than the loop bandwidth, such that no dynamics are introduced and its model reduces to a simple gain $G(C)$. The amplitude $A_o(t)$ of the output signal is extracted by a detector D and compared to a desired reference amplitude A_{ref}. Loop filter L (approximately an integrator) amplifies the error signal and applies the signal C to the control input of the VGA. Given a stable loop, this system steers the VGA such that in equilibrium, the detected amplitude approximates A_{ref}. An AGC system is denoted *simple* when A_{ref} is zero, *delayed* otherwise [55].

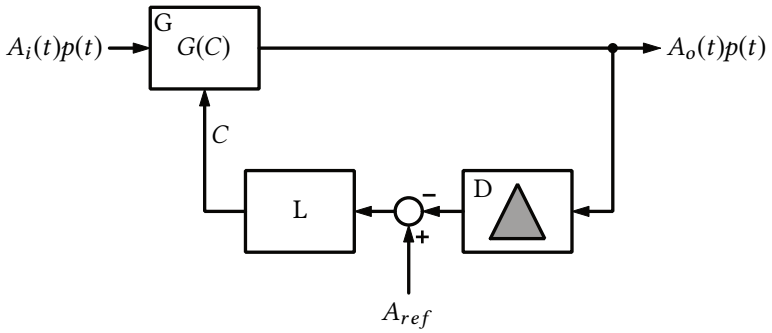


Figure 3.1: General continuous-time feedback AGC system.

3.1.2 Large-signal Transfer Characteristic

Figure 3.2 shows the static input-output characteristic of a general delayed feedback AGC system. Gain control action occurs for input amplitudes between A_1 and A_2, called the effective AGC range. Usually non-linearities such as amplifier clipping determine the limits, although they can be deliberately built in. Outside of the AGC range, the loop is effectively open and the VGA gain stays fixed. The following sections assume the input level to be within the effective AGC range.

A figure of merit, the *compression ratio*, stiffness or flatness factor M_c, is defined as the ratio of the change of input level in dB to the change of output level in dB [55]. It is related to the loop gain as will be shown below. Note that in the case of simple AGC, M_c cannot be arbitrarily high, since the system would produce no output at all.

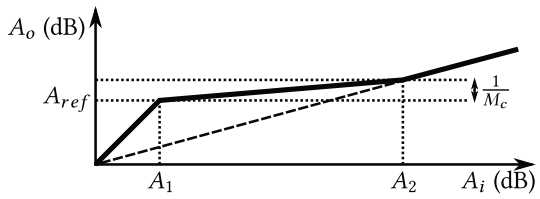


Figure 3.2: Large-signal static characteristic of a delayed AGC system.

3.1.3 Small-signal Linearization

Clearly, this system is non-linear, as the gain of the VGA depends on the amplitude of the input signal and the detector performs a non-linear operation. In general non-linear dynamics are difficult to analyze exactly [57, 58]. One way of tackling this problem is to linearize the system for incremental changes around a given stable steady state, assuming one exists, such that linear analysis tools can be applied. Large-signal behavior should still be confirmed, if not analytically, then at least numerically.

As a first step, note that the system essentially operates on the (low frequency) envelope of the input signal, hence the dependence on the signal shape through $p(t)$ can be omitted. In addition, the explicit time argument will be dropped for notational clarity. Now, if the detector is made much faster than the loop bandwidth, it can be replaced by its incremental gain, which is assumed signal-independent and constant. As we pursue a general analysis approach, let's include a continuous, invertible function $F(\cdot)$ that lumps together the equivalent detector gain and other, possibly non-linear, operations in the feedback path (see below). Oftentimes, $F(\cdot)$ is also applied to the reference input in order to obtain a direct mapping between A_{ref} and A_o. This is depicted in Fig. 3.3.

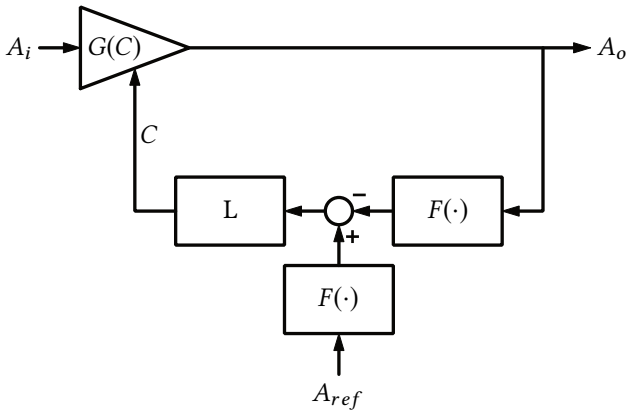


Figure 3.3: General feedback AGC system in the amplitude domain. The $F(\cdot)$ function lumps together the equivalent detector gain and other, possibly non-linear, operations.

The next step involves linearizing this system around a stable steady state, which will be indicated by the subscript q. Denote any signal X in the system by the sum of a steady-state component and a small disturbance:

$$X = X_q + \Delta x \tag{3.1}$$

Explicit increment notation (Δx) is retained for clarity. The goal is to find the equivalent system that operates on the increments. The function describing the VGA has essentially two inputs, the input amplitude and the control signal:

$$A_o = G(C)\, A_i \tag{3.2}$$

Calculation of the total differential of Eq. (3.2), which expresses the change of the output amplitude Δa_o due to a change of the control signal Δc or the input Δa_i, yields:

$$\Delta a_o = A_{i,q} \left.\frac{\mathrm{d}G(c)}{\mathrm{d}c}\right|_{c=C_q} \Delta c + G(c)|_{c=C_q}\, \Delta a_i \tag{3.3}$$

$$= A_{i,q} \frac{\mathrm{d}G(C_q)}{\mathrm{d}c} \Delta c + G(C_q) \Delta a_i \tag{3.4}$$

where Eq. (3.4) uses simplified Leibniz notation for brevity. The dependence on the operating point is readily apparent. Equally, evaluating the derivative of $F(x)$ in the appropriate operating points ($A_{ref,q}$ and $A_{o,q}$) results in an incremental gain.

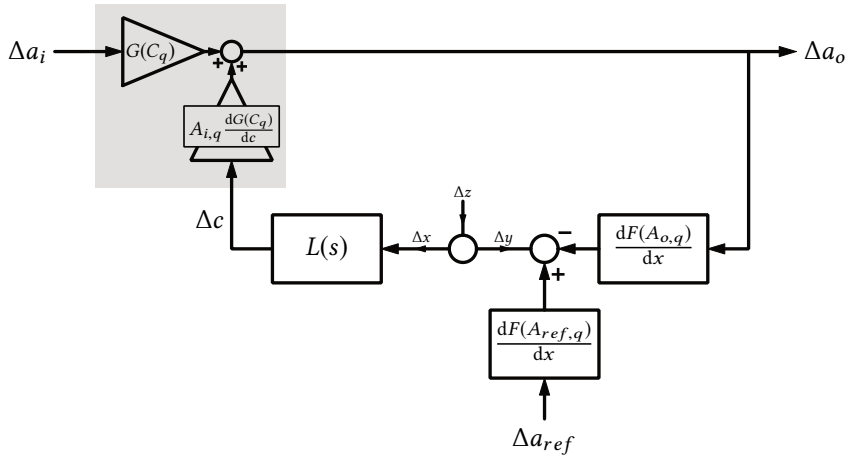


Figure 3.4: Equivalent linearized general continuous-time feedback AGC system and 1-GFT test signal injection.

Figure 3.4 shows the equivalent linear AGC system for small changes. This is a two-input (Δa_i and Δa_{ref}), one-output (Δa_o) linear system. As superposition now holds, let's apply the 1-GFT to both closed-loop transfer functions in this feedback system. The aim is to decompose each closed-loop transfer H function in the following form:

$$H = H_\infty \frac{T}{1+T} + \frac{H_0}{1+T} \tag{3.5}$$

The definitions of the various lower-level transfer functions are summarized in Appendix A.1.1. To not overly complicate notation, the symbols H_∞ and H_0 will not be further specified and the context should clarify to which decomposition they apply.

The natural injection point is at the error summing point, as indicated in Figure 3.4 and results in the desired H_∞ in both cases. Set both inputs to zero. The loop gain $T(s)$ is

given by:

$$T(s) = \left.\frac{\Delta y}{\Delta x}\right|_{\Delta a_i=0,\Delta a_{ref}=0} \tag{3.6}$$

$$= A_{i,q}\frac{\mathrm{d}G(C_q)}{\mathrm{d}c}\frac{\mathrm{d}F(A_{o,q})}{\mathrm{d}x}L(s) \tag{3.7}$$

Consider now the transfer function from Δa_{ref} to Δa_o, so Δa_i must be set to zero. The second-level transfer functions become:

$$H_\infty = \left.\frac{\Delta a_o}{\Delta a_{ref}}\right|_{\Delta a_i=0,\Delta y=0} \tag{3.8}$$

$$= \left[\frac{\mathrm{d}F(A_{ref,q})}{\mathrm{d}x}\right]\left[\frac{\mathrm{d}F(A_{o,q})}{\mathrm{d}x}\right]^{-1} \tag{3.9}$$

$$\approx 1 \tag{3.10}$$

$$H_0 = \left.\frac{\Delta a_o}{\Delta a_{ref}}\right|_{\Delta a_i=0,\Delta x=0} \tag{3.11}$$

$$= 0 \tag{3.12}$$

The approximation in Eq. (3.10) holds as in equilibrium $A_{ref,q} \approx A_{o,q}$ and $F(\cdot)$ is assumed smooth in that region. Equation (3.12) indicates no direct forward transmission. Indeed, $F(x)$ is unilateral in this ideal block diagram.

Now, let Δa_{ref} be zero. The corresponding results for the transfer function from Δa_i, which is the 'intuitive' system input, to Δa_o are:

$$H_\infty = \left.\frac{\Delta a_o}{\Delta a_i}\right|_{\Delta a_{ref}=0,\Delta y=0} = 0 \tag{3.13}$$

$$H_0 = \left.\frac{\Delta a_o}{\Delta a_i}\right|_{\Delta a_{ref}=0,\Delta x=0} = G(C_q) \tag{3.14}$$

Equation (3.13) holds because when both Δa_{ref} and Δy are zero, the output of the detector and hence the system output is necessarily zero. It expresses that, ideally, the system does not change the output amplitude when the input amplitude changes. This is clearly desired in an AGC system. However, direct forward transmission is not zero, as indicated by Eq. (3.14). This is the system response when the loop gain is zero, e.g. the initial response to a step of Δa_i.

It follows from Eqs. (3.10), (3.14) and (A.24) that the complete system of Fig. 3.4 can be described as:

$$\Delta a_o = \frac{T(s)}{1+T(s)}\Delta a_{ref} + \frac{G(C_q)}{1+T(s)}\Delta a_i \tag{3.15}$$

with $T(s)$ given by Eq. (3.7). As usual, the return difference $1+T(s)$ determines the stability of the closed-loop system[1], while the dynamic responses to the system inputs additionally include the other factors.

[1] Assuming implicitly that the system with T killed is stable.

Now, consider only the response due to Δa_i. Equations (3.7) and (3.15) show that, as both the VGA gain and the loop gain depend on the operating point, in general the settling time of the closed loop depends on the input amplitude. This is undesired in most applications, e.g. when a receiver has to adjust its output amplitude during a fixed-length data preamble. However, it is possible to achieve constant settling time by appropriate choice of $G(C)$ and $F(x)$ (and $L(s)$) [59–63]. Let's consider two practically useful implementations of $F(x)$.

Ideal detector The detector extracts a scaled version of the amplitude of the output signal. It follows that:

$$F(x) = k_D x \tag{3.16}$$

$$\frac{\mathrm{d}F(A_{o,q})}{\mathrm{d}x} = k_D \tag{3.17}$$

and the loop gain (Eq. (3.7)) reduces to

$$T(s) = A_{i,q} \frac{\mathrm{d}G(C_q)}{\mathrm{d}c} k_D L(s) \tag{3.18}$$

Clearly, the loop stability and closed-loop dynamics still depend on the input amplitude, as well as the VGA function $G(C)$ and detector gain k_D.

Ideal detector and logarithmic operation Usually, one adds a logarithmic amplifier (logamp) after the detector. If it is assumed much faster than the loop action, then:

$$F(x) = k_M \ln\left(\frac{k_D\, x}{j_M}\right) \tag{3.19}$$

$$\frac{\mathrm{d}F(A_{o,q})}{\mathrm{d}x} = \frac{k_M}{A_{o,q}} = \frac{k_M}{A_{i,q}\, G(C_q)} \tag{3.20}$$

in which k_M is the logamp large-signal gain and j_M is called the intercept, as depicted in Fig. 3.5. Remark that Eq. (3.20) expresses a simple gain, though still dependent on the input amplitude.

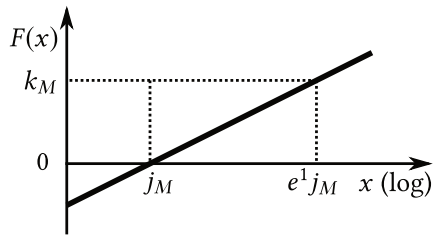


Figure 3.5: Transfer curve of a logarithmic amplifier.

Invoking Eq. (3.7) yields:

$$T(s) = \left[\frac{1}{G(C_q)} \frac{\mathrm{d}G(C_q)}{\mathrm{d}c}\right] k_M L(s) \tag{3.21}$$

Clearly, if the factor between brackets is made constant, the loop gain is independent of the operating point (given a fixed loop filter). Solving the resulting differential equation for the required VGA function and loop gain, yields:

$$G(C) = k_G \exp\left(\frac{C}{j_G}\right) \tag{3.22}$$

$$T(s) = \frac{k_M}{j_G} L(s) \tag{3.23}$$

The VGA must be an exponential function of its control signal (see Fig. 3.6). In Eq. (3.22), k_G and j_G are the VGA large-signal gain and intercept, respectively. Together with the

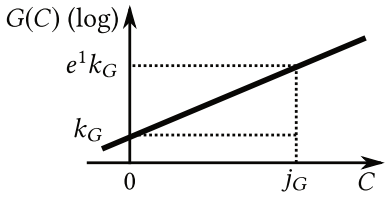


Figure 3.6: Transfer curve of a exponential amplifier.

loop filter L, these parameters determine the dynamics of the loop. For stability reasons, the loop gain should cross the unity-gain frequency with an approximate slope of 20 dB per decade. If L is approximately an integrator with unity-gain frequency f_L, the unity-gain frequency of the loop becomes:

$$f_{0\text{dB},T} = \frac{k_M}{j_G} f_L \tag{3.24}$$

which is a function of the logamp gain k_M and VGA intercept j_G, but not of the logamp intercept j_M, the detector gain k_D, nor the VGA gain k_G. In the circuit implementation (Chapters 4 and 5), those parameters will be used as scaling constants to obtain practical node voltages or branch currents.

The closed-loop frequency response with respect to Δa_i is constructed graphically from Eq. (3.15), as depicted in Fig. 3.7 and exhibits high-pass behavior. We write:

$$\frac{\Delta a_o}{\Delta a_i} = \frac{G(C_q)}{1 + \frac{2\pi f_{0\text{dB},T}}{s}} \tag{3.25}$$

with:

$$G(C_q) = k_G \exp\left(\frac{C_q}{j_G}\right) \tag{3.26}$$

in which the inverted zero due to non-infinite dc loop gain is neglected. The construction shows that an amplitude-independent inverted pole is introduced at the loop unity-gain frequency. In other words: the system filters out any variation of the amplitude envelope of the input signal. $G(C_q)$ is now the midband gain and is evidently still a function of the input amplitude. The settling time is constant and given by (2 %):

$$T_s \approx \frac{4}{2\pi f_{0\text{dB},T}} \tag{3.27}$$

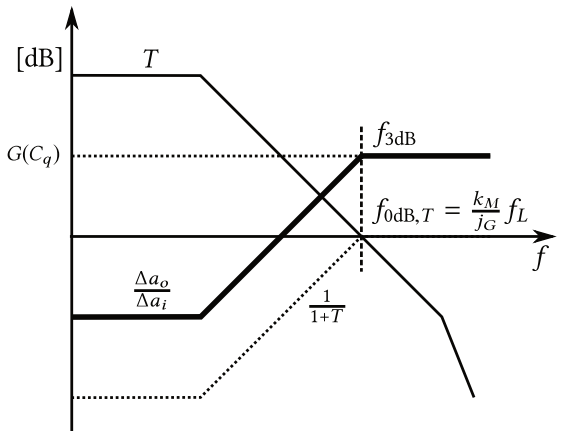


Figure 3.7: Graphical construction of the small-signal response due to Δa_i, at a given equilibrium.

In contrast to the settling time, the static output amplitude error when a unit input amplitude step is applied to Δa_i, is not constant. Using Laplace's limit theorem and referring to Eq. (3.15):

$$\epsilon_\infty = \lim_{s\to 0} s \frac{1}{s} \frac{G(C_q)}{1+T(s)} \approx \frac{G(C_q)}{T(0)} \tag{3.28}$$

The compression ratio M_c, introduced in Section 3.1.2, can be approximated by the ratio of relative change of the input level to the relative change of the output level [55]:

$$M_c = \frac{\left.\frac{A_i}{A_{i,q}}\right|_{\text{dB}}}{\left.\frac{A_o}{A_{o,q}}\right|_{\text{dB}}} \approx \frac{\frac{\Delta a_i}{A_{i,q}}}{\frac{\Delta a_o}{A_{o,q}}} \tag{3.29}$$

which, given Eq. (3.15) and Eq. (3.2), yields in steady-state:

$$M_c \approx 1 + T(0) \tag{3.30}$$

Hence the loop gain actually *is* the compression factor. Equation (3.28) and Eq. (3.30) show that in a delayed AGC system, a high dc loop gain is desirable for minimum output amplitude deviation over the effective AGC range.

3.1.4 Conclusion

Figure 3.8 summarizes these findings. In order to obtain constant settling time of the closed-loop system with a fixed integrator in the loop, the VGA transfer curve must be exponentially dependent on its control signal and a logamp must be included after the detector. In a sense, these operations convert the multiplication of the VGA to addition and subtraction. For convenience, the detector gain and logarithmic operation are also applied to the reference.

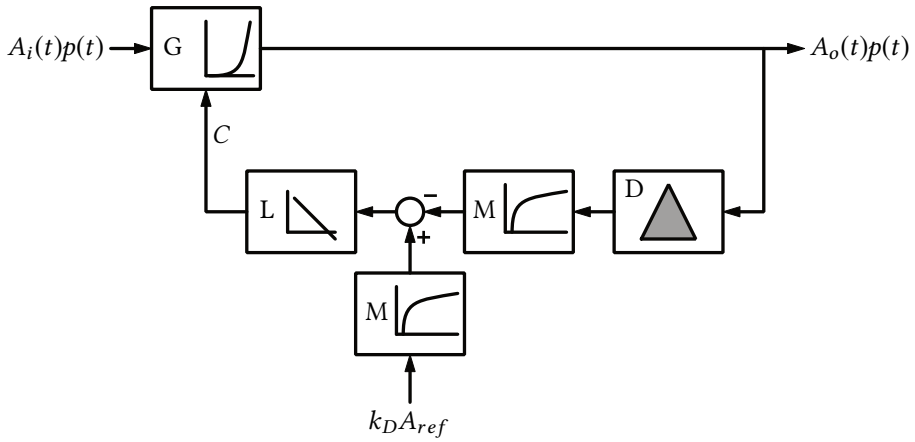


Figure 3.8: Continuous-time feedback AGC system, including exponential VGA and logamp to obtain constant settling time.

3.2 Event-driven Digitally Controlled AGC

The AGC system is to be embedded in a very high-speed linear optical receiver operating on multilevel modulation formats (see Chapter 1). As such, the VGA in Fig. 3.8 represents the high speed datapath of the receiver. It is similar to the one described in Chapter 6. Its input stage is a single-ended transimpedance amplifier, connected to one input of a differential main amplifier. A control loop removes the dc offset between the differential output signals by adjusting the dc voltage at the other input of the main amplifier, thus providing balanced differential output signals.

The VGA should be exponentially dependent on its control signal. While significant work is already required for a single stage, imposing a combined exponential response to the totality of the stages in an pure analog continuous fashion requires an inordinate amount of effort. The total (noise) bandwidth should remain constant to preserve signal fidelity across the input range and multilevel modulation formats require controlled overshoot behavior [18]. This adds considerable difficulty as several interacting circuit elements must be changed in tandem according to a possibly complex function. Also, in some feedback-based topologies, a change of gain could change the local loop gain, potentially leading to undesired peaking or even instability. Furthermore, trimming the control law on-chip, e.g. with zener zapping, would be inconvenient, costly and permanent. Clearly, pure analog control is undesired and not flexible enough in the light of the high-speed amplifier with high-frequency performance dominated by parasitics.

The solution to reduce complexity, proposed in this work, is to approximate the exponential VGA curve by a discrete set of fixed gains, which are selected by a digital controller governed by the detected signal. Indeed, by incorporating some sort of look-up table (LUT), both the gain and the frequency response of each separate stage can be shaped in order to obtain an overall desired response. This is akin to digitally assisted analog systems [64, 65].

The following sections will describe the system-level aspects of such an implementation. Section 3.2.1 introduces a VGA controlled by a LUT. In Section 3.2.2, the quantization of the AGC loop and its implications on closed-loop behavior are discussed. A quasi-

linearization of the system, based on describing functions, is presented in Sections 3.2.3, 3.2.4 and 3.2.6, while a method to avoid limit cycles is discussed in Section 3.2.5. Secondary aspects are briefly touched upon in Section 3.2.7.

3.2.1 LUT-based VGA

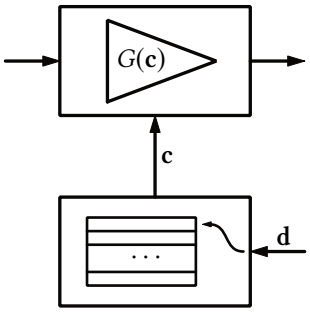


Figure 3.9: LUT-based VGA.

Figure 3.9 depicts the idea of a datapath—the VGA—controlled by a LUT, which is essentially a memory block. The VGA is composed of multiple stages. The LUT contains a number of N_c-bit code words, denoted *states*. A control word $\mathbf{d}$ with length N_d bits, representing a quantized signal related to the detected output amplitude, selects the appropriate state to be applied, as a word $\mathbf{c}$, to the VGA. As a result, the gain and frequency response (and possibly other parameters) of the datapath are programmed. Compared to a pure analog implementation and depending on the process technology, potentially huge silicon area and power savings are possibly when the LUT is implemented in CMOS logic. Another advantage is the possibility to reprogram the LUT in the field. Also, reuse in other designs is promoted when a behavioral description of the LUT in a hardware description language is available as this can easily be re-synthesized. A disadvantage, however, is that control accuracy is reduced as the system is quantized and measures must be taken to rule out limit cycling (Section 3.2.5).

3.2.2 Quantization of the AGC Loop

At some point in the loop, a quantized version of a signal related to the detected output amplitude is to be extracted, the choice of which dictates what blocks must be implemented in the quantized domain. Referring back to Fig. 3.8, several possible locations can be observed.

System output The quantization of the high-speed output requires a high-speed quantizer. Implementation at the projected data rate would be extremely difficult and power-hungry.

Detector output The detector output is a lower-speed signal. Indeed, it detects (a function of) the envelope of the output signal, which requires some kind of integration. In this case both the logamp and loop filter must be implemented in the quantized domain.

Loop filter output The integrated error signal is the slowest signal. The logamp and loop filter must be analog.

In this work, the quantizer is inserted at the loop filter output as this avoids the use of clocks or delays. This will be explained next.

Up to this point, no mention has been made of discretization in time. Indeed, the proposed LUT-based VGA imposes essentially no constraints on the time-behavior of the control words, as long as the VGA gain is allowed to settle between code word changes. This leaves the choice whether to include some kind of uniform clocked sampling or let the sampling rate be dictated by the detected signal itself.

A fixed system clock allows straightforward implementation of the loop filter and logamp in the digital domain (CMOS logic). However, the digitized signal is subject to aliasing. Hence the (constant) sampling the rate must be chosen based on the highest expected frequency. Assuming this can work, a lot of unneeded samples would be generated once settled, resulting in a waste of power, as the incident power to an optical receiver is fairly constant[2]. In addition, running a full-swing CMOS clock next to a sensitive (transimpedance) amplifier is greatly undesired, although a differential implementation and other measures (Section 2.5.4) could mitigate the effects to some extent. Moreover the required clock distribution tree would increase the interference coupling mechanisms.

In a clockless design, a new sample is generated only when the signal crosses a quantization threshold (level-crossing quantization [66]). As the sample rate is now proportional to the activity of the input signal, this is denoted *event-driven sampling* [67]. This results in activity-dependent power dissipation for the clock-less digital parts. In addition, the quantizer output spectrum is alias-free and the omission of a clock distribution tree reduces the number of coupling mechanisms. The umbrella term "event-driven" has been coined for these kind of systems [68]. The quantizer is an event-driven analog to digital converter (ADC) and the resulting signal is processed using event-driven digital signal processing (DSP). The quantized signal can be represented in parallel form (e.g. binary or thermometer) or as a kind of asynchronous delta encoding. The event-driven processing can be clocked, defeating the purpose, or work in continuous-time. It uses conventional asynchronous logic and delays, the latter which must be tuned and matched [68, 69].

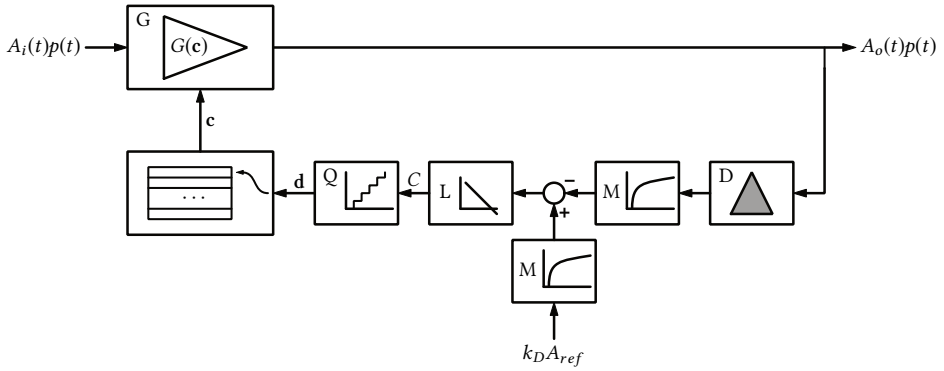


Figure 3.10: Event-driven digitally controlled AGC loop.

A clockless design is chosen for the advantages stated above. Furthermore, delay elements were deemed undesired as the implementation of these delays, required for the digital processing, either uses an external time reference, inverter chain or a clock after all and requires tuning. Hence, as shown in Fig. 3.10, the quantizer is inserted at the loop

[2] In the case of the continuous-wave (CW) transmission in the context of this work.

filter output while the loop filter itself and logamp remain analog. The amplified loop error signal C is quantized. The LUT is a pure combinatorial circuit implemented in CMOS. The system is now an event-driven digitally-controlled AGC loop.

3.2.2.1 Quantizer

For analysis purposes, the LUT including conversion to binary code words **d** can be bypassed by denoting the continuous-time output of the quantizer $Q(C)$ as D and assuming it to be directly connected to an equivalent continuous-time VGA $G(D)$. The output of the loop filter is C.

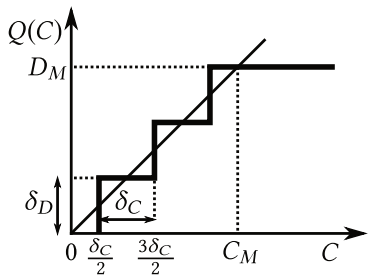


Figure 3.11: Uniform quantizer input-output characteristic.

The quantization is uniform. The input-output characteristic of the quantizer can be described as (Fig. 3.11):

$$Q(C) = \sum_{i=1}^{N_Q} \delta_D \mathcal{H}\left(C - i\delta_C + \frac{\delta_C}{2}\right) \tag{3.31}$$

in which $\mathcal{H}(\cdot)$ is the Heaviside step function [70]. The quantizer input and output step (bin) are δ_C and δ_D, respectively. The number of steps equals N_Q. It is asymmetric (not odd) and of the mid-thread type. As the number of quantization steps increases, the quantizer curve approaches a straight line to a progressively better degree, its slope being:

$$\tilde{k}_Q = \frac{\delta_D}{\delta_C} \tag{3.32}$$

In Section 3.2.6, Eq. (3.32) is shown to be the approximate equivalent gain of the quantizer. The residual difference between the output and the input of the quantizer can be modeled as additive noise (although not independent from the input signal) and is called quantization noise.

3.2.2.2 Limit Cycles

The quantizer represents a new hard nonlinearity in the system and the possibility of limit cycles must be investigated. An intuitive argument will now be given to indicate the existence of these oscillatory responses. A more rigorous study follows in the next sections.

Without loss of generality, the logamps can be ignored. The loop filter output is quantized. However, the control loop will still steer the VGA gain such that the average of

the detected signal minus (the average of) the desired signal is zero. In other words, such that the average of the detected signal equals (the average of) the desired signal. This means that variations of the detected signal are still allowed, as long as they are fast enough (compared to the loop bandwidth). It follows that it is possible that the VGA cycles between two adjacent gains, as long as the average gain makes the error signal small enough. These periodic limit cycles give rise to periodic quantization noise. Section 3.2.5 goes deeper into this matter.

3.2.3 Describing Functions of the Quantizer

The linearization approach for *incremental* changes (calculating the total derivative in a given nominal operating point), used for the continuous-time system (Section 3.1.3), can not be applied to this new system which contains hard discontinuities. Indeed, the derivative of the quantizer input-output relation $Q(C)$ is a sum of shifted Dirac delta functions, which is not continuous nor bounded. The resulting loop gain would either be zero or infinite.

Instead the operation of the nonlinear element for changes of *finite* size is approximated by a linear operation. This is called quasi-linearization as the resulting linear system depends also on signal amplitude. When the signal presented to the nonlinearity is assumed to have a certain form, the quasi-linearization results are called *describing functions* [58]. Three conditions must be satisfied for the describing functions to be meaningful. First, the nonlinear element is time-invariant. Second, no subharmonics are generated by the nonlinear element in response to sinusoidal input. Both are true for the quantizer. Third, the input to the nonlinearity is approximately sinusoidal (the *filter hypothesis*). This condition is true to the extend that the loop filter, logamp and detector dynamics provide enough attenuation of higher order harmonics of a potential limit cycle oscillation. Let's assume this to be the case and seek validation when required.

Assume a limit cycle exists. Then the input of the quantizer can be approximated as the sum of two components: the amplified average error signal C_ϵ and a filtered periodical signal C_s[3]. If the system input does not change, C_ϵ is only slowly changing with respect to the limit cycle amplitude A over the cycle period T_c:

$$T_c \left| \frac{\mathrm{d}C_\epsilon}{\mathrm{d}t} \right| \ll A \tag{3.33}$$

As Eq. (3.33) is satisfied, the dual-input describing function (DIDF) analysis can be applied. The quantizer input is assumed to be:

$$C(t) = C_\epsilon + C_s(t) \tag{3.34}$$

$$= B + A \sin\left(2\pi \frac{t}{T} + \theta\right) \tag{3.35}$$

The describing functions for both components (in the presence of each other) represent an equivalent quantizer (Fig. 3.12). It can be shown that the describing functions for

[3] Another way to see that the quantizer input must include a bias term is the following: if the system is limit cycling, a bias term must necessarily develop because of the asymmetric nonlinearity, by rectification action.

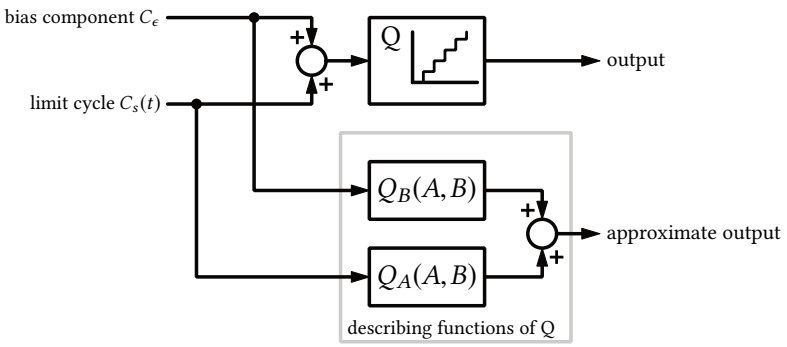


Figure 3.12: Describing functions as approximators of the quantizer (based on [58]).

the limit cycle C_s (amplitude A) and the bias component C_ϵ (level B) can be written as [58]:

$$Q_A(A,B) = \frac{\delta_D}{A}\sum_{i=1}^{N_Q}\left[q\left(\frac{(2i-1)\frac{\delta_C}{2}+B}{A}\right)+q\left(\frac{(2i-1)\frac{\delta_C}{2}-B}{A}\right)\right] \tag{3.36}$$

$$Q_B(A,B) = \frac{\delta_D}{B}\sum_{i=1}^{N_Q}\left[p\left(\frac{(2i-1)\frac{\delta_C}{2}+B}{A}\right)-p\left(\frac{(2i-1)\frac{\delta_C}{2}-B}{A}\right)\right] \tag{3.37}$$

with helper functions:

$$p(x) = \begin{cases} \frac{-1}{2} & x < -1 \\ \frac{1}{\pi}\arcsin x & |x| <= 1 \\ \frac{1}{2} & x > 1 \end{cases} \tag{3.38}$$

$$q(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2} & |x| <= 1 \\ 0 & |x| > 1 \end{cases} \tag{3.39}$$

Equations (3.36) and (3.37) are a function of both the bias component level B and the limit cycle amplitude A (at the input of the quantizer). They are non-phase-shifting as the quantizer is memoryless. They are also independent of frequency as the quantizer output does not depend on the derivative of its input (it does not matter how fast the input changes).

Graphical representations of Eqs. (3.36) and (3.37), normalized to $\tilde{k}_Q$ (Eq. (3.32)), are shown in Figs. 3.13 to 3.16 for a two-step quantizer ($N_Q = 2$). Consider the describing function of the limit cycle $Q_A(A,B)$ without bias ($B = 0$). For amplitudes smaller than half a step, the output of the quantizer does not change, hence its gain[4] is zero. For a normalized input of $\sqrt{2}/2$, the normalized gain reaches a maximum of $4/\pi$. For large input values within the input range of the quantizer, the gain approaches 1, after which it starts to roll off for even larger inputs. Note however, that for nonzero B, the gain can be significantly higher as the bias term approaches a multiple of $\delta_C/2$. For smaller amplitudes, maximum gain occurs as B approaches a transition point (Fig. 3.15). For B equal to a multiple of $\delta_C/2$, the gain is infinite for an infinitesimal small amplitude as the

[4] In this context 'gain' and 'describing function' will be used interchangeably.

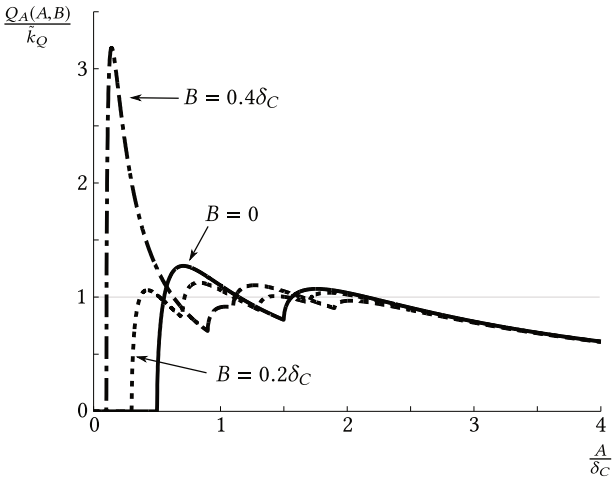


Figure 3.13: Normalized describing function of the limit cycle for various dc biases, as a function of limit cycle amplitude (two-step quantizer).

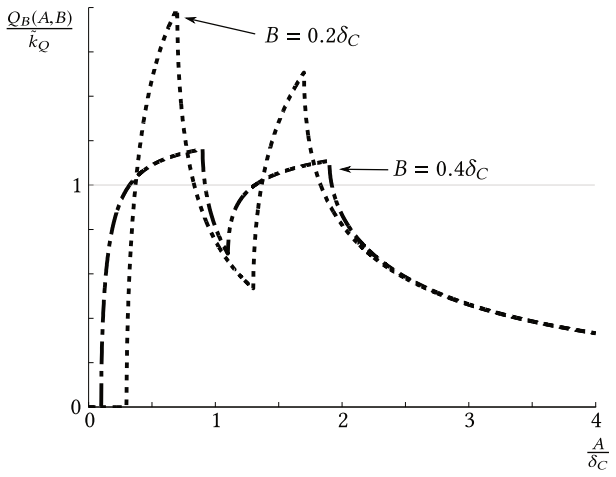


Figure 3.14: Normalized describing function of the bias term for various dc biases, as a function of limit cycle amplitude (two-step quantizer).

limit cycle will be centered around a transition point of the quantizer and even a very small disturbance will change its output.

Now consider $Q_B(A, B)$, the describing function of the bias term in the presence of a limit cycle. For B equal to a transition point, it is unity as the average quantizer input equals the average output (Fig. 3.16). If the average output of the quantizer is greater than B, the bias gain is higher than unity and vice versa. It follows that if B approaches a transition point from below (above), the describing function of the bias term is always smaller (higher) than unity for any amplitude.

Remark that the incremental-input describing function (IIDF) cannot be used, as no restrictions are applied on the bias term. This once again shows that incremental lin-

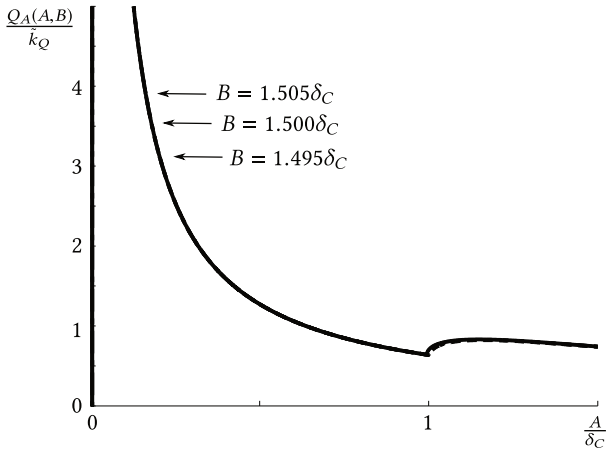


Figure 3.15: Normalized describing function of the limit cycle for dc biases around $1.5\delta_C$, as a function of limit cycle amplitude (two-step quantizer).

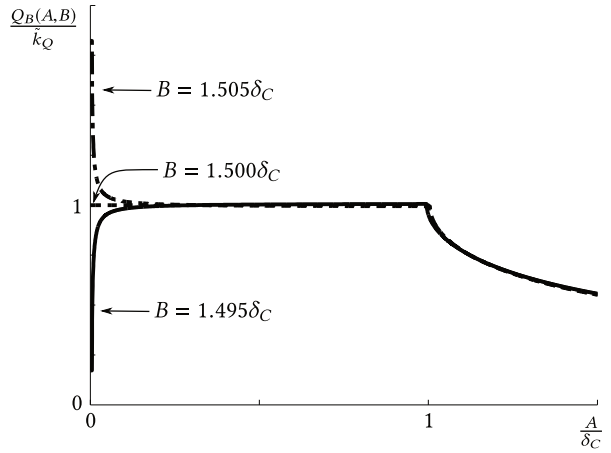


Figure 3.16: Normalized describing function of the bias term for dc biases around $1.5\delta_C$, as a function of limit cycle amplitude (two-step quantizer).

earization does not apply in this system.

3.2.4 Linearization of the Event-Driven AGC Loop

Using the describing functions developed in Section 3.2.3, the linear model of Section 3.1.3 will be extended to include the effect of the quantizer on the AGC loop. Referring to Fig. 3.10 and Eq. (3.23), denote the linear part of the loop gain as T_{lin}:

$$T_{\text{lin}}(s) = \frac{k_M}{j_G} L(s) \tag{3.40}$$

3.2.4.1 Existence of Limit Cycles

In the given system, for a limit cycle to exist, both the bias term and the limit cycle sinusoid must propagate around the loop unattenuated [58]. This leads to a dynamic as well as a static condition on the existence of limit cycles.

The *dynamic* condition is evaluated on the incremental system approximation of Section 3.1.3 (in the presence of the bias term B). Consider Fig. 3.4 and set the incremental inputs zero. The VGA and logamp are replaced by their incremental linearization, while the quantizer is replaced by the describing function of the sinusoid $Q_A(A,B)$ (large-signal linearization). The detector was already replaced by a simple gain when the amplitude-domain model was set up. A sustained sinusoidal oscillation with frequency f_0 requires:

$$Q_A(A,B)\,T_{\text{lin}}(j2\pi f_0) = -1 \tag{3.41}$$

or:

$$Q_A(A,B) = \frac{1}{|T_{\text{lin}}(j2\pi f_0)|} \tag{3.42}$$

$$\angle T_{\text{lin}}(j2\pi f_0) = -180^\circ \tag{3.43}$$

As $Q_A(A,B)$ is non-phase-shifting, Eq. (3.43) shows that the potential limit cycle frequencies will be those where the phase of the linear part of the loop gain crosses -180°. In a physical implementation such frequencies always exist. Note that $Q_A(A,B)$ is the gain margin of $T_{\text{lin}}(j2\pi f)$ when limit cycling. Ultimately, the non-dominant poles of the loop gain determine the limit cycle frequency.

The *static* condition is evaluated on the static, large-signal model of the AGC system (in the presence of the limit cycle A). Consider the large-signal amplitude-domain in Fig. 3.3 and mentally add the quantizer by including the describing function for the bias component $Q_B(A,B)$ after the loop filter. Note that the VGA is exponential and the two logamps are included (Fig. 3.8). When a test signal is injected at the input of the quantizer, it is seen that a dc signal B can propagate around the loop when (invoking Eqs. (3.19) and (3.23)):

$$B = L(0)\left[k_M \ln\left(\frac{k_D}{j_M}A_{ref,q}\right) - k_M \ln\left(\frac{k_D}{j_M}A_{i,q}G(C_q)\right)\right] \tag{3.44}$$

with (Eq. (3.22)):

$$G(C_q) = k_G \exp\left(\frac{C_q}{j_G}\right) \tag{3.45}$$

$$C_q = B\,Q_B(A,B) \tag{3.46}$$

This yields:

$$B = L(0)k_M\left[\ln\left(\frac{A_{ref,q}}{A_{i,q}}\right) - \frac{C_q}{j_G}\right] \tag{3.47}$$

$$= L(0)k_M\left[\ln\left(\frac{A_{ref,q}}{A_{i,q}}\right) - \frac{B\,Q_B(A,B)}{j_G}\right] \tag{3.48}$$

Solving this condition for $Q_B(A, B)$ yields:

$$Q_B(A,B) = \frac{j_G}{B} \ln\left(\frac{A_{ref,q}}{A_{i,q}}\right) - \frac{j_G}{k_M L(0)} \tag{3.49}$$

$$\approx \frac{j_G}{B} \ln\left(\frac{A_{ref,q}}{A_{i,q}}\right) \tag{3.50}$$

in which the approximation holds for high dc loop gain. The bias term depends on both the amplitude of the reference signal and the input signal, which agrees with intuition.

It can be deduced graphically that for any given input amplitude $A_{i,q}$ within the effective AGC range, a set (A, B) exists that satisfies both conditions (Eqs. (3.41) and (3.49)) for a given $A_{ref} = A_{ref,q}$. A limit cycle will eventually develop with amplitude (at the input of the quantizer) smaller than a quantization step. The VGA will toggle between exactly two states. The pathological situation where the detected amplitude exactly equals the reference amplitude will quickly vanish due to noise and other imperfections, such that in practice a limit cycle will always develop (again, for input signals within the effective AGC range). This has been confirmed with simulations.

3.2.4.2 Limit Cycle Stability

The number of potential limit cycles is bounded by Eq. (3.41). Although not applied here, methods exist to determine their stability[5] [58]. The intuitive explanation given in Section 3.2.2 suggests that at least one limit cycle must be stable. This has been confirmed by numerical simulations.

3.2.5 Avoiding Limit Cycles

At this point, one should wonder whether or not these limit cycles should be avoided. For a large number of quantization steps (fine granularity) the system approximates better its continuous-time equivalent, the gain steps are small and the quantization noise at the output is buried in the noise generated by the datapath of the receiver. However, it takes more effort to populate the entire LUT, which increases linearly in size with each extra step, as for each code word the gain and frequency response must be tailored. This could be dealt with by using automatic calibration, though this is not further considered. In any case, complexity increases again, which needs to be avoided. For a small number of quantization steps (coarse granularity), the quantization noise is dominant and will deteriorate sensitivity of the receiver. In addition, the resulting big output amplitude swing can potentially upset subsequent circuits. All this impacts performance negatively.

Furthermore, in both cases the continuous rail-to-rail switching of voltages in a rather large digital block consumes power to no avail and is a source of interference. This further impacts performance, not only of the datapath under consideration itself, but also of adjacent channels in a multichannel receiver. These are the same reasons a clockless design was chosen in the first place (Section 3.2.2).

In summary, limit cycling in an optical receiver is not desired. One way of achieving better behavior is to allow a range of amplitudes at the system output. Indeed, most data processing systems tolerate a certain margin. Figure 3.17 depicts the operation principle.

[5] A limit cycle is stable if it returns to its original state in response to a perturbation of its amplitude or frequency. It is called unstable otherwise.

A window comparator W indicates if the detected signal lies in the allowed range, in order to inhibit the LUT from changing its output—thereby fixing the VGA gain:

$$W(x) = \mathcal{H}(x - A_{o,\min}) - \mathcal{H}(x - A_{o,\max}) \tag{3.51}$$

The window bounds are defined as:

$$A_{o,\min} = A_{ref}(1 - \alpha) \tag{3.52}$$
$$A_{o,\max} = A_{ref}(1 + \alpha) \tag{3.53}$$

with $0 < \alpha < 1$. The window is centered around A_{ref} and is $2\alpha A_{ref}$ wide. $\mathcal{H}(\cdot)$ is the Heaviside step function [70].

The system is now time variant, precluding the further use of describing functions. In essence, the control loop is broken for signals within range; there $Q_A(A, B)$ and $Q_B(A, B)$ are effectively reduced to zero. When the loop is closed—i. e. when the system is settling toward a steady-state—the dynamics can still be adequately predicted by the theory developed in the previous sections. However, as soon as `inh` is asserted, the settling response is cut short. This results in a shorter effective settling time for a smaller number of quantization steps. The settling time, however, is still upper-bounded by its equivalent continuous-time value. Hence, infinitely long transient response cannot occur as in the case of a general AGC system. As already pointed out in Section 3.1.3, this is a desirable feature.

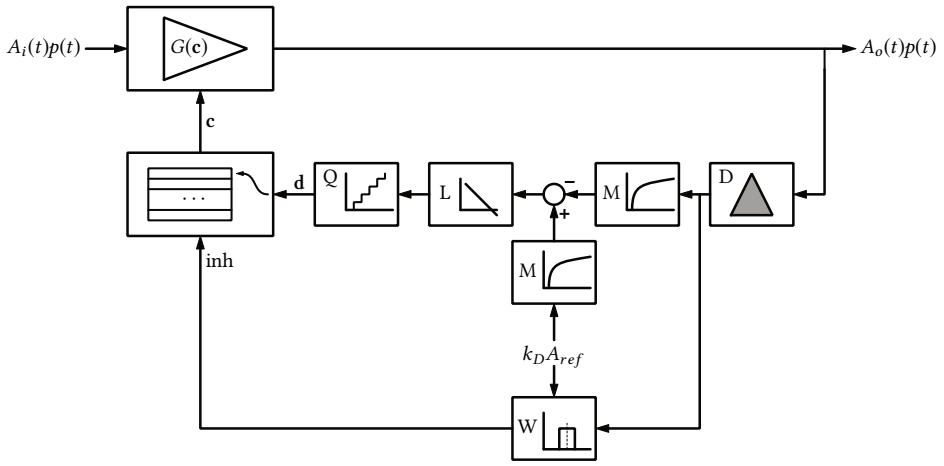


Figure 3.17: Event-driven digitally controlled AGC loop with window comparator to avoid limit cycling.

3.2.5.1 Minimum Number of Quantization Steps

The allowed output amplitude window is usually fixed by higher-level specifications, thus the minimum number of quantization steps N_Q must be determined. For this and the following analyses, the LUT will be bypassed, as described in Section 3.2.2.1. At this point, it should be repeated that gain control only occurs for input amplitudes between $A_{i,\min}$ and $A_{i,\max}$; outside of this range the gain does not change (see Fig. 3.2). For a

given A_{ref}, minimum and maximum VGA gains can be associated with the input interval $[A_{i,\min}, A_{i,\max}]$:

$$G_{\min} = \frac{A_{ref}}{A_{i,\max}} \tag{3.54}$$

$$G_{\max} = \frac{A_{ref}}{A_{i,\min}} \tag{3.55}$$

These gains are selected when the quantizer output D is at its extremes. For each quantization level k, the output $D^{(k)}$ of the uniform quantizer is applied to the VGA gain $G(C)$, which is an exponential function of C. The resulting quantized gain is denoted $\hat{G}^{(k)}$. It follows that the logarithm of the resulting gain is uniformly quantized with step size of:

$$\ln \hat{G}^{\Delta} = \ln \hat{G}^{(k+1)} - \ln \hat{G}^{(k)}, \qquad k = 0 \ldots N_Q - 1 \tag{3.56}$$

$$= \frac{1}{N_Q} (\ln G_{\max} - \ln G_{\min}) \tag{3.57}$$

Limit cycling is avoided when the window size is at least as big as the output amplitude step size, for any input. In other words: for any input amplitude A_i (in the active AGC range), a quantized gain $\hat{G}^{(k+1)}$ must exist such that the resulting output amplitude A_o lies within the range $[A_{o,\min}, A_{o,\max}]$. Formulated:

$$\hat{G}^{(k)} A_i < A_{o,\min} \tag{3.58}$$

$$\hat{G}^{(k+1)} A_i \geq A_{o,\min} \tag{3.59}$$

$$\hat{G}^{(k+1)} A_i < A_{o,\max} \tag{3.60}$$

Taking the logarithm of Eq. (3.60) and introducing Eq. (3.56) yields:

$$\ln \hat{G}^{(k+1)} + \ln A_i < \ln A_{o,\max} \tag{3.61}$$

$$\ln \hat{G}^{(k)} + \ln \hat{G}^{(\Delta)} + \ln A_i < \ln \frac{A_{o,\max}}{A_{o,\min}} + \ln A_{o,\min} \tag{3.62}$$

From Eq. (3.58) and Eq. (3.62) follows:

$$\ln \hat{G}^{(\Delta)} < \ln \frac{A_{o,\max}}{A_{o,\min}} \tag{3.63}$$

Or with Eq. (3.57):

$$N_Q > \frac{\ln \frac{G_{\max}}{G_{\min}}}{\ln \frac{A_{o,\max}}{A_{o,\min}}} \tag{3.64}$$

which yields:

$$N_Q > \frac{\ln \frac{A_{i,\max}}{A_{i,\min}}}{\ln \frac{A_{o,\max}}{A_{o,\min}}} \tag{3.65}$$

$$N_Q > \frac{\ln \frac{G_{\max}}{G_{\min}}}{\ln \frac{1+\alpha}{1-\alpha}} \tag{3.66}$$

$$N_Q > \frac{\ln \frac{A_{i,\max}}{A_{i,\min}}}{\ln \frac{1+\alpha}{1-\alpha}} \tag{3.67}$$

Equation (3.65) expresses that the number of required quantization steps depends only on the ratios of the valid input and output dynamic range.

If the quantized output D is to be encoded in a binary word $\mathbf{d}$, it makes sense to exhaust all possible values of $\mathbf{d}$ and choose the number of quantization levels to be a power of 2:

$$N_Q \leftarrow 2^{\lceil \log_2 N_Q \rceil} \tag{3.68}$$

in which $\lceil \cdot \rceil$ represents the ceiling function.

3.2.6 Approximative Linearization of the Event-Driven AGC Loop

The linearization described in Section 3.2.4 is used to predict the existence of limit cycles. It is, however, less convenient for design because of the signal dependence of the describing functions. Figures 3.13 and 3.15 suggests that $Q_A(A, B)$ can be approximated by δ_D/δ_C for signals within the quantizer input range. The biggest deviation from δ_D/δ_C occurs for B around the quantizer transition points. However, this region of operation is avoided by the system described in Section 3.2.5.

It can be shown formally that, as the number of quantization steps tends to infinity, the equivalent continuous-time system appears and the describing function tends to the constant ratio:

$$\lim_{N_Q \to \infty} Q_A(A, B) = \frac{\delta_D}{\delta_C} = \tilde{k}_Q \tag{3.69}$$

This is readily seen, as the quantizer characteristic approaches a straight line with the given ratio as slope. Although in the implemented system the number of quantization steps is far from infinite, the approximation will be adopted for design guidance. The loop gain becomes:

$$T(s) \approx \tilde{k}_Q T_{\text{lin}}(s) \tag{3.70}$$

$$= \frac{\delta_D}{\delta_C} \frac{k_M}{j_G} L(s) \tag{3.71}$$

This approximation will be used, in conjunction with the expressions derived in Section 3.1.3, to implement the event-driven AGC system in Chapter 4.

3.2.7 Secondary Aspects

An implementation of the basic structure of the event-driven digitally controlled AGC (Fig. 3.17) demands further study of a number of additional aspects, to ensure system-level specifications can be met as well as to derive circuit-level requirements. The frequency response of the VGA, non-dominant poles in the loop, propagation delay in the logic building blocks and accuracy of the VGA exponential approximation are shortly discussed next.

3.2.7.1 High-pass Response of the VGA

Frequently, the VGA has a high-pass frequency response because of internal ac coupling or offset compensation loops. In Section 3.1, it was stated that the bandwidth of the VGA is assumed so high, that it can be simplified to a simple gain for the purpose of this work.

Ideally, this should have no influence on the detected signal. Indeed, the high-pass action reduces the time average of the time-domain output signal to zero, whereas the AGC loop works in the amplitude domain. However, in Sections 4.6 and 5.1, it will be shown that the detector operates on a balanced differential signal (the differential output), such that any transient response resulting from the balancing will prolong the detector settling time. One way to cope with this, is to make the AGC loop much slower than the balancing loop, as explained in Section 4.5.

3.2.7.2 Nondominant Poles

The finite speed of the analog blocks can be expressed by including extra factors in the linearizations presented in Section 3.1.3 and Section 3.2.4. The limited bandwidth is expressed by modeling each block as a low-pass filter of appropriate order (neglecting zeros). Note that, in general, the poles depend on the input signal amplitude. The linear part of the loop gain becomes:

$$T_{\text{lin}}{}'(s) = T_{\text{lin}}(s)\left(\prod_i \frac{1}{1+\frac{s}{\omega_i}}\right)\left(\prod_i \frac{1}{1+\frac{s}{Q_i\omega_{0i}}+\frac{s^2}{\omega_{0i}^2}}\right) \tag{3.72}$$

$$\approx \frac{k_M}{j_G}\frac{1}{\frac{s}{\omega_L}}\left(\prod_i \frac{1}{1+\frac{s}{\omega_i}}\right)\left(\prod_i \frac{1}{1+\frac{s}{Q_i\omega_{0i}}+\frac{s^2}{\omega_{0i}^2}}\right) \tag{3.73}$$

The closed-loop response is affected according to Eq. (3.15). Stability and transient response concerns call for worst case evaluation of the pole frequencies for a given input amplitude range. The additional poles should be placed far enough from the unity-gain frequency for a given closed-loop transient response specification.

The effects of finite logamp bandwidth and VGA gain switching speed can be included in this manner.

3.2.7.3 Propagation Delay and Glitches

When a code word is applied to the LUT-based VGA, the output word does not change instantaneously. The finite propagation delay τ_d of the combinatorial logic corresponds to an excess phase lag in the loop gain:

$$T_{\text{lin}}{}'(s) = T_{\text{lin}}(s)\exp^{-s\tau_d} \tag{3.74}$$

Together with the stability requirements, Eq. (3.74) constrains the maximum propagation delay.

Propagation delay mismatch of internal logic paths could result in glitching or even forbidden gain sequences as the LUT output code settles to its final value. However, in practice the delay mismatch is much smaller than the loop settling time and can be ignored.

3.2.7.4 Limited Accuracy of VGA Exponential Approximation

As explained in Section 3.1.3, the gain characteristic of the VGA should be exponentially dependent on its control signal. In the event-driven system, the VGA control signal is

quantized, which results a discrete set of valid gains lying on the ideal curve. Equally, the finite resolution of a LUT-based VGA introduces quantization of the gain values. Referring to Fig. 3.9, the finite word length of control signal **c** limits the accuracy. The quantization error can be modeled as an additive gain error dependent on the input signal. Small errors that preserve monotonicity correspond to shifted transition points of the quantizer. It follows that a non-uniform quantizer can be introduced to capture the quantization effects. Repeating the describing function analysis of Section 3.2.3 would reveal additional dependence of the closed-loop behavior on the input amplitude.

This is not pursued in detail as the description of the VGA as the datapath of an optical receiver is beyond the scope of this work. Furthermore, an exact match of the realized gains with the theoretical curve is not feasible nor required. Indeed, the system should be robust and tolerant to small imperfections. Suffice it to say that the resolution is chosen adequately high.

3.2.8 Conclusion

In this section, an event-driven implementation of the automatic gain control loop introduced in Section 3.1 was proposed from a system-level point-of-view. Both the structural and behavioral aspects were discussed. The control loop includes a LUT-based VGA and quantizer as the event-driven parts, in addition to analog blocks such as the detector, logarithmic amplifier and loop filter. The choice of location of the quantizer was discussed as well as the benefits of using a clockless design.

The non-linear nature of the quantizer was dealt with by applying a linearization technique known as describing function analysis. In this way, linear system design methods could be used. Additionally, the existence of limit cycles was predicted. They can be avoided by allowing a range of output amplitudes. This led to a requirement on the number of quantization steps. For design guidance, an approximate linearization of the system was introduced. Secondary aspects such as additional poles, propagation delay and finite resolution were shortly discussed.

The result is a system model of the quantized AGC loop. This model will serve as the basis for the system-level design and circuit implementation, described in Chapters 4 and 5, respectively.

Chapter 4

Event-driven AGC System-level Design

Based on the analysis in Section 3.1, a system-level design of the event-driven automatic gain control (AGC), embedded in a high-speed linear optical receiver, is now presented. Quantities involved will no longer be dimensionless, instead physical voltages and currents will be used. In the light of a feasable physical implementation, knowledge of potential circuit topologies including their strengths and weaknesses is used to maximum advantage. Although presented in a linear fashion, the design decisions require interaction between the system level and circuit implementation level. The result of the system-level design is the requirements on the loop gain and specifications for key lower-level building blocks: the variable gain amplifier (VGA), quantizer and loop filter.

A top-level block diagram is presented in Section 4.1, while the top-level specifications are discussed in Section 4.2. The important specifications for the VGA, quantizer and loop filter are derived Sections 4.3 to 4.5, followed by a short discussion of the detector and transconductor in Section 4.6.

4.1 Overview

Consider the block diagram in Fig. 4.1. VGA G converts input current I_i to a balanced differential voltage $V_o = V_{op} - V_{om} = G(\mathbf{c})\, I_i$. Detector D detects the output level (extracts a measure of the differential output amplitude) and has a differential output $V_{\text{det},p} - V_{\text{det},m}$. The logamps M require a current as input signal (Section 5.3). Hence, they are driven by a linear transconductor stage G_m which has multiple identical outputs I_{det}. The reference signal of the loop is current I_{ref}, which is a measure of the desired output amplitude. It has been implicitly multiplied by the (linearized) gains of the detector and transconductor, such that the desired output level is obtained. I_{ref} is generated by a digital to analog converter (DAC). Loop filter L, approximately an integrator, provides loop compensation. Its output is connected to quantizer Q, that converts the amplified loop error signal C to a digital code word $\mathbf{d}$. This signal selects a state $\mathbf{c}$ in the look-up table (LUT), which in turn governs the gain and frequency response of the VGA.

Both I_{ref} and I_{det} are also applied to the window comparator W. This block asserts digital signal `inh` as soon as the detected output amplitude is within the desired range

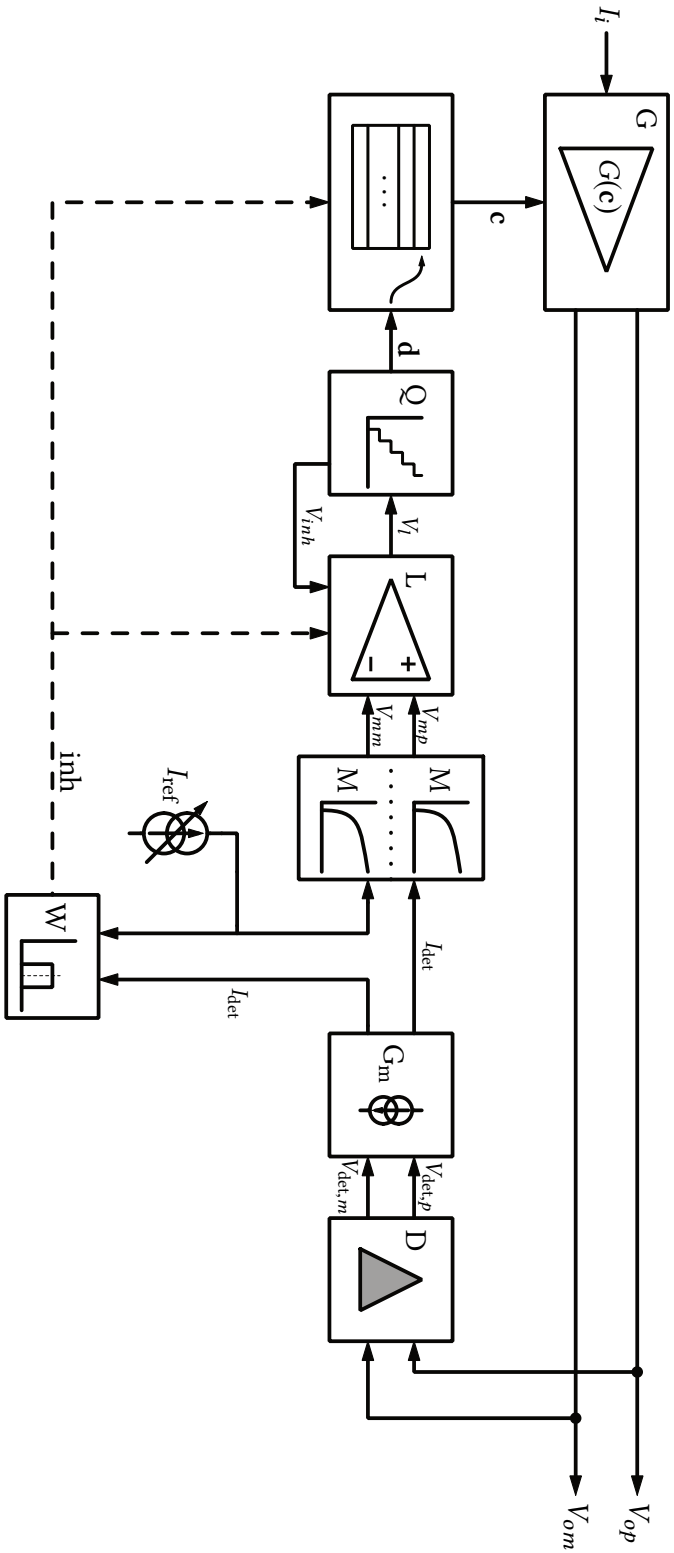


Figure 4.1: Functional block diagram of the implemented AGC system.

and inhibits the LUT from changing its state, effectively breaking the loop and keeping the VGA gain fixed. This avoids limit cycling (Section 3.2.5).

Transient Loop Restart Behavior When the loop is broken (`inh` asserted or input signal outside of the effective AGC range (Section 3.1.2)), the output voltage of the loop filter V_l will saturate against one of the supply rails. When the loop is reinstated (`inh` deasserted and input signal within the effective AGC range), it takes time for the amplifier to resume linear operation as its internal nodes recover from the overload condition. This situation can potentially prolong settling time and should be avoided. Therefore, two additional mechanisms are included around the loop filter.

First, when `inh` is asserted, the loop filter is automatically reconfigured as a follower, such that its output follows the midpoint voltage V_{inh} of the current transition bin in the quantizer. This voltage is easily generated in the quantizer (Section 5.5). Second, the maximum output voltage of the amplifier is limited to the full-scale input voltage of the quantizer by including a precision clamp. This circuit will not be further discussed.

4.2 Top-level Design Objectives

Table 4.1 shows the top-level specifications of the linear optical receiver. Specifications of no interest to the AGC system, such as jitter and sensitivity, are omitted. The receiver converts a single-ended photo current to a differential peak-to-peak output voltage of typically 400 mV. Any high-pass pole should be no higher than 500 kHz to allow for long strings of consecutive identical digits without appreciable power penalty due to baseline wander (Section 2.5.3). Ultimately the receiver is to be used with multilevel modulation formats (e.g. 4-PAM), however, all testing will be done using non-return-to-zero (NRZ) data. This is easier to generate with high quality at the required speeds and is readily available from measurement equipment. Furthermore, Chapter 5 will show that only the behavior of the detector depends on the shape of the data signal. No explicit limits on silicon area are imposed, in contrast with the four-channel receiver of Chapter 6. The typical NRZ data rate is 26 Gb/s.

Specification	Symbol	Unit	Min.	Typ.	Max.
Bandwidth	$f_{3\text{dB}}$	GHz		20	26
Bit rate (NRZ)	R_b	Gb/s		26	
Baud rate (4-PAM)		GBd		20	
Transimpedance range	R_T	Ω	125		2000
Dynamic range (optical; electrical)	DR	dB	12.5; 25		
Differential peak-to-peak output voltage	V_{od}^{pp}	mV	320	400	480
Common-mode output voltage	V_{oc}	V		2.1	
Low cut-off frequency	f_{pL}	kHz			500
Core supply voltage	V_{DD}	V		2.5	

Table 4.1: Top-level design objectives of the linear receiver.

The system is designed in the SiGe BiCMOS 0.13 µm process technology described in

Section 1.3. The main core is supplied by a 2.5 V rail. An auxiliary supply of 1.2 V is available for most digital logic and also provides a convenient mid-rail reference voltage for analog building blocks. Heavy local filtering and sensible layout are required to minimize interference.

4.3 VGA

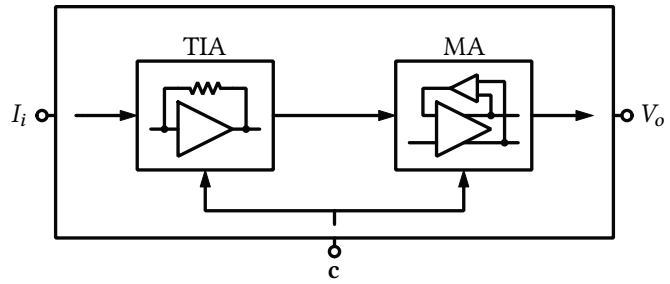


Figure 4.2: The datapath of an optical receiver is the VGA in the AGC control loop context.

The VGA $G(\mathbf{c})$ is the high speed datapath of the receiver (Fig. 4.2). In contrast with the receiver described in Chapter 6, this receiver is not hard limiting (for signals within the specified input range). A single-ended transimpedance amplifier (TIA) linearly converts unipolar photo current I_i to a voltage that is, ideally, *linearly* amplified by a main amplifier (MA). The MA is composed of three stages and a 50 Ω output driver.

Using an internal control loop, single-ended to differential conversion is provided such that the output $V_o = V_{op} - V_{om}$ is balanced. As shown in Section 6.5, this loop introduces a gain-dependent high-pass pole at f_{pL} in the input-output transfer function. Its frequency range $[f_{pL.\,\min}; f_{pL,\max}]$ is given by:

$$f_{pL,\min} = f_{pL,\max}\frac{G_{\min}}{G_{\max}} = 31\,\text{kHz} \tag{4.1}$$

with $f_{pL,\max} = 500\,\text{kHz}$, $G_{\min} = 125\,\Omega$ and $G_{\max} = 2000\,\Omega$ from Table 4.1. It is possible to obtain a constant pole frequency by making the gain-bandwidth product (GBW) of the balancing loop constant. This has not been implemented to limit the design complexity.

The VGA gain should be an exponential function of the control signal (Section 3.1). The digital signal **c** controls the gain, bandwidth and overshoot of the datapath. It is the quantized version of a conceptual continuous-time signal D (a voltage). The latter can be assumed in the range 0 V to 1 V such that, without any loss of generality, the minimum and maximum required gain are selected at $D = 0\,\text{V}$ and $D = 1\,\text{V}$, respectively. From Eq. (3.22), it follows that:

$$k_G = G(0) = G_{\min} \tag{4.2}$$

$$j_G = \ln^{-1}\frac{G_{\max}}{G_{\min}} \tag{4.3}$$

which yield $k_G = 125\,\Omega$ and $j_G = 0.36$ (V).

Table 4.2 summarizes the specifications of interest in this context for the datapath. They are calculated above or follow directly from the top-level design objectives (Table 4.1). It is beyond the scope of this work to discus the design of the datapath in detail.

4.4 Quantizer

This block quantizes the amplified loop error signal: $\mathbf{d} = Q(V_l)$. The specifications of the VGA suffice to select the number of quantization steps N_Q. The single-ended output amplitude window width α is 20 %, centered around 400 mV. Equations (3.66) and (3.68) yield:

$$N_Q = 8 \tag{4.4}$$

Hence the output of the quantizer is encoded in a 3-bit binary word $\mathbf{d}$.

As the conceptual continuous-time VGA control signal D is normalized to the range 0 V to 1 V, the quantizer output step is given by:

$$\delta_D = \frac{1\,\mathrm{V} - 0\,\mathrm{V}}{N'_Q - 1} = 143\,\mathrm{mV} \tag{4.5}$$

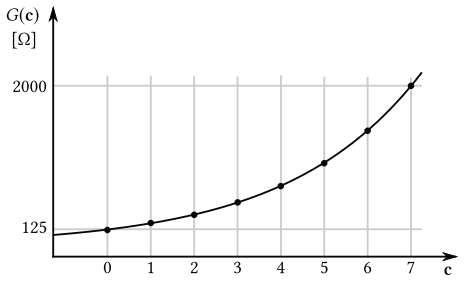
The quantizer input step is *chosen* as:

$$\delta_C = 200\,\mathrm{mV} \tag{4.6}$$

This results in a convenient input range of 1.6 V and, given the supply voltage of 2.5 V, allows for headroom and relaxed offset requirements in the circuits that realize the quantizer, such as internal comparators (Section 5.5). The approximate quantizer gain follows from Eqs. (3.69), (4.5) and (4.6):

$$\tilde{k}_Q = 0.715 \tag{4.7}$$

Table 4.2 summarizes the quantizer specifications while Fig. 4.3(a) shows a graphical depiction of the quantized exponential gain behavior of the VGA, as well as the ideal discrete VGA gains.



(a) Graphical representation.

$\mathbf{c}$	$G(\mathbf{c})$ [Ω]
0	125
1	186
2	276
3	410
4	610
5	906
6	1346
7	2000

(b) Ideal discrete VGA gain per code word.

Figure 4.3: Specification of the VGA and the quantizer.

4.5 Loop Design and Loop Filter

The design of the AGC control system (as any other feedback system) involves selecting the dc loop gain and gain-bandwidth product and designing the frequency response of the loop gain and null loop gain such that a desired closed-loop response is obtained over the operating range. The null loop gain(s) (that depend on the chosen input and output) can safely be assumed infinite, as any appreciable direct forward transmission through either the VGA or the logamp and detector can be assumed non-existent at this point.

In Section 3.1.1, it was argued that the loop gain should approximate an ideal integrator response. The choice of the loop GBW is governed by the following considerations: first, as mentioned in Section 3.1.1, the detector is assumed much faster than the loop. Second, as will be shown in Section 5.1, the detector requires a balanced differential signal at its input. This implies that the balancing loop must be much faster than the AGC loop, such that any transient response due to settling of the balancing loop has negligible effect on the VGA gain. Third, in order to make the settling time of the datapath predominantly dictated by the AGC loop (Eq. (3.27)), it makes sense to make the AGC loop much slower than the balancing loop. Hence, the typical GBW of the AGC loop is chosen conservatively a decade below the worst-case GBW of the balancing loop:

$$f_{0\text{dB},T} = \frac{f_{pL,\min}}{10} = 3.1\,\text{kHz} \tag{4.8}$$

which gives a 2 % settling time upper bound of 205 µs (Eq. (3.27)).

As the loop filter is not a perfect integrator, the dc loop gain will be finite such that:

$$T = \frac{T_0}{1 + \frac{s}{2\pi f_{0\text{dB},T}}} \tag{4.9}$$

with:

$$T_0 = \tilde{k}_Q \frac{k_M}{j_G} L_0 \tag{4.10}$$

$$f_{0\text{dB},T} = \tilde{k}_Q \frac{k_M}{j_G} f_L \tag{4.11}$$

where Eq. (4.11) has already been derived in Eq. (3.71). Non-dominant poles are ignored in Eq. (4.9). T_0 must be chosen high enough such that the integrator approximation holds. In addition, it determines the accuracy of the control loop. It is sensible to require at least $T_0 = 40$ dB dc loop gain. In practice, it is easy to obtain a higher value.

The VGA intercept j_G and the quantizer steps are already chosen in Sections 4.3 and 4.4. The remaining parameters to select are k_M, L_0 and f_L, the logamp large-signal gain, dc gain and GBW product of the loop filter, respectively. In Section 5.3, it is shown that the logamp forces a current through a silicon PN-junction such that $k_M = U_t$, the thermal voltage (26 mV at 300 K). Equations (4.8), (4.10) and (4.11) yield:

$$L_0 = 1840 = 65\,\text{dB} \tag{4.12}$$

$$f_L = 60\,\text{kHz} \tag{4.13}$$

In principle, the non-dominant poles of the loop filter could be placed before its GBW as the loop gain of the system is an attenuated version of the frequency response of the loop

filter (Eq. (4.11)). However, as the loop filter can be reconfigured as a follower when `inh` is asserted (Section 4.1), it should also be stable under unity-gain feedback. Hence the non-dominant poles are placed sufficiently above the GBW.

Furthermore, the other non-dominant poles of the complete loop, determined by the poles of the building blocks (Section 3.2.7.2), are specified above 1 MHz. This is easy to achieve in practice and makes the poles completely ignorable, given their distance from the loop GBW.

4.6 Detector and Transconductor

The detector and transconductor do not directly govern the loop dynamics when they are much faster than the loop (Eq. (3.25)). However, their combined (linearized) gains ($k_D G_m$) determine the range of I_{det} and I_{ref}. In Section 5.1, it will be shown that the equivalent linear gain of the detector k_D spans roughly from 0.1 to 0.25, such that only the transconductance G_m of the transconductor is left as a scaling constant. A convenient current range is chosen: 0 µA to 120 µA. Given the specified differential output amplitude range of the VGA in Table 4.1, a transconductance of 900 µS is chosen.

4.7 Conclusion

The key specifications for the building blocks of the event-driven AGC system, integrated in a linear optical system (Section 4.2), have been derived based on the system-level model of Chapter 3 and are summarized in Table 4.2. These specifications form the foundation for the circuit level design, presented in Chapter 5.

It should be noted that the system could be compressed to some extent by not retaining the clear separation of functionality between the building blocks. However, to enable easier testing of the proof-of-concept, it was decided to directly implement the block diagram and consider further integration in a future, revised version.

Specification	Symbol	Unit	Min.	Typ.	Max.
VGA					
Transimpedance	R_T	Ω	125		2000
Large-signal gain	k_G	Ω		125	
Intercept	j_G	(V)		0.36	
Low cut-off pole frequency	f_{pL}	kHz	31		500
Differential peak-to-peak output voltage	V_{od}^{pp}	mV	320	400	480
Common-mode output voltage	V_{oc}	V		2.1	
Detector					
Small-signal gain	k_D	·	0.1		
Transconductor					
Transconductance	G_m	µS		900	
Output current range	I_o	µA	0		120
Logamp					
Large-signal gain	k_M	V		U_t	
Intercept	j_M	A			
Loop					
Dc gain	T_0	dB	40		
Unity-gain frequency	$f_{0\text{dB},T}$	kHz		3.1	
2 % settling time	T_s	µs		205	
Non-dominant poles		MHz	1		
Loop Filter					
Dc gain	L_0	dB	65		
Unity-gain frequency	f_L	kHz		60	
Non-dominant poles		kHz	$4f_{0\text{dB}}$		
Quantizer					
Number of quantization steps	N_Q			8	
Input step	δ_C	mV		200	
Approximate equivalent gain	$\tilde{k}_Q$	·		0.725	

Table 4.2: Specifications of the key building blocks of the AGC system embedded in the linear optical receiver.

Chapter 5

Event-driven AGC Implementation

Based on the system model developed in Chapter 3 and the block-level specifications derived in Chapter 4, this chapter describes the implementation of the building blocks of the event-driven automatic gain control (AGC), embedded in a high-speed linear optical receiver, on the circuit level. The reader is referred to Fig. 4.1 for the block diagram.

In Sections 5.1 to 5.3, the detector, transconductor and logarithmic amplifier are presented, respectively. Section 5.4 introduces the loop filter, based on the analysis of the basic topology in Appendix B. The quantizer and its building blocks (comparator, monoflop, digital control) are discussed in Section 5.5, while the window comparator is presented in Section 5.6. Finally, Section 5.7 concludes the description of the system with the look-up table (LUT). Some peripheral building blocks which are not an inherent part of the AGC functionality, but are nonetheless critical for the system are shortly described in Section 5.8: biasing, current DACs, test tree and digital interface and control. To conclude, Section 5.9 presents experimental results that confirm the functionality of the proposed system.

5.1 High-speed Detector

The task of the detector is to extract the peak-to-peak amplitude of the differential output signal of the variable gain amplifier (VGA). An ideal detector is accurate, high-speed and power-efficient. Accuracy in a classic peak or valley detector is achieved with ideal diodes implemented with high gain amplifiers in a negative feedback loop [31]. At the projected data rate of 26 Gb/s, the required loop gain-bandwidth product (GBW) must be extremely high and is not viable. Several approaches have been presented to mitigate the speed limitation while still retaining reasonable accuracy [71–74]. However, none is sufficiently fast for this application. Therefore, accuracy must be traded for speed and instead of detecting the peak-to-peak amplitude, another quantity indicative of the ac energy content, the mean absolute value, is extracted [75]. Unfortunately, the detector output will be dependent on the signal shape (no free lunch).

In this work, a detector is proposed that takes advantage of the balanced differential nature of the output signal of the VGA. Figure 5.1 depicts the simplified circuit diagram.

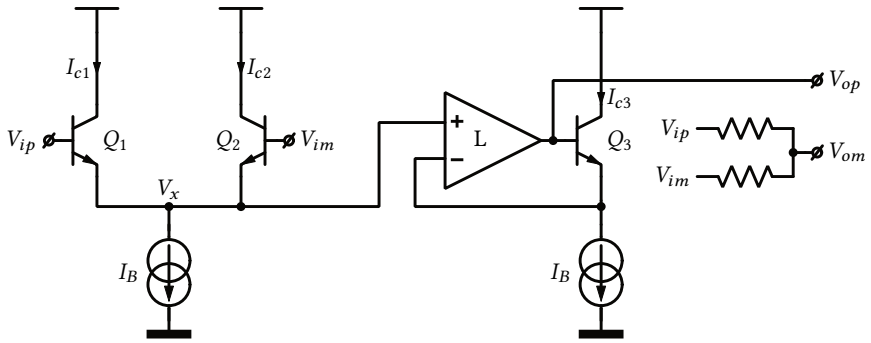


Figure 5.1: Simplified circuit diagram of the proposed high-speed detector.

The balanced differential input signal $V_{id} = V_{ip} - V_{im}$ is applied to a bipolar differential pair Q_1, Q_2 with tail current I_B. The voltage at the common node, V_x, is applied to the emitter of transistor Q_3 (also biased with I_B), which together with opamp L constitutes a negative feedback loop. The closed-loop behaves as a low-pass filter. The output V_{op} is taken at the base of Q_3 and referred to the common-mode input voltage V_{om}: $V_o = V_{op} - V_{om}$.

5.1.1 Non-linear Model

Assume infinite current gain and output resistance of the bipolar transistors (Section 1.3) and infinite loop gain at any operating point ($\beta \to \infty$, $r_o \to \infty$, $T \to \infty$). In addition, assume Q_1 and Q_2 identical and Q_3 A times bigger than Q_1 or Q_2. All transistor are assumed in the active region.

5.1.1.1 Static Behavior

Consider the system in static equilibrium. Let the differential- and common-mode decomposition of the input signals be:

$$V_{id} = V_{ip} - V_{im} \tag{5.1}$$

$$V_{ic} = \frac{1}{2}(V_{ip} + V_{im}) \tag{5.2}$$

$$= V_{om} \tag{5.3}$$

Equation (5.3) follows from the circuit diagram. Kirchoff's mesh law around Q_1 and Q_2 yields:

$$V_{ip} - V_{be1} + V_{be2} - V_{im} = 0 \tag{5.4}$$

Also, with $V_{be1}, V_{be2} \gg U_t$ and I_{s0} the reverse saturation current of the transistors:

$$V_{be1} = U_t \ln \frac{I_{c1}}{I_{s0}} \tag{5.5}$$

$$V_{be2} = U_t \ln \frac{I_{c2}}{I_{s0}} \tag{5.6}$$

$$I_B = I_{c1} + I_{c2} \tag{5.7}$$

It follows from Eqs. (5.4) to (5.6) that:

$$\frac{I_{c2}}{I_{c1}} = \exp\left(-\frac{V_{id}}{U_t}\right) \tag{5.8}$$

Eliminating I_{c2} or I_{c1} between Eq. (5.8) and Eq. (5.7) yields:

$$I_{c1} = \frac{I_B}{1 + \exp\left(-\frac{V_{id}}{U_t}\right)} \tag{5.9}$$

$$I_{c2} = \frac{I_B}{1 + \exp\left(\frac{V_{id}}{U_t}\right)} \tag{5.10}$$

The voltage at the common point of the differential pair, V_x (with Eqs. (5.1) and (5.2)),

$$V_x = V_{ip} - U_t \ln\left(\frac{I_{c1}}{I_{s0}}\right) \tag{5.11}$$

$$= V_{ic} + \frac{V_{id}}{2} - U_t \ln\left(\frac{I_B}{I_{s0}} \frac{1}{1 + \exp\left(-\frac{V_{id}}{U_t}\right)}\right) \tag{5.12}$$

is imposed on the emitter of Q_3 by the feedback loop, such that the output V_{op} becomes:

$$V_{op} = V_x + U_t \ln\left(\frac{I_B}{A\, I_{s0}}\right) \tag{5.13}$$

$$= V_{ic} + \frac{V_{id}}{2} + U_t \ln\frac{1}{A} + U_t \ln\left(1 + \exp\left(-\frac{V_{id}}{U_t}\right)\right) \tag{5.14}$$

Referring the output voltage to the common-mode input voltage and substituting Eq. (5.3):

$$V_{op} - V_{ic} = V_{op} - V_{om} = V_o \tag{5.15}$$

Or:

$$V_o = \frac{V_{id}}{2} + \underbrace{U_t \ln\frac{1}{A}}_{\text{offset term}} + \underbrace{U_t \ln\left(1 + \exp\left(-\frac{V_{id}}{U_t}\right)\right)}_{\text{error term}} \tag{5.16}$$

The output voltage is half of the differential input voltage plus an offset and error term that depend on the input voltage and the relative emitter areas. Both are a function of temperature.

The first-order truncated asymptotic expansion of Eq. (5.16) for magnitudes of the input voltage a few times the thermal voltage, results in the following approximation:

$$V_o \approx \begin{cases} \frac{V_{id}}{2} + U_t \ln\frac{1}{A} & , V_{id} \gg U_t \\ -\frac{V_{id}}{2} + U_t \ln\frac{1}{A} & , -V_{id} \gg U_t \end{cases} \tag{5.17}$$

Or:

$$V_o \approx \left|\frac{V_{id}}{2}\right| + U_t \ln\frac{1}{A} \tag{5.18}$$

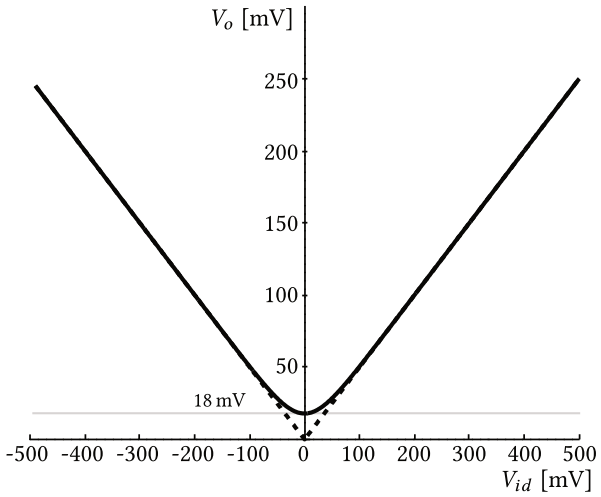


Figure 5.2: Static input-output characteristic (solid line) and approximation (dashed line) of the detector at room temperature and equal emitter areas.

Hence, the circuit of Fig. 5.1 implements a (scaled) absolute value function.

Equation (5.16) indicates that, for zero differential input, the offset term plus error term attains a maximum value of $U_t \ln(2/A)$ which, for equal emitter areas ($A = 1$), equals 18 mV at room temperature and 24 mV at 105 °C. As the input voltage increases, Fig. 5.2 shows that the error quickly vanishes, leaving only the offset term $U_t \ln(1/A)$, which is linearly dependent on temperature. This is illustrated in Fig. 5.2, for equal emitter areas at room temperature.

5.1.1.2 Dynamic Behavior

Now let the differential-mode input signal be time-dependent. Assume that the differential pair is fast enough such that the common node voltage V_x instantaneously follows the inputs, which is relatively easy to achieve in practice. In addition, express the low-pass filter action of the closed-loop as the convolution of $V_x(t)$ with the impulse response $h(t)$. Equations (5.12) to (5.15) can now be written as follows:

$$V_x(t) = V_{ic} + \frac{V_{id}(t)}{2} - U_t \ln\left(\frac{1}{1 + \exp\left(-\frac{V_{id}(t)}{U_t}\right)}\right) - U_t \ln\frac{I_B}{I_{s0}} \tag{5.19}$$

$$V_{op}(t) = h(t) * \left[V_x(t) + U_t \ln\left(\frac{I_B}{A\, I_{s0}}\right)\right] \tag{5.20}$$

This yields:

$$V_o(t) = h(t) * \left[\frac{V_{id}(t)}{2} + U_t \ln\left(1 + \exp\left(-\frac{V_{id}(t)}{U_t}\right)\right) + U_t \ln\frac{1}{A}\right] \tag{5.21}$$

$$\approx h(t) * \left(\left|\frac{V_{id}(t)}{2}\right| + U_t \ln\frac{1}{A}\right) \tag{5.22}$$

in which the approximation holds for input voltages a few times the thermal voltage, which is reasonable as the desired output voltage of the datapath is a few hundred millivolts[1] (Table 4.2). The output voltage is then equal to the sum of an offset and the low-pass filtered absolute value of *half* the differential input. This is called the *mean absolute value* and is dependent on the signal shape [76].

In this work, A is chosen equal to 1, such that the temperature-dependent offset vanishes. The detector output is then given by:

$$V_o(t) \approx h(t) * \left| \frac{V_{id}(t)}{2} \right| \tag{5.23}$$

5.1.2 Linear Model in the Amplitude Domain

Section 3.1.3 presented a linearized model of the AGC control loop, in which the $F(\cdot)$ function in Eq. (3.20) includes the linearized detector gain in the amplitude domain. As explained in the previous section, accuracy is traded for speed in the proposed detector. The detected signal and hence linearized gain are signal-dependent. This introduces some uncertainty into the system regarding signal ranges and accuracy, but as Eq. (3.20) shows, does not affect the loop dynamics. Let's specifically denote the linearized ratio of the output voltage[2] to the peak-to-peak differential input voltage as the detector gain k_D. In order to establish a feeling for the output range of the detector, k_D will be derived for different signal shapes, using Eq. (5.23), by calculating the periodic steady-state (pss) output voltage.

Square wave Let the differential input be a perfect square wave with peak-to-peak amplitude $2A_{id}$. Dividing by two and taking the absolute value results (Eq. (5.23)) in a dc voltage with magnitude $A_{id}/2$, which remains invariant under low-pass filtering. Hence:

$$k_D^{\text{sq}} = \frac{\frac{A_{id}}{2}}{2A_{id}} = \frac{1}{4} \tag{5.24}$$

Sine wave For a sine wave input with differential peak-to-peak amplitude $2A_{id}$ and period T:

$$V_{id}(t) = A_{id} \sin(2\pi \frac{t}{T}), \tag{5.25}$$

the pss detector output is calculated as:

$$V_o = \frac{1}{T} \int_T \frac{A_{id}}{2} \left| \sin(2\pi \frac{t}{T}) \right| \mathrm{d}t \tag{5.26}$$

$$= \frac{2}{T} \int_{T/2} \frac{A_{id}}{2} \sin(2\pi \frac{t}{T}) \, \mathrm{d}t \tag{5.27}$$

$$= \frac{A_{id}}{\pi} \tag{5.28}$$

[1] Although no rigorous proof is given, an intuitive argument is the following: as most of the time, $V_{id}(t)$ is greater than a few thermal voltages, the result of the convolution is not greatly affected.

[2] More precisely, the dc component of the pss output voltage, as ripple is superimposed on the output voltage.

Hence:

$$k_D^{\sin} = \frac{\frac{A_{id}}{\pi}}{2A_{id}} = \frac{1}{2\pi} \approx 0.16 \tag{5.29}$$

Note that the linearized detector gain is independent of frequency.

Figure 5.3 illustrates both the exact results and the approximate results from the linear model. It is confirmed that the incremental detector gain is a function of input amplitude and signal shape and that $k_D^{\sin}$ and k_D^{sq} are only indicative, yet practically useful measures. The output amplitude is not a linear function of the AGC setpoint value. However, at a typically desired differential peak-to-peak output amplitude of around 400 mV, the approximation is close enough.

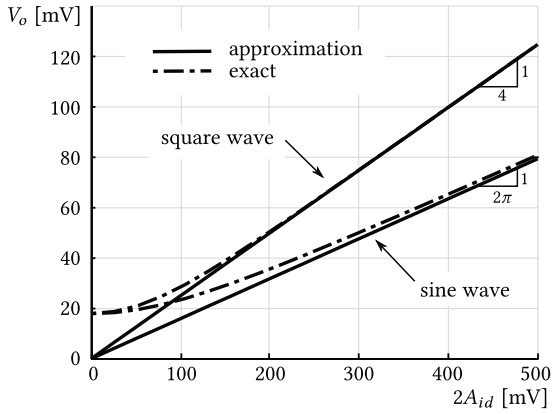


Figure 5.3: Exact (dash-dot) and approximate (solid) periodic steady state input-output characteristic of the detector for a square wave and sine wave input at room temperature and equal emitter areas.

Pseudo-random Data Sequences No formal derivation will be given for random data sequences encoded in different modulation formats. Instead, a qualitative argument will be given.

In a dc-balanced pseudo-random bit sequence (PRBS) non-return-to-zero (NRZ) stream with ideal rectangular pulses, the two signal levels are equally likely. Clearly this results in the same detector output and gain as for the periodical square wave: $k_D = 1/4$. Likewise, the four levels in a dc-balanced quaternary pulse-amplitude modulation (PAM) stream with ideal rectangular pulses are equally probable. The detector output will be 2/3 that of a NRZ stream: $k_D = 1/6$.

When the pulse shape differs from the ideal, the effective detector gain will be accordingly lower. This is hard to quantify a priori. However, it is reasonable to assume that the effective gain will lie between the values for the square wave and sine wave, multiplied by the scaling factor for the relevant modulation format.

5.1.3 Non-idealities

The influence of the detector's non-linear nature on the incremental gain has been described in the previous section. In addition, multiple other aspects affect the detector

output: the detector filter bandwidth, detector offset and response of the VGA.

5.1.3.1 Detector Filter Bandwidth and Output Ripple

In order to limit the ripple superimposed on the static detector output, the cut-off frequency of the low-pass filter f_D should be low enough in order not to significantly alter the frequency content of the signal at the filter input.

A NRZ PRBS sequence with bit period T_b and pattern length n has spectral lines at multiples of $1/(nT_b)$ [40], resulting in a minimal ripple frequency of $2/(nT_b)$ at the input of the filter through the absolute value operator. The same argument holds for a 4-PAM data stream, with 'digits' and 'bit period' replaced by 'symbols' and 'symbol period', respectively.

For a 26 Gb/s NRZ data stream, the bit period is 38.5 ps. The lowest spectral line is at 12 Hz for a PRBS $2^{31} - 1$ pattern. This is too low to implement on chip. Moreover, the detector must be must faster than the AGC loop, a requirement stated in Section 3.1.3. It was experimentally determined that a bandwidth of around 10 MHz gives satisfying results. Note that the limited bandwidth of the subsequent transconductor stage (Section 5.2) will also provide filtering.

5.1.3.2 Detector Offset

Mismatch between the bipolar devices, current sources, source followers (Section 5.1.4) and offset of the opamp in the detector results in an extra dc offset error term at the detector output. This is mitigated in the circuit by appropriate amplifier topology, device sizing and layout. Furthermore, residual offset can be calibrated out by adjusting the AGC setpoint current I_{ref}.

5.1.3.3 Influence of the VGA on the Signal Shape

The detector output voltage depends on the signal shape, which is partially determined by the datapath (VGA). First, the low-pass frequency response of the VGA has a smoothing effect on the signal pulse shape. As the bandwidth is reduced, the output eye gradually closes vertically. This results in lower detector output compared to the value derived solely from the eye amplitude.

Second, residual offset of the datapath results in an amplitude-independent detector output offset and at the same time desensitizes the mean absolute value function for differential changes. Indeed, in the limit one side of the differential pair is completely cut-off indefinitely and only the offset of the input signal remains after low-pass filtering. For this reason the total residual offset at the output of the datapath (and hence the input-referred offset of the balancing amplifier, see Sections 4.3 and 6.5) should not exceed 10 mV.

Third, noise inherent to the system input signal or added by the VGA results in a higher detector output via the rectifier action, as shown in [77] for a sine wave with gaussian noise.

In summary, the exact output of the detector is, a priori, unknown. However, it can be assumed that it will be in the range derived in this section. This has been confirmed by simulations. In practice, adjusting the desired set point of the AGC loop can compensate for these uncertainties.

5.1.4 Circuit

Figure 5.4 shows the detailed schematic of the detector with annotated bias currents and device sizing. The important specifications are: common-mode input range around 2.1 V (output common-mode voltage of datapath), differential input range at least 500 mV, small-signal bandwidth of the filter f_D 1 MHz to 12.9 MHz. Source followers M_1 and M_2 (both low-V_t devices) shift down intermediate nodes by ~320 mV in order to mitigate the output voltage range requirements of the opamp. Bulk and source are connected to avoid threshold voltage increase due to the bulk effect. For a typical V_t of 260 mV, V_{EFF} must be 60 mV, i.e. $IC_{1,2} \approx 1.5$. Voltage sources implemented as MOS diodes in the collector and drain leads protect the active devices from overvoltage. The current source devices operate at $IC \approx 20$ for low g_m/I_D and g_m to minimize drain referred thermal noise but $V_{DS,sat}$ still low enough. Drain lengths are chosen 1 µm to bring down drain current mismatch[3], for moderately high output resistance and layout convenience.

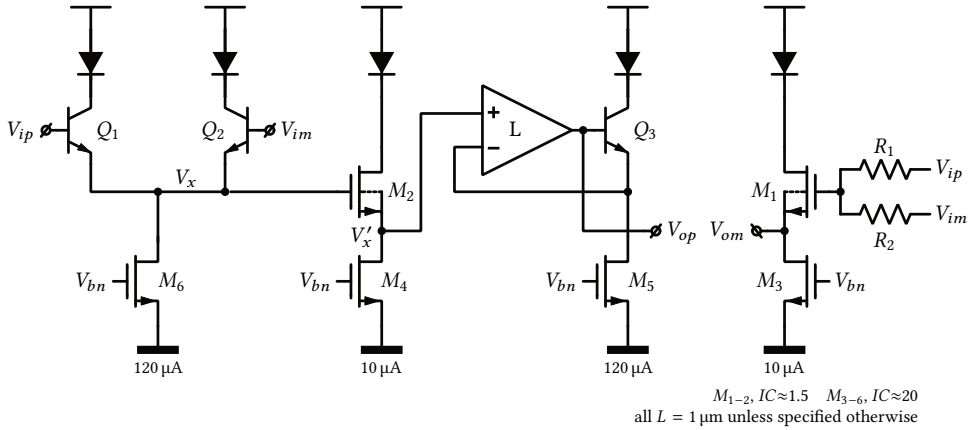


Figure 5.4: Detailed circuit diagram of the high-speed detector, including level shifters to shift the output voltage range requirement of the opamp down to a reasonable level.

In Section 5.1.1.2 it was mentioned that the differential pair should be able to follow its inputs instantaneously. Although the circuit is used in large-signal regime, resulting in widely varying operating point parameters, a small-signal argument hints at the required bias collector current. Let the instantaneous input be such that one side of the differential pair is in cut off (see Fig. 5.2). The active transistor now behaves as an emitter follower, with a small-signal bandwidth in the order of f_T of the transistor, given low enough total source impedance R_S (25 Ω plus the ohmic base resistance in this case) [31]. However, the pole formed at the emitter by the resistance and capacitance at the emitter dominates the high-frequency response:

$$f_p \approx \frac{1}{2\pi} \frac{1}{\left(\frac{1}{g_m} + \frac{R_S}{\beta}\right) C_L} \tag{5.30}$$

[3] In general, drain current mismatch is minimized in weak inversion as the gate area increase dominates over the increase in g_m. However, for a fixed gate area, drain current mismatch is minimized in strong inversion, where g_m is low.

C_L is essentially the sum of the gate-drain plus bootstrapped gate-source capacitance of M_2 and the gate-drain plus gate-bulk capacitance of M_6, ~20 fF. A bias current of 120 µA produces sufficient g_m to place the pole above 30 GHz. To minimize the loading on the datapath, Q_1 and Q_2 have minimal emitter area.

Summing resistors R_1 and R_2 directly load the output of the datapath and should be much higher than 25 Ω. A value of 5 kΩ is selected. They constitute a parasitic low-pass filter with the capacitances to ground, lightly affecting the settling time of the detector output V_{om}.

5.1.4.1 Loop Design

The specified closed-loop bandwidth f_D is determined by the feedback loop consisting of opamp L and emitter follower Q_3. The transfer function from v_x to v_{op} can be written as the general feedback theorem (GFT) decomposition (Appendix A):

$$H = \frac{v_{op}}{v_x} = H_\infty \frac{T}{1+T} + \frac{H_0}{1+T} \tag{5.31}$$

$$\approx \frac{T}{1+T} = D \tag{5.32}$$

with T the loop gain. Equation (5.32) holds as the gain of the followers (M_2 and Q_3) is nearly unity ($H_\infty \approx 1$) and sensible design will make direct transmission H_0 negligible. As shown in Fig. 5.5, the loop GBW $f_{0\text{dB}}$ approximates f_D when peaking is negligible. H is essentially equal to the discrepancy factor D. A reasonable static error of less than 1 % calls for at least 40 dB dc loop gain over the entire input and output range. A single-stage, dc optimized opamp is sufficient. Stability is ensured by parallel compensation at the output node v_{op} [78]. Time-domain overshoot is avoided by ensuring H_0 negligible and placing the non-dominant poles of the loop gain at least 4 times higher than the GBW for negligible peaking.

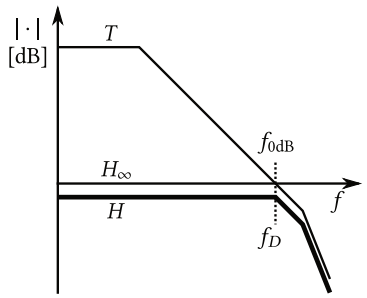


Figure 5.5: GFT decomposition of the detector loop voltage transfer function.

In order to mitigate input-referred offset, a symmetrical opamp topology is chosen (Fig. 5.6). The low-side input common-mode range demands a PMOS input differential pair M_{1-2}, operating in moderate inversion $IC \approx 1$ for high g_m and g_m/I_D for high gain and low input-referred offset, with $L = 0.8$ µm for smaller area. A bias current of 20 µA each and $C_c \approx 9$ pF gives $f_{0\text{dB}} \approx g_{m1,2}/(2\pi C_c) = 360\,\mu\text{S}/(2\pi\, 9\,\text{pF}) = 6.4$ MHz that meets the specification. The rail devices operate at the onset of strong inversion (M_{9-12}) to strong inversion (M_{3-6}) and long L for reduced current mismatch and limited area. The high resulting $V_{DS,sat}$, particularly for the NMOS transistors, poses no problem as the

output range of the amplifier is centered around mid-supply. Cascodes M_{7-8}, operating in moderate inversion at $IC \approx 4$, match the output conductance of the PMOS to the NMOS transistors (increasing gain), while reducing systematic offset. As their contribution to input-referred mismatch is low, area is saved by using smaller channel lengths. The biasing circuits are not shown.

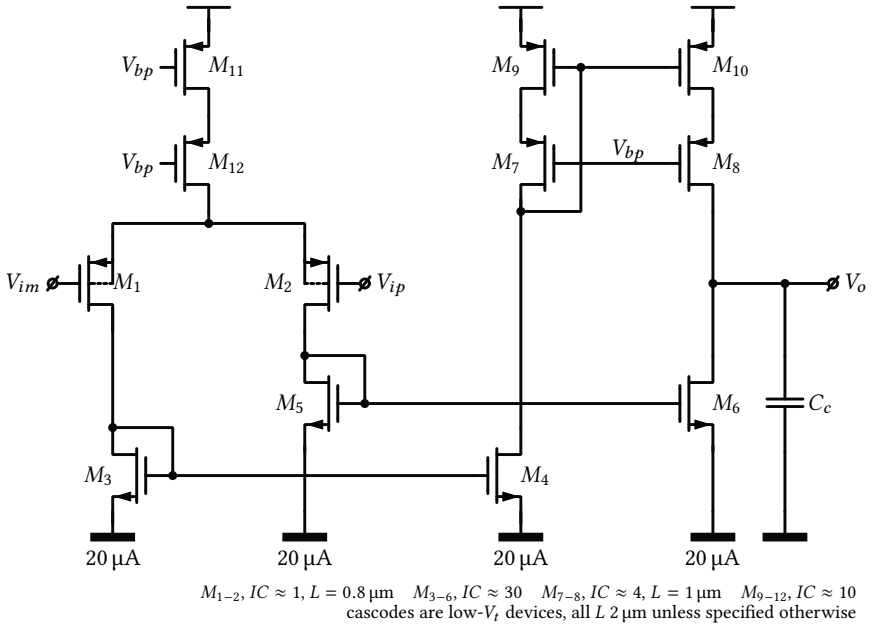


Figure 5.6: Circuit diagram of the symmetrical opamp in the high-speed detector. Bias voltage generation is omitted.

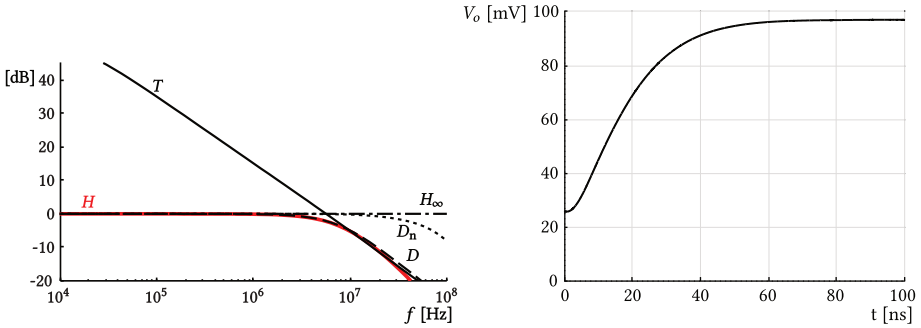
5.1.4.2 Simulation Results

In Fig. 5.7(a), the simulated GFT decomposition of the transfer function H (Eq. (5.31)) of the internal filter is shown. The discrepancy factor D starts at essentially 0 dB and ultimately rolls off with a tiny amount of peaking, indicating high dc loop gain and sufficient stability margin. The null discrepancy factor D_n is negligible for the frequencies of interest, indicating no direct forward transmission ($H_0 \approx 0$).

The theoretical model is confirmed in the simulated step response of the entire detector (Fig. 5.7(b)), for slightly filtered PRBS NRZ input data with peak-to-peak amplitude of 400 mV. This results in a output voltage slightly less than 100 mV. The simulated linearized gains (Table 5.1) for both a NRZ and a 4-PAM stream approach the theoretical values of 1/4 and 1/6 for larger inputs, thereby verifying the linear approximation.

5.2 Transconductor

The transconductor converts the floating output voltage of the detector to multiple identical currents, that are applied to the logamp (Section 5.3) and window comparator (Section 5.6). The circuit (Fig. 5.8) is a differential version of the classical opamp-plus-transistor



(a) GFT decomposition of the voltage transfer function H (internal filter).

(b) Step response for NRZ PRBS data ($2A_{id}$ = 400 mV) (entire detector).

Figure 5.7: Simulation results for detector.

$2A_{id}$ [mV]	50	100	150	200	250	300	350	400	450	500
PRBS	0.55	0.32	0.27	0.25	0.24	0.24	0.24	0.24	0.24	0.24
4-PAM	0.53	0.29	0.22	0.20	0.18	0.18	0.18	0.18	0.18	0.18

Table 5.1: Simulated detector gain for PRBS and 4-PAM as a function of differential peak-to-peak input amplitude.

current source circuit [29] and is essentially a linearized differential pair with current mirror as differential to single-ended converter. Output voltage $V_L \approx 1.2\,\text{V}$ is imposed by the subsequent circuits. The opamps and the devices M_{1-2} constitute feedback loops that impose the input voltage across the resistors R. Assuming linear operation and ideal devices and infinite loop gain, the output current can be written as:

$$I_o = \frac{V_{ip} - V_{im}}{R} \tag{5.33}$$

$$= G_m(V_{ip} - V_{im}) \tag{5.34}$$

with $G_m = 1/R$ the effective transconductance, specified at 900 µS or $R = 1.1\,\text{k}\Omega$ (Table 4.2). Tail current source I_B fixes the output current range at 0 µA to 120 µA (Section 4.6). Remark that the output current is always nonzero, typically 15 µA or higher, as the output voltage of the detector is always greater than 0 V.

5.2.1 Circuit

To create the required multiple outputs (Fig. 4.1), parallel cascaded PMOS and NMOS mirrors $M_{3-8,a,b,c}$ are used, as shown in the detailed circuit diagram (Fig. 5.9). A simple PMOS mirror load cannot be used as the imposed output voltage V_L would not allow M_{1-2} to operate in the saturation region. The operating point of the devices changes widely as the current through the input transistors and main current mirrors changes substantially over the output current range. Hence M_{1-2} are chosen to operate in moderate inversion at nominal bias current for high g_m and high loop gain. The opamps are implemented

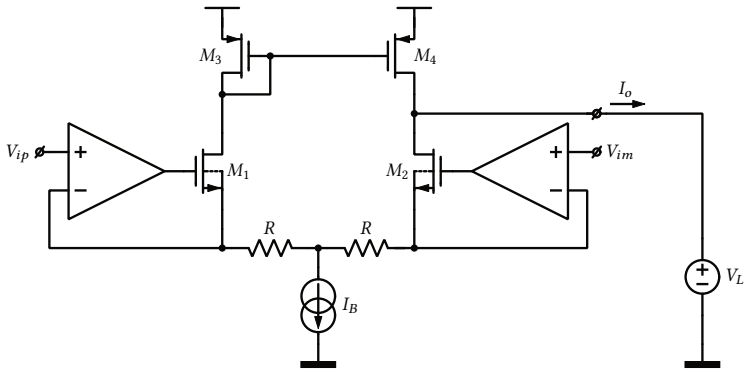


Figure 5.8: Working principle of transconductor.

using bipolar differential pairs Q_{1-2}, for high loop gain and less systematic static error, with PMOS current mirror load M_{9-10}. Base current can be ignored as the current gain is very high (Section 1.3). The current mirror rail devices all operate at the onset of strong inversion and long L for reduced current mismatch. The tail current sources are cascoded and sized for low drain current mismatch. The design of the feedback loops is similar to Section 5.1.4.1 and will not be repeated.

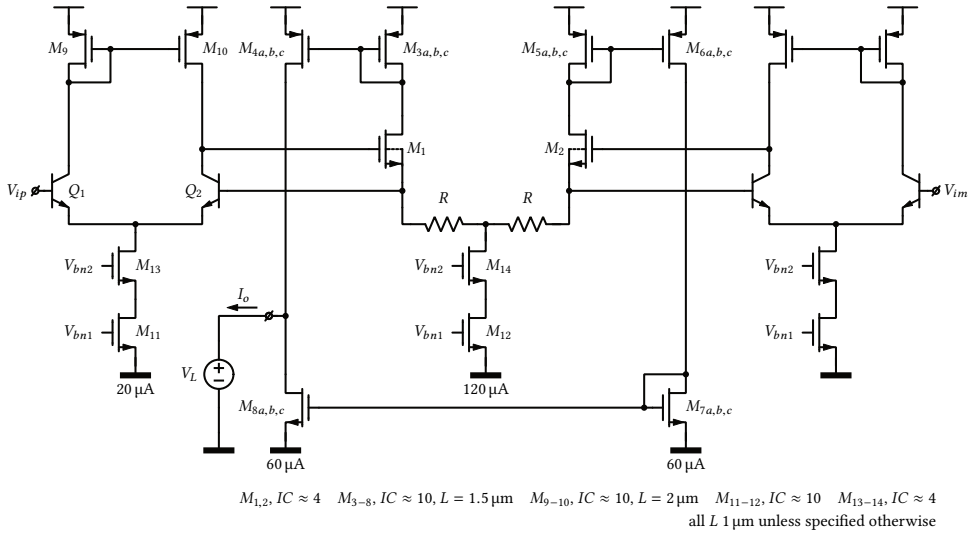


Figure 5.9: Circuit diagram of the transconductor.

Clearly, G_m tracks process and temperature variations of R and affects the accuracy of the AGC loop. Therefore the most stable resistor type (high-resistive poly) is chosen. Any residual offset can be tweaked out (within limits) by adjusting the AGC loop setpoint. Furthermore, offset between the multiple outputs is limited to 1 % at the nominal output current. This has been confirmed with simulations.

5.3 Logarithmic Amplifier

Including a logarithmic amplifier in a continuous-time AGC loop in which the VGA gain depends exponentially on its control signal, makes the settling time independent of the input amplitude (Section 3.1.1). In the quantized version of this system, the settling time is upper bounded.

The commonly used logamp implementation types can be partitioned roughly depending on the required bandwidth. High speed, high dynamic range-amplifiers usually implement some form of piecewise linear approximation and are relatively complex. Low-speed logamps use the inherent logarithmic properties of silicon junction devices and are easier to implement. The latter option is chosen for this application considering only a few MHz of bandwidth is required (Chapter 4) [76, 79, 80].

The purpose of this building block is to realize the logarithmic function of Eq. (3.19), repeated here for convenience:

$$F(x) = k_M \ln\left(\frac{x}{j_M}\right) \tag{5.35}$$

in which k_M is the large-signal gain and j_M the intercept. Two logamps are needed (Section 3.1), one for the reference input current and one for the detected output voltage, converted to a current. After subtraction in the loop filter (Fig. 4.1), j_M vanishes, as is also implied by Eq. (3.20). It follows that the actual value of j_M is of no importance, as long as both logamps are matched.

5.3.1 Working Principle

Consider the ideal exponential diode. When a current $I_d \gg I_{s0}$ is forced in the anode, the V-I characteristic is:

$$V_o = U_t \ln\left(\frac{I_d}{I_{s0}}\right) \tag{5.36}$$

Where I_{s0} and U_t are the reverse bias saturation current and thermal voltage, respectively. Real diodes deviate from this ideal model by non-idealities such as ohmic series resistances and a change in slope. Comparing Eq. (5.35) with Eq. (5.36) shows that the ideal diode implements $F(x)$, with $k_M = U_t$ and $j_M = I_{s0}$.

Usually, the diode is a bipolar transistor, connected in the feedback loop of an operational amplifier as shown in Fig. 5.10. Assume a MOS input stage to eliminate errors due to non-zero input bias current. The negative feedback loop establishes voltage V_{ref} at the input node. At the same time it lowers the input and output impedance, creating a decent current sink and voltage source. The amplifier must be able to sink the entire input current range. When the base is connected to the collector, the transistor is said to be diode-connected. Otherwise, the device is denoted as a transdiode. This configuration allows more freedom in setting the dc output voltage.

The emitter current of a bipolar transistor, as a function of the base-emitter and base-collector voltage V_{be} and V_{bc}, is described by the Ebers-Moll equation [31]:

$$I_e = -\frac{I_{s0}}{\alpha_F}\left(\exp\frac{V_{be}}{U_t} - 1\right) + I_{s0}\left(\exp\frac{V_{bc}}{U_t} - 1\right) \tag{5.37}$$

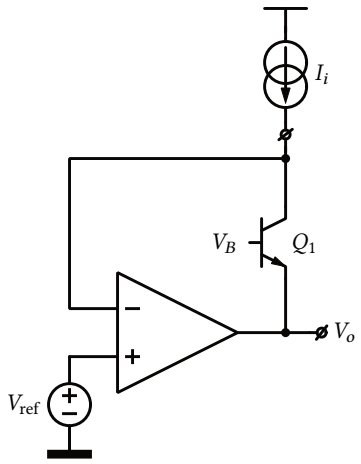


Figure 5.10: Classic transdiode-based logamp.

in which α_F is the grounded-base current gain. In the given process technology and under usual bias conditions, $\beta \gg 1$ such that $\alpha_F \approx 1$ and $I_e \approx I_c$ (Section 1.3):

$$I_c = -I_{s0}\left(\exp\frac{V_{be}}{U_t} - 1\right) + I_{s0}\left(\exp\frac{V_{bc}}{U_t} - 1\right) \tag{5.38}$$

Solving Eq. (5.38) for V_{be} yields:

$$V_{be} = U_t \ln\left(\frac{I_c}{I_{s0}} + \exp\frac{V_{bc}}{U_t} - 1\right) \tag{5.39}$$

$$V_{be} \approx U_t \ln\left(\frac{I_c}{I_{s0}}\right) \tag{5.40}$$

For $V_{bc} \neq 0$, additional currents will contribute an error to V_{be}. V_{bc} is zero when $V_{\text{ref}} = V_B$ or in the diode-connected configuration. If no equal base-collector voltage can be maintained, making $V_{bc} < 0$ or $V_{\text{ref}} > V_B$ introduces a negligible error. Equation (5.40) shows that the bipolar transistor approximates the ideal diode equation Eq. (5.36).

Temperature-dependence

The output voltage of the logamp (Eq. (5.40)) depends on temperature in two ways. First, the saturation current I_{s0} of a bipolar transistor is quite sensitive to temperature, approximately doubling for each 10 °C increase. Second, the thermal voltage U_t increases linearly with temperature, about 0.33 %/°C at room temperature. As, ideally, both the detected signal (converted to a current) and the AGC reference signal (a current) is subject to the same function (Eq. (5.35)), this temperature dependence has no influence on the static behavior of the AGC system as long as the dc loop gain remains high (see Eq. (3.21)). However, the dynamic behavior and settling time changes with temperature because the GBW of the AGC loop, and hence the closed-loop 3-dB bandwidth, depends on $k_M = U_t$ (Eq. (3.24)).

Various ways exist to compensate for the temperature dependence of the thermal voltage [76]. Most involve the weighted addition of a voltage or current that is complementary

dependent on temperature, e.g. via a suitable resistor, to the logamp output. A more modern approach is digital compensation. In this application, those solutions were deemed too complex or unpredictable and no effort was spent to implement temperature compensation. The resulting change in dynamic behavior is acceptible as the AGC specifications are chosen quite conservative.

5.3.2 Circuit

Figure 5.11 shows the implementation of the logamp. For simplicity, the amplifier is implemented as a common-emitter stage Q_2 with a NMOS follower M_4. Device Q_1 converts the input current to a voltage and acts as a cascode of Q_2. The static output voltage is given by:

$$V_o = V_B - V_{be1}(I_i) = V_B - U_t \ln\left(\frac{I_i}{I_{s0}}\right) \tag{5.41}$$

Reference voltage V_{ref} (Fig. 5.10) is now implicit and depends on the input current. It is equal to the sum of the drive voltages of transistors Q_2 and M_4:

$$V_{\text{ref}}(I_i) = V_{be2}(I_i) + V_{GS4} \tag{5.42}$$

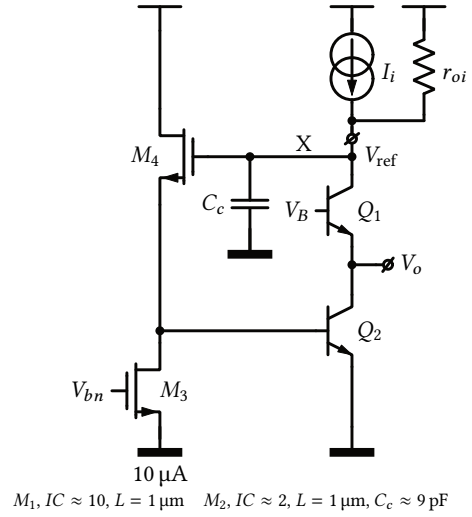


Figure 5.11: Implementation of the logamp.

In Section 5.2, it was pointed out that I_i is always higher than 15 µA, essentially due to the detector output that saturates at a positive voltage at low datapath output amplitudes, hence V_{be2} is always higher than 600 mV and current source M_3 never enters the triode region. This configuration can easily sink the maximum input current of 120 µA without large changes in V_{ref} (~60 mV/decade).

M_4 operates in moderate inversion for high g_m, partially undoing the reduced voltage gain of the follower due to non-zero g_{mbs} caused by non-zero V_{SB}. The bulk effect increases V_{TH2} and hence V_{GS2}. This guarantees $V_{\text{ref}} > V_B$ at all times, preventing saturation of Q_1 and large output voltage errors. An alternative is to operate M_4 in strong

inversion for sufficient gate-source voltage and shorting bulk and source, resulting in similar gain. This, however, is inconvenient for layout as a local well is required, precluding compact collective interdigited layout of both logamps. The current source transistor M_3 is sized for low current mismatch.

It is easily shown that the small-signal input-output transfer function $H = v_o/i_i$ can be written as the following GFT decomposition, by test signal injection at the error signal of the major loop (location X). Remark that this is a non-ideal injection point:

$$H = \frac{v_o}{i_i} = H_\infty \frac{T}{1+T} + \frac{H_0}{1+T} \tag{5.43}$$

$$H_\infty = -\frac{1}{g_{m1}(I_i)} \tag{5.44}$$

$$T \approx \frac{g_{m4}}{g_{m4}+g_{mbs4}} \frac{g_{m2}(I_i) r_{oi}}{1+sr_{oi}C_c} \approx \frac{g_{m2}(I_i)}{sC_c} \tag{5.45}$$

$$H_0 \approx 0 \tag{5.46}$$

in which the output resistance of the devices has been ignored (makes H_0 zero), as well as all capacitances and associated poles and zeros except for C_c which is the compensation capacitance creating the dominant pole in the loop gain T. Device r_{oi} is the driving point resistance at the input node and is dominated by the output resistance of the input source. The explicit dependence on I_i in g_{m2} (and g_{m1}) is retained to stress that the loop GBW and hence closed-loop bandwidth depends on the input current. With the given device sizes, the bandwidth is between 3.5 MHz and 85 MHz across the input current range, at room temperature, which is much higher than the AGC GBW, as required. Care has been taken to place the non-dominant poles sufficiently above the maximum GBW.

5.3.3 Simulation Results

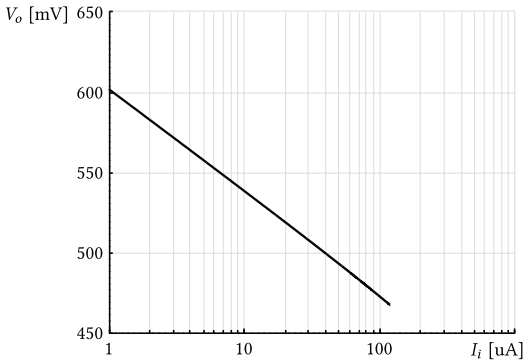


Figure 5.12: Simulated input-output characteristic of a logamp at room temperature.

The devices in both logamps are laid out in an interdigited fashion (MOS devices) or quad structure (bipolar devices). This ensures optimal process and thermal matching. The simulated output characteristic (Fig. 5.12) shows excellent logarithmic behavior. This is confirmed by a typical *log conformity*, defined as the peak deviation from the best fit straight line of the output voltage versus the log of the input current [81], of 0.33 %. Both

logamps are matched to within a few millivolt across process, supply and temperature (PVT) variations and local mismatch.

5.4 Loop Filter

The loop filter compensates the AGC loop and partly determines the settling time of the closed-loop system (Chapter 3). Its linearized model should approximate an integrator with unity-gain frequency f_L:

$$L \approx \frac{L_0}{1 + \frac{s}{2\pi f_L}} \tag{5.47}$$

The specified dc gain, unity-gain frequency and common-mode input range and output range are (Section 4.5):

$$L_0 = 1840 = 65\,\text{dB} \tag{5.48}$$

$$f_L = 60\,\text{kHz} \tag{5.49}$$

$$V_{ic} = 100\,\text{mV to } 1000\,\text{mV} \tag{5.50}$$

$$V_o = 200\,\text{mV to } 1800\,\text{mV} \tag{5.51}$$

in which the input range follows from the logamp implementation (Section 5.3) and the output range from the quantizer input specification (Section 4.4).

The unity-gain frequency of a typical opamp topology with parallel or Miller compensation can be written as $f_{0\text{dB}} = g_m/(2\pi C_c)$, with g_m the transconductance of the input stage and C_c the dominant capacitance. For $C_c = 10\,\text{pF}$, which is already quite big for an integrated capacitor, Eq. (5.49) results in $g_m = 4.3\,\mu\text{S}$. This low transconductance value, although easily achievable, has a negative impact on input-referred offset and noise and requires very high resistance levels (even in a two stage amplifier) to obtain the required gain. As an off-chip capacitance is not desired, a technique known as Miller compensation with capacitance multipliers is used. A detailed linear analysis of such a system is given in Appendix B. The result is repeated in Fig. 5.13 and constitutes the basis of the implementation discussed next.

5.4.1 Circuit

The loop filter is implemented as a two-stage amplifier, based on [82] (Fig. 5.14). The input stage, a differential pair with folded-cascode load, is followed by a common-source output stage. Both sides of the folded cascode are biased with a dedicated current mirror with current gain k. Compensation capacitor C_c is connected between the output node and a branch of the folded cascode. The effective compensation capacitance can be shown to be multiplied by both the gain of the second stage (Miller effect) and the current gain k. Hence a low unity-gain frequency can be obtained without sacrificing silicon area or input stage transconductance.

The linear model, derived in Appendix B, consist of one zero, one pole and one complex pole pair (Fig. 5.13):

$$L = \frac{v_o}{v_{id}} = L_0 \frac{\left(1 + \frac{s}{\omega_z}\right)}{\left(1 + \frac{s}{\omega_1}\right)\left(1 + \frac{s}{\omega_n Q} + \frac{s^2}{\omega_n^2}\right)} \tag{5.52}$$

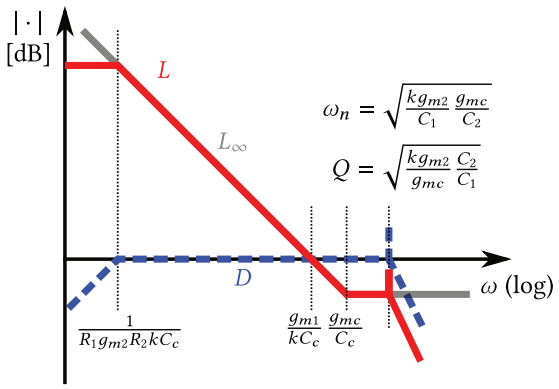


Figure 5.13: Linear model of the loop filter, a two-stage opamp with multiplied Miller capacitance (derived in Appendix B).

The dc gain, zero, real and complex poles and Q-factor are given by:

$$L_0 = g_{m1} R_1 g_{m2} R_2 \tag{5.53}$$

$$\omega_z = \frac{g_{mc}}{C_c} \tag{5.54}$$

$$\omega_1 = \frac{1}{R_1 g_{m2} R_2 k C_c} \tag{5.55}$$

$$\omega_n = \sqrt{\frac{kg_{m2}}{C_1}\frac{g_{mc}}{C_2}} \tag{5.56}$$

$$Q = \sqrt{\frac{kg_{m2}}{g_{mc}}\frac{C_2}{C_1}} \tag{5.57}$$

with g_{m1}, R_1, C_1 and g_{m2}, R_2, C_2 the effective transconductance, output resistance and output capacitance of the first and second stage, respectively. The input conductance of the current amplifier is denoted g_{mc}. The complex pole pair ω_n is created by the internal feedback loop through C_c and the current amplifier. In order to guarantee sufficient stability of the internal loop, $Q < 1$ is imposed for its associated Q-factor.

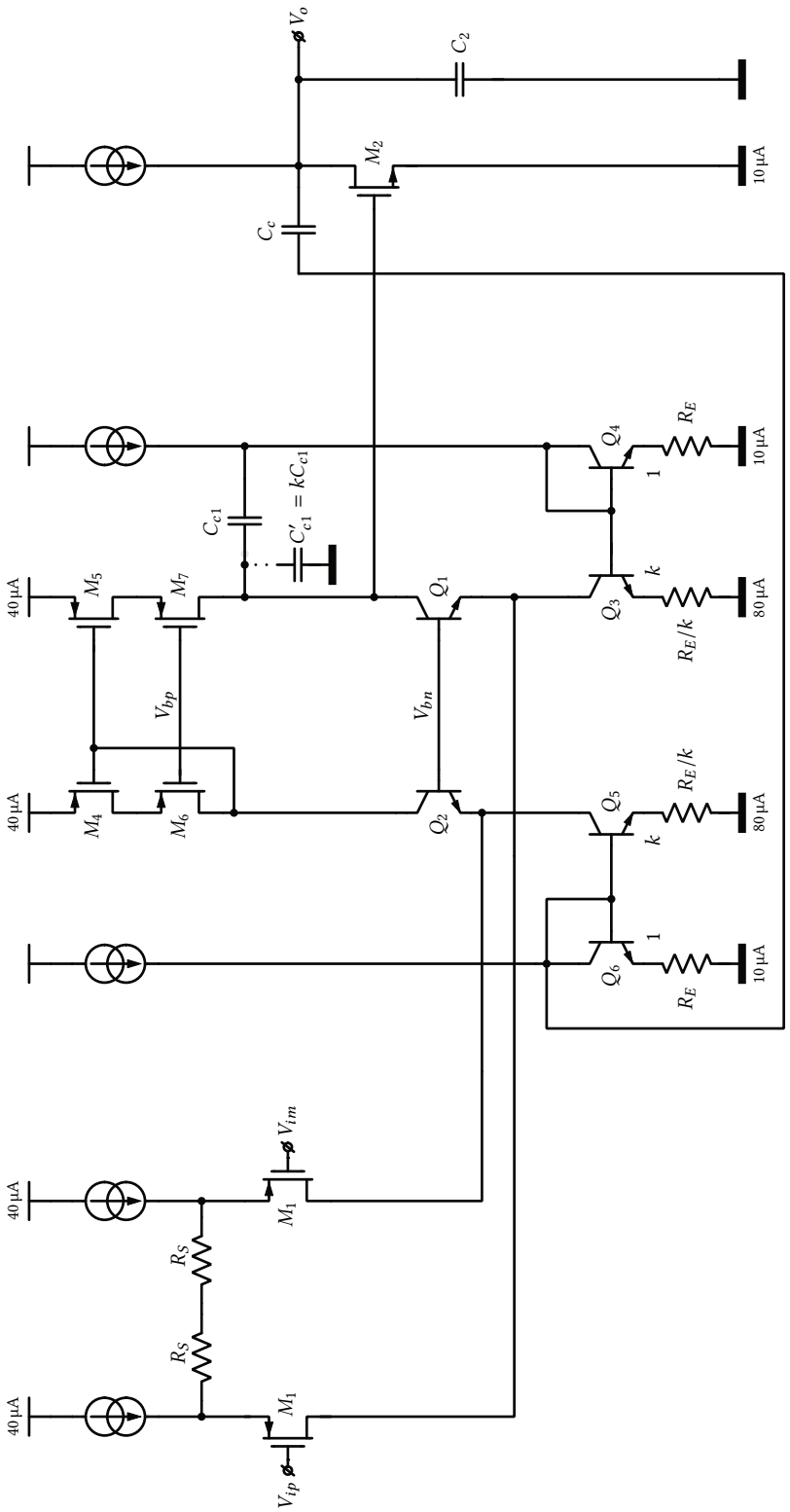
The unity-gain frequency is:

$$\omega_L = \frac{g_{m1}}{kC_c} \tag{5.58}$$

and is specified in Eq. (5.49).

5.4.2 Design Considerations

Given the low-side input range, a PMOS differential pair $M_{1,2}$ is chosen as input transconductor, with N-type folded cascode load and PMOS mirror for reduced systematic input-referred offset. Both sides of the folded cascode are independently biased with a current mirror Q_{3-4} and Q_{6-5} with gain k. The displacement current through C_c is sensed in the mirrored branch Q_6, multiplied with k in Q_5 and injected back into the output node of the first stage via current follower Q_2 and the PMOS mirror M_{4-5}.



M_1, $IC \approx 60$, $L = 2\,\mu\text{m}$ M_2, $IC \approx 20$, $L = 5\,\mu\text{m}$ $M_{4\text{-}5}$, $IC \approx 40$, $L = 2\,\mu\text{m}$ $M_{6\text{-}7}$, $IC \approx 3$, $L = 0.3\,\mu\text{m}$
$R_S = 3\,\text{k}\Omega$, $k = 8$, $R_E = 10\,\text{k}\Omega$, $C_c = 30.6\,\text{pF}$, $C_1 = 1\,\text{pF}$
current sources sized for optimally low input-referred offset: SI, large L
$Q_{4\text{-}6}$ sized for optimally low input-referred offset: big emitter area and degeneration (small g_m)
$Q_{1\text{-}2}$ sized for optimally high g_m (reduced kirk effect)

Figure 5.14: Circuit diagram of the loop filter.

The output resistance of the input stage R_1 can be written as:

$$R_1 \approx \left(1 + g_{m7}\left(r_{o7} \,||\, r_{o5}\right)\right)\left(r_{o5} + r_{o7}\right) \,||\, \beta_{Q_1} r_{oQ_1} \tag{5.59}$$

$$\approx g_{m7} r_{o7} r_{o5} \,||\, \beta_{Q_1} r_{oQ_1} \tag{5.60}$$

which is high due to boosting by local feedback. The output resistance of the output stage, R_2, is the parallel combination of the output resistances of the bias current source and M_3. Load capacitance C_2, estimated at 100 fF, dominates the output node capacitance.

The internal loop is compensated by intentionally adding capacitance to the output node of the first stage, as explained in Appendix B.3.2. In the circuit diagram, ignore C_{c1} and assume C'_{c1} connected. Then the capacitance at the internal output node, C_1, is:

$$C_1 \approx C'_{c1} \tag{5.61}$$

The main parasitic poles are located at the emitter of Q_{1-2} and the gate of mirror M_{4-5}:

$$\omega_{nd1} \approx \frac{g_{m4-5}}{C_{gs4} + C_{gs5}} \tag{5.62}$$

$$\omega_{nd2} \approx \frac{g_{mQ_{1-2}}}{C_{gd1-2} + C_{be1-2} + C_{cbQ_{3-5}} + C_{csQ_{3-5}}} \tag{5.63}$$

The mirror introduces a zero as the effective transconductance of the loaded differential pair halves at high frequencies, in addition to a feedforward zero:

$$\omega_{znd1} \approx \frac{2g_{m4-5}}{C_{gs4} + C_{gs5}} \tag{5.64}$$

$$\omega_{znd2} \approx \frac{g_{m4-5}}{C_{gd4}} \tag{5.65}$$

Moreover a parasitic pole and zero are created in the compensation path by the bipolar current mirrors of the cascode load.

The multiplication factor is chosen $k = 8$ as this allows for easy common-centroid layout. An effective input transconductance g_{m1} of 100 µS is aimed for. Equations (5.49) and (5.58) yield $C_c \approx 30$ pF which, although big, is still reasonable.

The folded cascode bias current was chosen 40 µA to place the first parasitic pole of the PMOS mirror (Eq. (5.62)) high enough. The bias current of the differential pair devices was chosen identical.

Contrary to usual dc optimized sizing, $M_{1,2}$ operate in very strong inversion, $IC \approx 60$. Combined with the resistive source degeneration, this results in an effective input transconductance $g_{m1} \approx 100$ µS (Eq. (5.58)). Degradation of the input-referred offset is countered by long $L = 2$ µm. PMOS mirror devices M_{4-5} operate in very strong inversion and long gate length, $IC \approx 40$, $L = 2$ µm, to reduce drain current mismatch without sacrificing area.

In order to place the second parasitic pole (Eq. (5.63)) at high enough frequency, the cascodes Q_{1-2} are implemented with bipolar transistors (for high g_m and low capacitances). In addition, degenerated bipolar transistors are used for the N-type current mirrors. This allows for sufficiently high input conductance of the current amplifier g_{mc} ($\approx g_{mQ_6}/(1 + g_{mQ_6} R_E)$), placing ω_n high enough, while still keeping the parasitic mirror pole and zero negligible due to lower parasitic capacitance compared to MOS transistors (remember $k = 8$). Stability of the internal loop (Eqs. (5.56) and (5.57)) and the closed-loop

zero (Eq. (5.54)) are directly affected by g_{mc}. The degeneration lowers the mirror current mismatch and hence total input-referred offset of the amplifier.

The output stage device M_3 operates at 10 µA and $IC \approx 20$ for reduced transconductance to lower the Q-factor of the internal loop (Eq. (5.57)). The long gate length $L = 5\,\mu m$ makes up for lost gain, in particular at lower output voltages.

Internal Loop Compensation

C'_{c1} was intentionally added to compensate the internal loop as it reduces its unity-gain frequency (Appendix B.3.2). A value of 8 pF was selected. For symmetry reasons the right branch of the folded cascode is also biased with a current amplifier with gain k, that can also be exploited as a capacitance multiplier. By replacing C'_{c1} by $C_{c1} = C'_{c1}/k = 1\,pF$, as indicated in the circuit diagram, further silicon area is conserved. It can be shown that this additional feedback loop introduces an extra left hand-side plane (LHP) zero and pole in the discrepancy factor D and hence in L. Although no further analytical treatment is given, this will be observed in the Spectre simulation results.

5.4.3 Transient Restart Behavior

As discussed in Section 4.1, two measures are taken to prevent saturation of the loop filter's output voltage when the loop is disabled (`inh` asserted). First, when `inh` is logic high, the loop filter is automatically reconfigured such that it follows the midpoint voltage V_{inh} of the current transition bin of the quantizer. This is implemented using analog multiplexers and some digital control. Second, the maximum output voltage of the amplifier is limited to the full-scale input voltage of the quantizer $V_{FS} = 1.8\,V$, by including a precision clamp. This clamp is an additional feedback circuit that is automatically engaged when the output voltage comes near V_{FS}. It reuses part of the folded cascode to implement an analog minimum selector. These circuits will not be discussed further.

5.4.4 Simulation Results

This section presents numerical Spectre simulation results of the nested GFT decomposition of the *loop gain of the loop filter in follower configuration* (Section 4.1), denoted the external loop gain. This loop gain is (approximately) the open-loop transfer function of the opamp $L = v_o/v_{id}$ and will be decomposed as a function of its internal loop, as was done analytically in Appendix B. The follower configuration presents the worst-case scenario for the non-dominant opamp poles as the external loop gain is not subject to a net attenuation as in the case of the AGC loop (Eq. (3.23)). The dc input voltage was 1 V.

The results for the decomposition of the external loop gain are shown in three situations: no compensation of the internal loop, straightforward compensation of the internal loop and compensation with capacitance multiplier. Finally the transient step response of the follower (the external closed-loop gain) is shown.

5.4.4.1 Decomposition of the External Loop Gain L

No Internal Compensation To validate the model of Fig. B.22, the simulated GFT decomposition of L, when C_{c1} or C'_{c1} are omitted, is shown in Fig. 5.15. L_∞ and the null discrepancy factor D_n are not shown for clarity. Clearly, the loop gain T and the discrepancy factor D fit the model up to 1 GHz. At that point parasitic poles and zeros, which are

not included in the model, determine the frequency response. The peaking in D caused by the placement of the second normal pole of T which is too low, is a sign of insufficient stability margin of the internal loop.

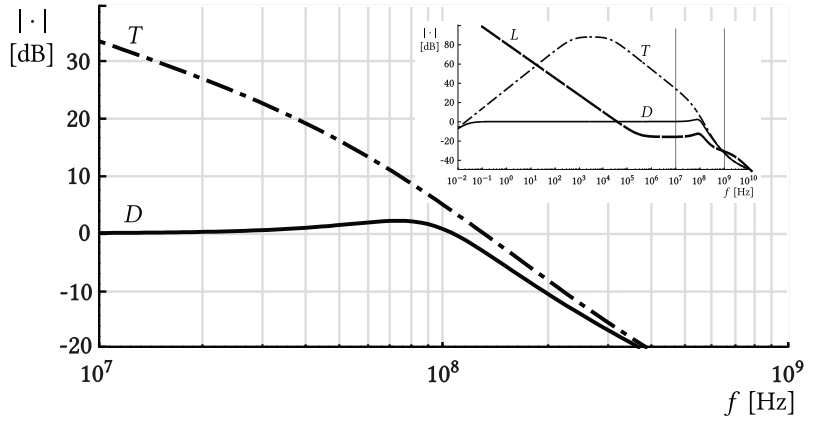


Figure 5.15: No internal compensation.

Internal Compensation In this and the next paragraph, L will also be omitted for clarity. Figure 5.16 shows the same Bode diagrams when C'_{c1} is reinstated to lower the first normal pole of T. The peaking in D has greatly diminished as the unity-gain frequency has moved down. At the same time the bandwidth of D and hence of the closed-loop transfer function L has dropped (Eq. (5.58), note the scale difference of the horizontal axis).

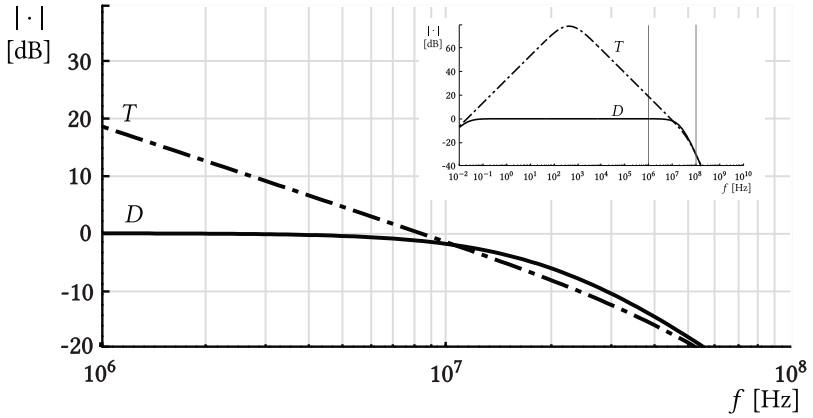


Figure 5.16: Internal compensation without using inherent capacitance multiplier.

Internal Compensation with Capacitance Multiplier Replacing C'_{c1} by C_{c1} to conserve silicon area, inserts a LHP zero and pole in T and has a beneficial effect on the stability of the internal loop. It behaves as a lead compensator (Fig. 5.17).

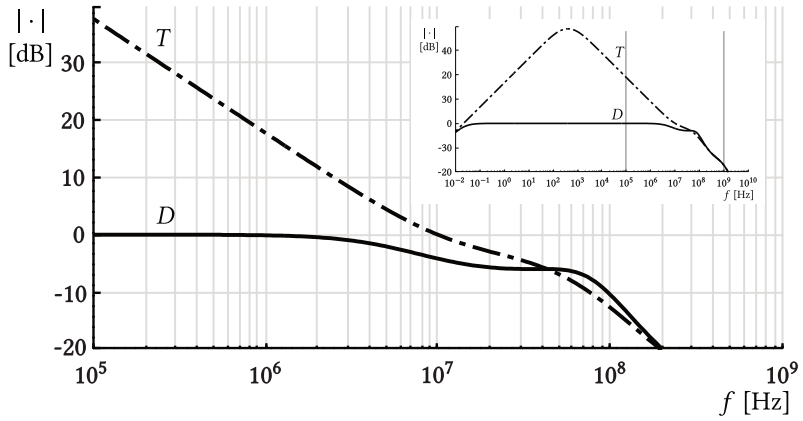


Figure 5.17: Compensation of the internal loop using the capacitance current multiplier.

Complete Decomposition The complete GFT decomposition of the external loop gain L (Fig. 5.18) shows that the very-high-frequency response is mainly shaped by the parasitics in L_∞ and D as the null discrepancy factor D_n is nearly flat, indicating negligible direct forward transmission. Post-layout Monte Carlo simulations with auto stop, including PVT variations and local mismatch, show that the typical unity-gain frequency is 55 kHz, close to the specification of 60 kHz.

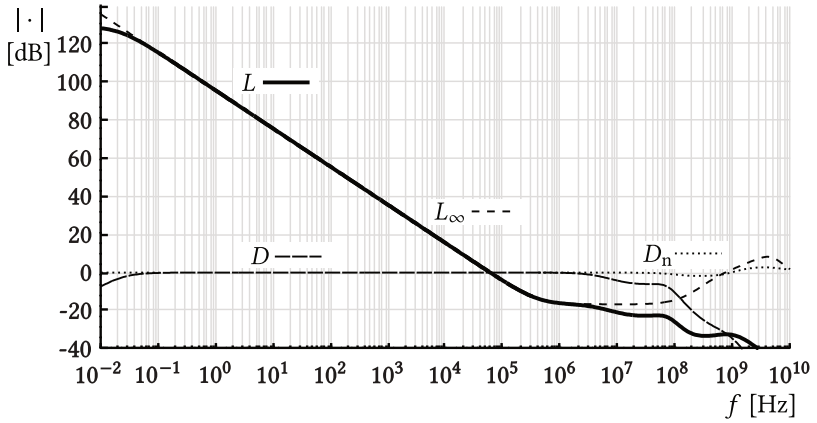


Figure 5.18: Simulated nested GFT decomposition of the external loop gain L of the loop filter in follower configuration.

5.4.4.2 Step Response of the External Closed-Loop Gain

The large-signal step response of the loop filter in follower configuration (Fig. 5.19), when V_i steps from 50 mV to 550 mV, is smooth without any overshoot and only negligible amount of preshoot. No significant slewing is apparent. Further simulations show that the output range of the loop filter (in follower configuration) extends down to 30 mV, the output voltage where the dc loop gain has dropped below 40 dB as the output device M_3 operates in deep triode.

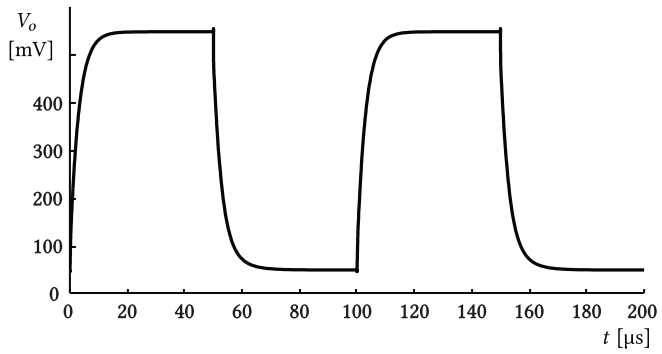


Figure 5.19: Simulated large-signal step response (V_i steps from 50 mV to 550 mV) of the loop filter in follower configuration.

5.5 Quantizer

Section 3.2.2 explained that a clockless event-driven architecture for the loop quantization is used to reduce digital-to-analog interference and lower the power consumption. A straightforward clockless flash analog to digital converter (ADC) topology is an obvious implementation candidate for the quantizer. However, this requires $N_Q - 1$ comparators for N_Q quantization levels, increasing power consumption and silicon area and hence makes scaling cumbersome.

Instead, a system based on asynchronous continuous-time delta modulation which does not suffer from these drawbacks, is implemented (Fig. 5.20) [68, 83, 84]. It consists of two comparators, a monostable multivibrator and a digital counter that retains the internal state $\mathbf{d}_{\mathrm{DAC}}$ and outputs the binary-encoded digital code word $\mathbf{d}$. The reference levels for the comparators are dynamically generated by a digital to analog converter (DAC) which is controlled by a feedback loop governed by the current state.

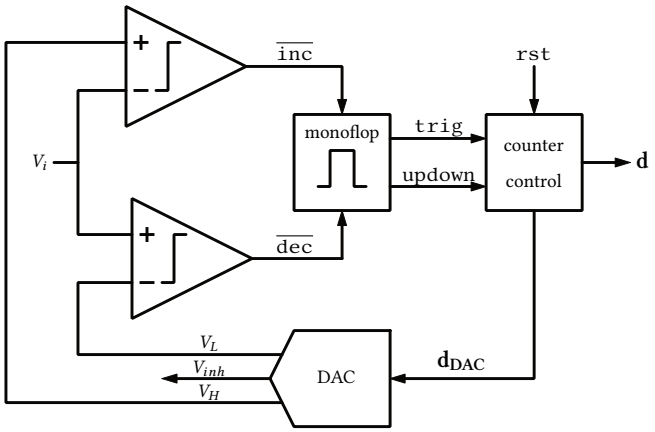


Figure 5.20: Block diagram of the quantizer, implemented as an asynchronous delta modulator.

The operation (Fig. 5.21) is as follows: assume the quantizer to be in a state $d_{DAC,k}$.

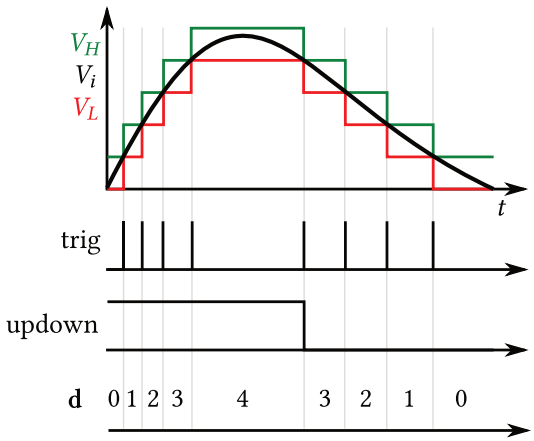


Figure 5.21: Working principle of the quantizer: asynchronous delta modulation.

This means that the input signal V_i is in the range $[V_L, V_H]$, associated with the k-th quantization interval and that the comparator outputs $\overline{\texttt{inc}}$ and $\overline{\texttt{dec}}$ are deasserted. When V_i increases and crosses V_H, $\overline{\texttt{inc}}$ asserts and the monoflop triggers: updown, indicating the direction of change, is set logic high and is applied to the counter, followed by a well-defined `trigger` pulse. The counter updates its internal state to $d_{DAC,k+1}$. The output **d** is incremented (with saturation) and the DAC updates its output voltage, such that V_i is now in the range $[V_L, V_H]$, associated with the $(k+1)$-th quantization interval. For decreasing inputs, updown is deasserted and the counter is decremented. Clearly, this is a stateful system. A pulse on `rst` resets the logic to a known state at power-on.

The propagation delay t_p of the quantizer is the time required for one conversion, i.e. the duration for a change of one of the comparator outputs to propagate around the loop such that the reference input voltages are settled again. The total propagation delay constitutes the sum of the propagation time of the comparators, monoflop, digital logic and settling time of the DAC. For correct operation, the input signal should not cross any quantization level during t_p. This is formalized as the *tracking condition*:

$$\max \left| \frac{\mathrm{d}V_i}{\mathrm{d}t} \right| < \frac{\delta_C}{t_p} \tag{5.66}$$

The input quantization step δ_C (Section 3.2.2.1) is specified as 200 mV (Section 4.4). Equation (5.66) expresses a speed limitation on V_i, which is the output of the loop filter. Assume now that only the loop filter (with unity-gain frequency f_L) determines the loop dynamics and that a rail-to-rail input step is applied at its input. The maximum rate of change at the loop filter output is given by $2\pi f_L V_{DD}$. Plugging the numbers of Tables 4.1 and 4.2 into Eq. (5.66) yields:

$$t_p < 200 \times 10^{-3} \frac{2.5}{2\pi\, 60 \times 10^3} \tag{5.67}$$

$$< 1.3\,\mu\mathrm{s} \tag{5.68}$$

This is an extremely conservative calculation as the bandwidth of the AGC loop itself is much smaller and a full-scale step input never occurs. As a t_p of 1.3 µs is easily achieved in practice in the given process technology, this is of no further concern.

Table 4.2 specified $N_Q = 8$ and $\delta_C = 200\,\text{mV}$. In order to mitigate the output range requirements of the loop filter (sufficient loop gain at low output voltages/codes), the reference voltage of the quantizer is chosen as 200 mV, which gives a full-scale voltage of 1.8 V. The quantization transition points lie at $400, 600, \ldots, 1600\,\text{mV}$. A strict requirement is monotonicity of the quantizer. This requires that the differential non-linearity (DNL) be smaller than one quantization step. Here we specify the integral non-linearity (INL) and DNL more stringent, at $0.1\delta_C$. This allows the quantizer to be easily scaled up to 40 quantization steps while still keeping INL $< 0.5\delta_C$ and DNL $< 0.5\delta_C$, without redesigning the comparators.

Sections 5.5.1 to 5.5.4 discuss the DAC, comparators, monoflop and digital logic in more detail, before simulation results are presented in Section 5.5.5.

5.5.1 DAC and Full-scale Voltage Generator

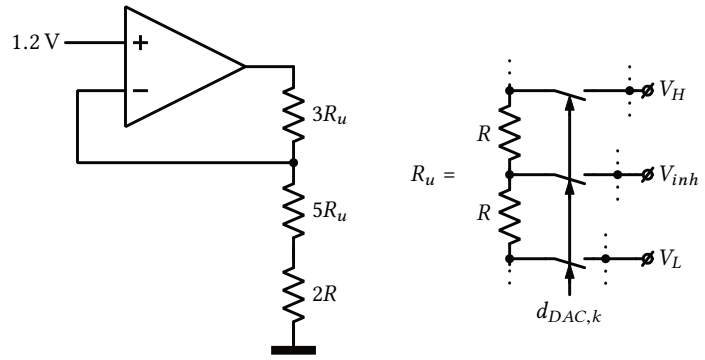


Figure 5.22: Resistor-string 3-bit DAC in the quantizer. The reference and full-scale voltage are 200 mV and 1.8 V, respectively.

The DAC in the quantizer is implemented as a simple resistive-string divider (Fig. 5.22) [85]. Each unit resistor R_u is made up of two resistors in series, with $2R = 1875\,\Omega$. Analog switches select the comparator reference voltages V_H and V_L associated with the one-hot input word $d_{DAC,k}$ (Section 5.5.4). A third output voltage is V_{inh}, the midpoint voltage between V_H and V_L. It is applied as input to the loop filter, reconfigured as a follower when the AGC loop is disabled (Section 4.1).

The resistor string is embedded in the feedback loop of a non-inverting amplifier configuration in order to create the full-scale voltage of 1.8 V when the auxiliary supply of 1.2 V is applied as input. The amplifier must source $1.8\,\text{V}/(9 \cdot 1875\,\Omega) \approx 110\,\mu\text{A}$ and provide sufficient loop gain. Although a high loop GBW would extend good power-supply rejection to a higher frequency, it would also increase the closed-loop bandwidth and reject less ripple of the input voltage. A GBW of a few MHz, combined with proper local filtering and decoupling of the auxiliary supply was chosen to avoid excessive interference from other (digital) blocks. The opamp (not shown) is implemented as a two stage Miller-compensated amplifier with bipolar input differential pair for higher input transconductance (compared to a MOS stage with the same bias current, even in weak inversion (WI)) and PMOS mirror load. Input bias current introduces a negligible error. The output stage is a PMOS transistor in common-source configuration with the resistor string as load.

As the resistor string is ratiometric, global process variations and global temperature dependence effects are negligible in this application. However, local mismatch increases the non-linearity of the divider. It can be shown that the INL is the limiting factor in determining the required matching of the resistors. Specifically, the worst case INL is given by [85]:

$$\max|\mathrm{INL}| = \frac{1}{2}\delta_C \cdot N_Q \cdot (\%\,\mathrm{matching}) \tag{5.69}$$

Under the given specification, the resistors need at least 2.5 % relative matching, which is easily achieved by increasing the area. As also other factors influence the non-linearity of the DAC and the quantizer, *dcmatch* (initially) and Monte Carlo (final check) simulations were performed to ensure sufficient performance.

5.5.2 Comparator

The comparator is the decision block in the quantizer and outputs a logic one when its differential input voltage $V_{id} > 0$ and a logic zero otherwise. It is essentially a very high gain limiting amplifier. Although a typical opamp structure could be used, a dedicated building block is designed that has hysteresis to reject noise and detector ripple on the input signal.

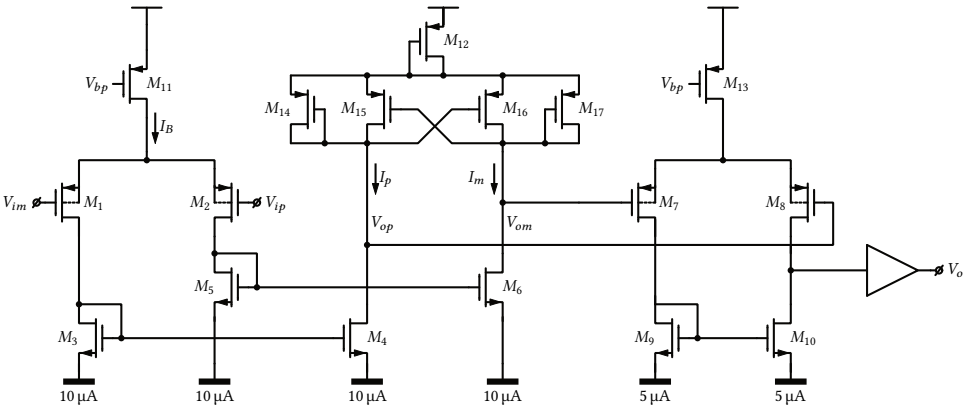


Figure 5.23: Clockless comparator in the quantizer: preamplifier, latch, analog buffer and CMOS buffer.

The comparator (Fig. 5.23) comprises four building blocks [85]. A preamplifier converts the differential input voltage to a differential current, which is applied to a positive feedback analog latch. The output voltage of the latch is differential, which is converted to a single-ended voltage by a buffer. Finally, a logic buffer generates a rail-to-rail logic output voltage. The preamplifier is a gain stage used to increase sensitivity (minimum voltage that can be discriminated) and reduce kickback to the input when the latch changes state. It also reduces the total input-referred offset as the latch has high intrinsic random offset (as positive feedback, in contrast with negative feedback, tends to increase sensitivity to device parameters).

5.5.2.1 Preamplifier

A PMOS input differential pair is chosen as the common-mode input range extends almost to the lower supply rail. Current mirrors apply the output currents to the latch. It follows from the schematic that the output currents are given by:

$$I_p = \frac{I_B}{2} + g_m \frac{v_{id}}{2} \tag{5.70}$$

$$I_m = \frac{I_B}{2} - g_m \frac{v_{id}}{2} \tag{5.71}$$

$$= I_B - I_p \tag{5.72}$$

with $v_{id} = v_{ip} - v_{im}$ the small-signal differential input voltage.

Input transistors M_{1-2} operate at the center of moderate inversion and moderate channel length, $IC \approx 1$, $L = 1\,\mu\text{m}$, for high g_m and high g_m/I_D. This results in optimum preamp gain and low input-referred offset, but increased input capacitance as the gate area is large. The non-input devices M_{3-6} operate at the onset of strong inversion, $IC \approx 10$, $L = 1\,\mu\text{m}$ as a trade-off between drain-referred current mismatch and gate area.

5.5.2.2 Latch

The latch is composed of transistors M_{14-17}. For analysis, consider the NMOS version (Fig. 5.24). Cross-coupled devices M_b are shunted by diode-connected transistors M_a. This is a strongly non-linear block. However, a linearized model exposes the basic positive feedback mechanism and will be used in the large-signal explanation of the latch operation[4].

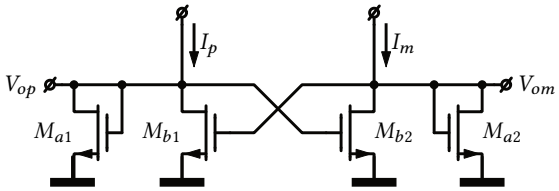


Figure 5.24: Schematic of the latch only, NMOS version.

The small-signal differential and common-mode half circuits of the latch are shown in Fig. 5.25[5]. Consider the transient operating point of the latch halfway during a state transition. Both outputs are approximately equal and all devices operate in the saturation region. Impedance $Z \approx (1/g_{ma}) \,||\, 1/(sC_o)$ lumps together the output resistance and output capacitance. The latter determines the speed of the latch, but this will not be discussed in detail. In the following sections, C_o will be ignored, such that:

$$Z \approx \frac{1}{g_{ma}} \tag{5.73}$$

[4] A different approach to explaining the large-signal operation is given in [85]

[5] The following definitions are used [18, 86]: at a port of a differential circuit, the differential voltage and current are defined as $V_d = V_p - V_m$ and $I_d = 0.5(I_p - I_m)$. The common-mode voltage and current are given by $V_c = 0.5(V_p + V_c)$ and $I_c = I_p + I_m$. It follows that $V_p = V_c + 0.5V_d$ and $I_p = 0.5I_c + I_d$. Spectre's *ideal_balun* component implements these equations.

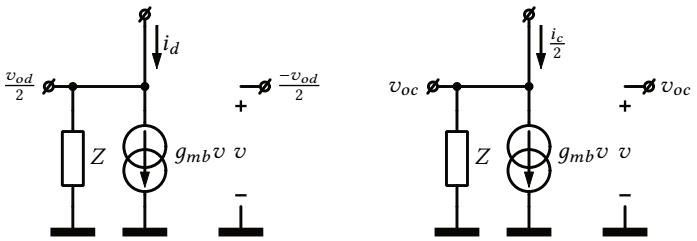


Figure 5.25: Small-signal differential and common-mode half circuits of the latch in the comparator.

The response v_{od} due to excitation with a differential current i_d can be calculated using the 1-GFT decomposition at the gate, an ideal injection point (Appendix A):

$$H_\infty = 0 \tag{5.74}$$

$$H_0 = 2Z \approx \frac{2}{g_{ma}} \tag{5.75}$$

$$T = -g_{mb}Z \approx -\frac{g_{mb}}{g_{ma}} \tag{5.76}$$

Under the condition

$$g_{mb} \geq g_{ma} \tag{5.77}$$

the negative loop gain indicates net positive feedback for differential signals. Equations (5.74) to (5.76) yield:

$$H = \frac{v_{od}}{i_d} = \frac{2}{g_{ma}} \frac{1}{1 - \frac{g_{mb}}{g_{ma}}} = \frac{2}{g_{ma} - g_{mb}} \tag{5.78}$$

H, the differential input resistance, is a negative resistance when $g_{mb} > g_{ma}$. In that case, the system is no longer memoryless which is exposed as hysteresis (see later).

For completeness, the common-mode response is calculated in a similar way and is given by:

$$H_\infty = 0 \tag{5.79}$$

$$H_0 = \frac{Z}{2} \approx \frac{1}{2g_{ma}} \tag{5.80}$$

$$T = g_{mb}Z \approx \frac{g_{mb}}{g_{ma}} \tag{5.81}$$

$$H = \frac{v_{oc}}{i_c} = \frac{1}{2g_{ma}} \frac{1}{1 + \frac{g_{mb}}{g_{ma}}} = \frac{1}{2} \frac{1}{g_{ma} + g_{mb}} \tag{5.82}$$

Equation (5.81) shows that for common-mode signals the feedback is always negative and the devices are effectively equivalent to one diode-connected transistor.

The small-signal model is approximately valid over the large-signal input current range—for incremental disturbances—as long as the instantaneous operating point parameters (effective g_m, etc.) are substituted. Note that the output range of the latch is given by the V_{gs} range of M_a, which is $[0, V_{gsa}(I_B)]$, with I_B the tail current of the preamplifier. The transient large-signal operation can be explained as follows:

- Assume $I_p \gg I_m$. All current flows in M_{a1}, such that $V_{op} = V_{gsa1}(I_B)$ and $V_{om} = 0$. The other devices $M_{b1,b2,a2}$ are in cut-off. The instantaneous differential loop gain is zero (Eq. (5.76)).
- As I_p decreases and I_m increases, the drain current of M_{a1} starts to drop while and the drain current of M_{b2} rises. The drain current of M_{b1} and M_{a2} remains zero and the loop gain starts to increase as g_{mb2} increases and the instantaneous impedance $Z \approx 1/g_{ma2}$ is very high. The feedback is net negative. The output voltages do not change substantially.
- As I_m increases, M_{b2} is gradually pulled out of the linear region into saturation, dramatically increasing its effective transconductance. The switching point occurs as soon as the (magnitude of the) instantaneous loop gain equals exactly unity and net positive feedback sets in. V_{om} increases exponentially as M_{a2} steals current from M_{b2}, while V_{op} decreases exponentially. The loop gain further builds up, almost instantaneously, to its maximum value (Eq. (5.76)). Then, as M_{b2} enters the linear region, the loop gain quickly drops again.
- When $I_m \gg I_p$, all current flows in M_{a2}, such that $V_{om} = V_{gsa2}(I_B)$ and $V_{op} = 0$. The other devices are in cut-off and the loop gain is zero once again.

The switching points of the comparator $\pm V_{id,sw}$ can be determined from the condition that loop gain must equal unity. In strong inversion[6], the transconductance can be approximated as:

$$g_m(I_D) \approx \frac{I_D}{nU_t\sqrt{IC}} \tag{5.83}$$

$$= \frac{\sqrt{I_D I_o \frac{W}{L}}}{nU_t} \tag{5.84}$$

Equating Eq. (5.76) to unity and solving yields:

$$-\frac{g_{mb}(I_{p,sw})}{g_{ma}(I_{m,sw})} = 1 \tag{5.85}$$

With $I_{p,sw}$ and $I_{m,sw}$ the currents at the switching point. Substituting Eq. (5.84) gives:

$$-\frac{I_{p,sw}\left(\frac{W}{L}\right)_a}{I_{m,sw}\left(\frac{W}{L}\right)_b} = 1 \tag{5.86}$$

Using Eqs. (5.70) to (5.72), the switching points can be computed as:

$$V_{id,sw} = \pm\frac{I_B}{g_m}\frac{1-\alpha}{1+\alpha} \tag{5.87}$$

with:

$$\alpha = \frac{\left(\frac{W}{L}\right)_b}{\left(\frac{W}{L}\right)_a} \geq 1 \tag{5.88}$$

[6]This is an approximation as the transistors operate from weak to strong inversion, given the variable drain current and fixed gate size.

in which the inequality follows from Eq. (5.77). Equations (5.87) and (5.88) shows that the hysteresis depends on the operating region of the input pair *of the preamplifier* and the ratio of the gate shape factors of the transistors *in the latch*.

Now consider Fig. 5.23 again. Diode-connected device M_{12} acts as a voltage source and brings the output range of the latch down, relaxing the input range requirements of the buffer.

The latch transistors operate at $IC \approx 2$ for high g_m. Following Spectre simulations, a hysteresis window of 10 mV was deemed sufficient to reject noise and ripple. The hysteresis window was made digitally programmable between 10 mV and 20 mV to err on the safe side.

5.5.2.3 Analog and CMOS Buffer

The analog buffer converts the differential latch output voltage to a single-ended output. It can be regarded as a mini-comparator, that only has to discriminate a rather large input voltage. The internal output node of the analog buffer is the major source of systematic offset when it is not biased at the threshold of the digital CMOS buffer. Hence, M_{9-10} are sized to develop a proper V_{GS} ($IC \approx 7$) and use long lengths ($L = 2\,\mu m$) to reduce systematic mirror mismatch. Input devices M_{7-8} operate in higher inversion to reduce area while still retaining decent g_m.

The CMOS buffer creates a full rail-to-rail swing for the subsequent logic.

5.5.3 Monoflop

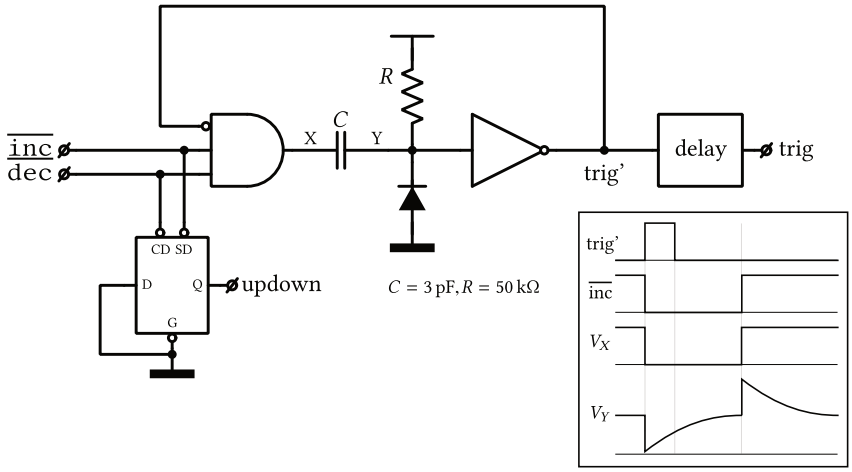


Figure 5.26: Non-retriggerable monoflop to generate `trigger` and updown signal in the quantizer. Level shifting is not shown.

The purpose of the monoflop, one-shot or monostable multivibrator is to generate a well-defined updown signal followed by a `trigger` pulse in response to a change of the output state of one of the comparators. Both signals represent an event upon which the counter and control logic acts.

The implemented monoflop (Fig. 5.26) is of the non-retriggerable type, which is however not mandatory in this context. A monoflop is retriggerable if any change of its inputs

during the trigger output pulse initiates a new timing cycle, thus extending the length of the output pulse. In a non-retriggerable monoflop the input event is simply ignored. In this application hysteresis in the comparators and the limited bandwidth of the loop filter prevent fast consecutive switching of the monoflop inputs.

The monoflop consists of an AND gate, a passive differentiator with pull-up and an inverter. The operation is as follows: initially, assume `trig`' low and $\overline{\texttt{inc}}$, $\overline{\texttt{dec}}$ high. The voltage at intermediate nodes X and Y equals the positive supply rail. When a negative edge occurs at either of the inputs, say $\overline{\texttt{inc}}$, the output of the AND gate toggles to a logic low. The step reponse of the differentiator is superimposed on V_Y and the inverter toggles, pulling `trig`' high. As soon as V_Y crosses the inverter threshold, `trig`' is pulled low. Some caution is required, because when $\overline{\texttt{inc}}$ deasserts, the positive going derived step response is imposed on V_Y and could reach twice the supply voltage. Therefore a mix of high and low voltage logic (not shown) is used to avoid damage. The diode prevents V_Y from undershooting.

The length of the trigger pulse should be at least as long as the total of the worst-case setup and hold time of the input gate of the subsequent logic, which is less than 1 ns in the given technology. It is easy to show that, with ideal logic, the pulse width is given by:

$$T_p = RC \ln \frac{1}{2} \tag{5.89}$$

Clearly, T_p depends highly on device tolerances, but this poses no problems in this application. With the given component values and over PVT variations, Monte Carlo analysis shows that T_p is centered around 100 ns with the extreme values almost 50 % higher or lower.

The D latch is connected such that updown asserts or deasserts on a $\overline{\texttt{inc}}$ pulse or $\overline{\texttt{dec}}$ pulse, respectively, indicating the direction of change.

The delay at the trigger output guarantees `trig` to occur later than any event in updown (including the hold time of the subsequent logic).

5.5.4 Counter and Digital Control

The digital logic and counter convert the delta-modulation pulses generated by the comparators and monoflop to a digital code word that drives the internal DAC and produces a binary encoded word that controls the look-up table (Section 5.7).

The block is described in the SystemVerilog hardware description language (HDL) [87] and is automatically and efficiently synthesized and placed-and-routed with a digital implementation flow [88, 89].

```
‘define NUM_BITS 3
‘define WIDTH (1<<‘NUM_BITS)

module QuantizerControl (trig, updown, rst, datao, dDAC);
input trig, updown, rst;
output reg [‘NUM_BITS-1:0] datao;
output reg [‘WIDTH-1:0] dDAC;
reg [‘NUM_BITS-1:0] cnt;

always @(posedge trig or posedge rst)
    if (rst)
        cnt = 0;
    else begin
        $display("triggered, updown: %b", updown);
```

```
        if (updown) begin // increment with saturation
            if (cnt != `WIDTH-1) cnt = cnt + 1;
        end else begin // decrement with saturation
            if (cnt != 0) cnt = cnt - 1;
        end
    end

always @(cnt) begin
    // convert binary to one hot
    dDAC <= 0;
    dDAC[cnt] <= 1'b1;

    datao <= cnt;
end

endmodule
```

When a positive edge on `trig` occurs, the internal counter is incremented or decremented with saturation, depending on updown. At the same time the counter value is one-hot encoded to generate the control signal $\mathbf{d}_{\mathrm{DAC}}$ for the DAC.

5.5.5 Simulation Results

Simulation results of the quantizer are shown in Fig. 5.27. An input up and down ramp between 0 V and 1.8 V is applied to the input of the quantizer. At each transition step the appropriate logic signals are generated and the comparator reference values are adapted. The inset details a transition point. The propagation time is 25 ns.

Monte Carlo simulations with auto stop, including PVT variations and local mismatch parameters, show that the 3σ INL and DNL is lower than 0.1 δ_C.

5.6 Window Comparator

The window comparator asserts `inh` as soon as the detected output of the VGA (converted to a current I_{det}) is within the allowed range. This disables the AGC loop and avoids limit cycling (Sections 3.2.5 and 4.1). The window comparator operates in the current domain. A block diagram is depicted in Fig. 5.28. Given a window centered around the AGC setpoint, expressed as current I_{ref} and $2\alpha I_{\mathrm{ref}}$ wide, the upper and lower window bounds I_H and I_L are derived as:

$$I_H = I_{\mathrm{ref}}\,(1+\alpha) \tag{5.90}$$

$$I_L = I_{\mathrm{ref}}\,(1-\alpha) \tag{5.91}$$

In this application α is typically 20 % (Section 4.4). Both I_H and I_L are subtracted from I_{det}. The resulting currents are compared to zero by two comparators which are wired such that the output `inh` is logic high when I_{det} is within the range $[I_H, I_L]$. The internal comparators exhibit (programmable) hysteresis to reject noise and detector ripple. The window bounds are generated with current mirrors and are also digitally programmable.

5.6.1 Current Comparator

The current comparator must support bidirectional input current and exhibit sharp transition edges with well-defined hysteresis. Figure 5.29 shows the circuit diagram, based

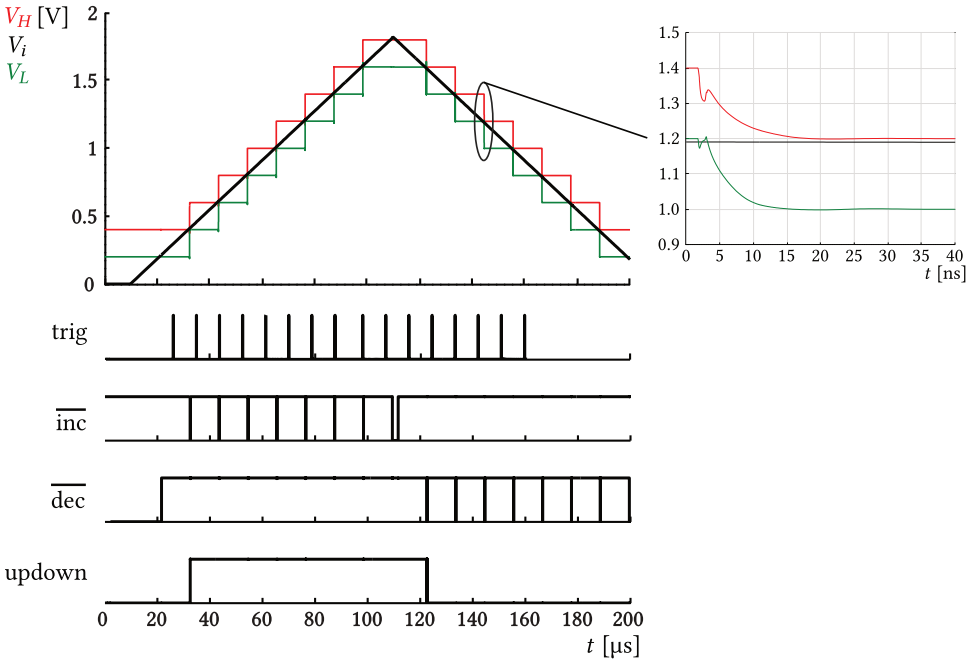


Figure 5.27: Simulation results of the quantizer.

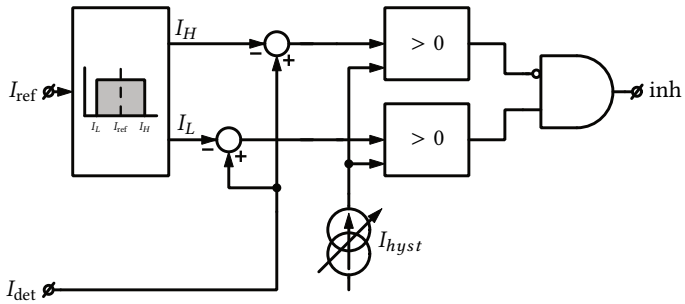


Figure 5.28: Block diagram of the window comparator, working in the current domain.

on [90]. It outputs a logic high when the input current $I_i > 0$ and a logic zero otherwise. It is composed of two stages. The input stage provides a virtual ground at a mid-rail dc voltage and a low input impedance. It applies a copy of the input current to output stage, which uses positive feedback to quickly decide on the logic output voltage. The worst case input range equals twice the output range of the transconductor, −120 µA to 120 µA. The hysteresis is programmable between 5 µA to 20 µA in steps of 2.5 µA.

5.6.1.1 Input Stage

Transistors M_{1-10} constitute the input stage, which is basically a push-pull current conveyor topology [91]. The bidirectional input current is transferred to the decision stage at extremely different impedance levels. It establishes a virtual ground for the preced-

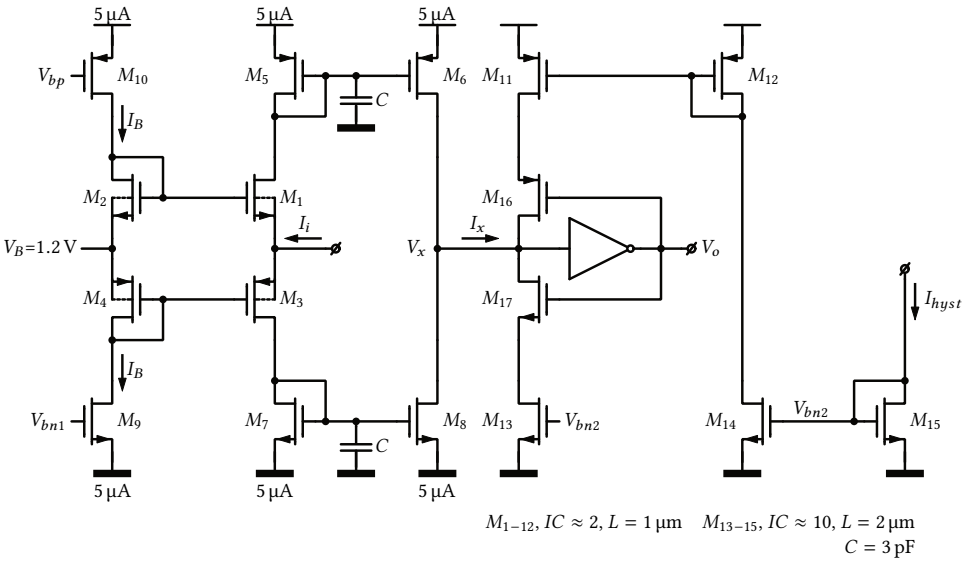


Figure 5.29: Circuit diagram of the current comparator with hysteresis used in the window comparator.

ing block (transconductor stage, Section 5.2) around 1.2 V. In a symmetrical circuit with identical devices, when $I_i = 0$, all branch currents are equal, $I_x = 0$ and V_x would float mid-rail. However, any small mismatch (e.g. in device output conductance) that results in non-zero I_x, will trigger the decision circuit and V_x will be pulled hard to one of the supply rails (due to the large incremental gain).

For $I_i \gg I_B$, M_1's drain current $I_{d1} = 0$ and the voltage established at the input node is:

$$V_i = V_B - V_{GS4} + V_{gs3}(I_i) \tag{5.92}$$

in which the dependence on the input current is made explicit. Conversely, for $I_i \ll -I_B$, $I_{d3} = 0$:

$$V_i = V_B + V_{GS2} - V_{gs1}(I_i) \tag{5.93}$$

M_{1-4} operate at the center of moderate inversion $IC \approx 2$ at the chosen bias current $I_B = 5\,\mu A$. This will limit the difference in effective gate-source voltage ΔV_{EFF} of the input transistors—and hence the deviation of V_i, the voltage applied to the preceding stage—to 250 mV across the input range. This sizing also results in beneficially low $V_{DS,sat}$.

Mirror devices M_{5-8} also operate at $IC \approx 2$ to limit the developed gate source voltage over the input range in order not to push $M_{1,4}$ in the triode region. Capacitance C is part of the noise and ripple rejecting low-pass filter. The cut-off frequency $g_{m5,7}/(2\pi C)$ is chosen at a few MHz, as a trade-off between propagation time and ripple rejection. Bias transistors $M_{9,10}$ are chosen equal to $M_{5,7}$ for layout convenience.

The output resistance of the input stage is $1/(g_{ds6} + g_{ds8})$. The input resistance is approximately $1/(g_{m1} + g_{m3})$, and depends on the input current.

5.6.1.2 Decision Stage

The decision stage is an amplifier with positive feedback gain boosting created by connecting two inverters (complementary common-source amplifiers) in a loop (redrawn in Fig. 5.30). Devices $M_{n2,p2}$ constitute the output inverter, while $M_{n1,p1}$ is the input inverter which is driven by two current sources I_P and I_N (implemented with MOS devices). The input current is applied to node X. The diodes model the cut-off operation of the current sources: as the voltage across the current source decreases beyond the compliance voltage, their output current drops to zero.

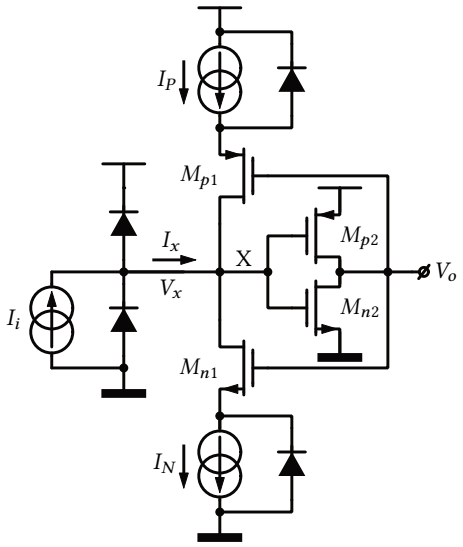


Figure 5.30: Principle circuit diagram of the decision stage in the current comparator.

The large-signal operation is as follows:

- Assume $I_i \ll 0$. Node X is discharged, V_x is pulled to the lower supply rail and the output V_o is high. The current through both M_{n1} and I_N is zero as they are in very deep triode region, $V_{DS} = 0$. In addition, neither M_{p2} nor I_P carry current, as they are in cut-off regime ($V_{GS,p1} = 0$). Hence, no current flows in or out of node X, $I_x = 0$. The loop gain is zero.
- As I_i increases, V_o initially remains high. M_{p1} and I_P remain in cut-off. V_x remains approximately zero until I_N outputs its nominal current. When $I_x = I_N$, node X charges as the input current has nowhere to flow anymore, V_x increases and M_{n1} is pulled quickly toward saturation. As the loop gain inflates to -1, net positive feedback sets in.
- When V_x crosses the threshold of the output inverter, V_o toggles pushing M_{n1} and I_N in cut-off as $V_{GS,n1} = 0$. Soon after, V_x has increased to the positive supply rail such that both M_{p1} and I_P enter deep triode region and carry no current. The loop gain vanishes again. No current flows in or out of node X, $I_x = 0$.
- Similar operation hold when I_i now drops again. The switching point occurs for $I_i = I_P$.

Clearly, the decision stage exhibits well-defined hysteresis with switching points I_N and I_P. The capacitance at node X determines the speed.

For small-signals, the instantaneous loop gain can be expressed as:

$$T(I_i) \approx -\frac{g_{m1,\text{eff}}(I_i)}{g_{ds1,\text{eff}}(I_i)} \frac{g_{m2,\text{eff}}(I_i)}{g_X(I_i)} \tag{5.94}$$

in which the negative sign implies positive feedback. In principle, it is possible to derive the switching points based on this expression, similar to what was done for the voltage comparator in the quantizer (Section 5.5.2). However, this is very tedious as it is difficult to express the quantities involved as a function of I_i as they are a complex function of multiple elements (degenerated g_m, g_X) or are difficult to model (g_{ds}) to begin with.

In Fig. 5.29, I_P and I_N are nominally equal and implemented with current mirrors M_{11-15} which operate at the onset of strong inversion and long gate length ($IC \approx 10$, $L = 2\,\mu\text{m}$) as a trade-off between current mismatch, area and drain referred noise. The inverters are taken from the standard cell library provided with the technology.

5.6.1.3 Simulation Results

Monte Carlo simulations with auto stop, including PVT variations and local mismatch parameters, show that the 3σ input-referred offset and propagation time of one comparator are 150 nA and 3.8 µs, respectively, which is sufficient for this application. A dc sweep of the input current illustrating the hysteresis is shown in Fig. 5.31, for $I_{\text{ref}} = 50\,\mu\text{A}$, $I_{hyst} = 10\,\mu\text{A}$.

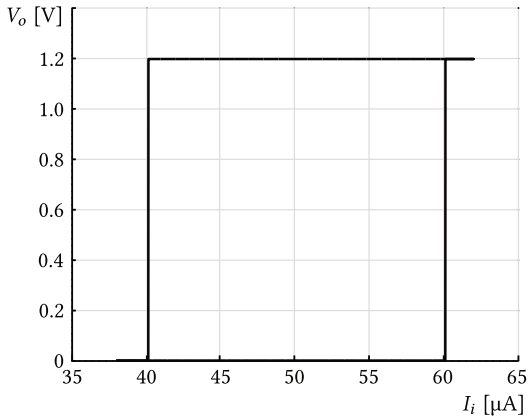


Figure 5.31: Output voltage of the current comparator as a function of input current. $I_{\text{ref}} = 50\,\mu\text{A}$, $I_{hyst} = 10\,\mu\text{A}$.

5.7 Look-up Table

The digital LUT uses the quantized AGC loop error signal **d** as an index for a code word **c** in a memory block. The code word represents a state of the VGA and programs its gain and frequency response (Section 3.2.1). The LUT is described in SystemVerilog, a HDL [88, 89]. Using a digital implementation flow with automatic synthesis to gate-level primitives and place-and-routing allows to include a degree of intelligence while saving massive amounts of time.

The realized building block (Fig. 5.32) includes control logic and a double look-up table (two *slices*). The latter allows to quickly reconfigure the datapath using either a dedicated package pin or a serial peripheral interface (SPI) command. This is intended for use in an automatic bit rate selection system. The block is implemented as a module that is connected to the internal SPI bus, just as other modules and registers. A second bus connects to the analog core—datapath and AGC system. The internal SPI bus is a parallel bus with separate address, forward and return data wires. The chip communicates with the outside world through a controller module that implements a 4-wire serial SPI protocol.

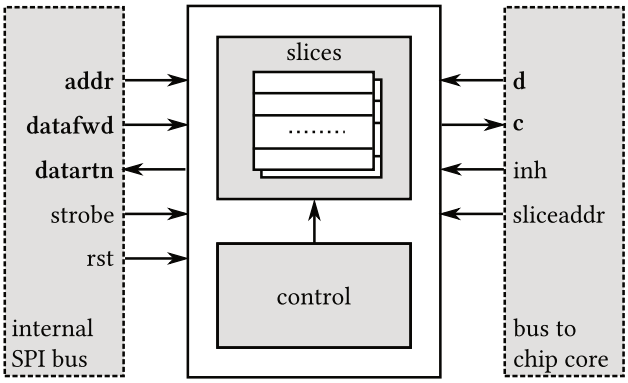


Figure 5.32: Block diagram of the digital lookup table, including control logic and interface with the internal SPI bus and the analog chip core.

The look-up table is pure combinatorial logic; any change in the 3-bit code word **d** selects a word **c** that is applied to the datapath, unless `inh` is asserted, as explained in Section 3.2.5. The relevant code is:

```
[...]
logic [`STATE_DATA_WIDTH-1:0] slices[`SLICE_CNT][`STATE_CNT];
[...]
assign intInh = (inhmode == AUTO) ? coreBus.inh
                                  : ((inhmode == MANUAL_HIGH) ? 1'b1
                                                              : 1'b0);
assign curSliceAddr = (sliceselmode == EXT) ? coreBus.sliceaddr
                                            : intSliceAddr;
always_latch begin
    if (inhmode == MANUAL_HIGH)
        coreBus.c <= slices[curSliceAddr][manualHighStateAddr];
    else
        if (~intInh) coreBus.c <= slices[curSliceAddr][coreBus.d];
    end
end
[...]
```

A difference in propagation delay between code paths can introduce glitches, which could translate into undesired datapath behavior. However, in the 0.13 µm CMOS logic (Section 1.3), any glitch duration is extremely small compared to the very slow AGC loop dynamics, such that they have essentially no impact at all (see also Section 3.2.7.3).

The control part is sequential logic driven by the SPI controller and is only active on request. It's main task is decoding and executing commands, as soon as the module is addressed. The following listing illustrates the idea:

```
[...]
assign spiBus.datartn = (spiBus.addr == addr) ? datartnbuf : 'z;

always_ff @(posedge spiBus.rst or posedge spiBus.strobe) begin : processpkt
    CC_pkt_s pkt;
    if (spiBus.rst) begin
        [...]
        resetSlices();
    end else if (spiBus.addr == addr) begin
        pkt = spiBus.datafwd;

        unique case (pkt.CMD)
            CMD_RSTATE:
                datartnbuf <= slices[pkt.PLD.data_R.sliceaddr][pkt.PLD.data_R.stateaddr];
            CMD_RINH:
                datartnbuf <= coreBus.inh;
            CMD_RSTATEADDR:
                datartnbuf <= coreBus.stateaddr;
            CMD_WSTATE:
                slices[pkt.PLD.data_W.sliceaddr][pkt.PLD.data_W.stateaddr]
                      <= pkt.PLD.data_W.statedata;
            CMD_INHMODE:
            begin
                inhmode <= pkt.PLD.data_INHMODE.inhmode;
                if (pkt.PLD.data_INHMODE.inhmode == MANUAL_HIGH)
                    manualHighStateAddr <= pkt.PLD.data_INHMODE.stateaddr;
            end
            CMD_SLICESELMODE:
                [...]
        endcase

    end
end : processpkt
```

The following commands are implemented:

- RSTATE: read a word from a slice at a given address.
- WSTATE: write a word into a slice at a given address.
- RINH: read the inh signal generated by the window comparator.
- INHMODE: allows to override inh. When the override value is logic high, the look-up table can be controlled externally instead of by the AGC loop.
- SLICESELMODE: internal or external slice selection.

The code shows that it is possible to reconfigure the datapath on the fly by writing new data to the look-up table and to enable or disable the AGC loop externally. This flexible implementation makes the receiver highly programmable, expanding its use cases and possibilities.

Note the vast use of SystemVerilog's `packed structs`, `packed unions` and enums, that allow easy manipulation of packets by addressing data using meaningful names. This higher data abstraction level helps to avoid mistakes.

Cadence's RTL Compiler and Encounter were employed for the gate-level translation and place-and-routing, respectively. A total of 1645 logic gates and an area of $255\,\mu m \times 200\,\mu m$ were used.

5.8 Peripheral Circuits

The AGC system is integrated in a dual-channel optical receiver array. Several peripheral building blocks are included such as bias current generation and distribution, test structures and a digital interface.

Biasing, Current DACs and Current Mirrors

Various currents are digitally programmable. For example, the window width and hysteresis in the window comparator are currents, as is the desired amplitude for the AGC loop. In addition, the bias network operates in the current domain to avoid malfunction due to IR-drop over the wires. As almost all bias currents are derived from a single reference current (derived from an external reference), the current distribution network necessarily consists of both global and local current mirrors. All these mirrors are dc optimized, for obvious reasons.

Test Tree

Comprehensive verification of an integrated system requires that key signals or signal derived thereof are observable to the external world. System inputs or outputs are generally easy to examine as they are attached to dedicated pins. However, to confirm more thoroughly the functionality of the individual building blocks, dedicated test structures that increase observability need to be built into the chip.

The implemented chip has a single analog test pin that is multiplexed such that selected low-speed voltages or currents are exposed. Examples include: local supply rails, a copy of local bias currents, average input photo current, the balancing voltage of the datapath, AGC loop error signal, detector outputs… Some signals are buffered to minimize disturbances. As multiple channels need to share a single test pin, the multiplexing is multilevel. It is digitally controlled via the SPI interface.

SPI Interface and Digital Registers

Each channel has a number of digitally programmable signals, test structures and switchable test routing. The analog multiplexers and digital settings can be programmed via a SPI protocol. Figure 5.33 depicts the logic block diagram of the internal structure.

An external device communicates with a controller via a 4-wire synchronous serial SPI protocol. The controller is the master of an internal parallel communication bus, which consists of a forward and return datapath and address, strobe and reset lines. Each channel contains the LUT for the AGC loop and a register. In addition a separate register controls some global settings. These blocks are slaves of the controller and have a unique address.

These digital blocks are written and verified in SystemVerilog and implemented with a digital flow.

5.9 Experimental Results

The AGC system was integrated in a dual-channel linear optical receiver. The chip was fabricated and designed in the 0.13 µm SiGe BiCMOS technology described in Section 1.3. A die micrograph, including photo diode array, is depicted in Fig. 5.34. The AGC part measures 600 µm × 680 µm, excluding bond pads, of which one quarter is due to supply

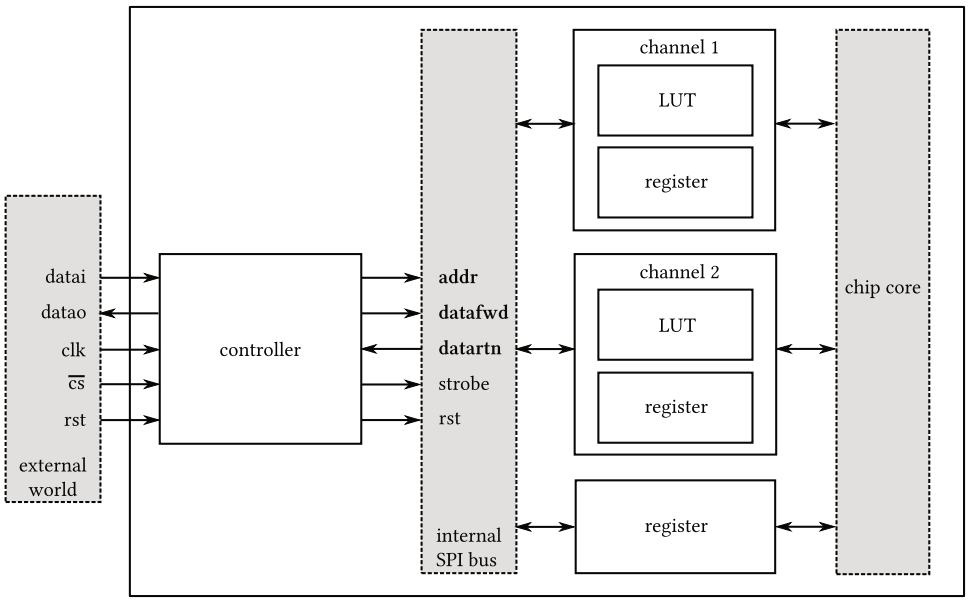


Figure 5.33: Logic block diagram of the digital SPI interface and internal structure.

filtering and decoupling. The chip is not packaged, but placed in a cavity and bonded directly to a test board. This avoids additional package parasitics that could degrade the performance of the high-speed datapath.

The test board supports two types of input current source. For low-speed tests (up to 500 MHz), an electrical photo diode emulation circuit is used (Fig. 5.35). The output signal of an RF arbitrary waveform generator is converted to a current by resistor R and superimposed on a variable dc current. This allows to apply various input waveform shapes to the receiver in order to measure input-output characteristics, step response of the AGC system… The impact of the input network on the transfer function of the transimpedance amplifier (TIA) input stage, via the loop gain (as in Section 6.3.2), was checked in simulation to make sure the functional changes were negligible. The input current can be measured across the resistor with a high-speed differential probe.

For full-rate optical tests, a photo diode array is bonded to the receiver (Fig. 5.34). However, the available high-speed equipment is not as flexible and no arbitrary pulse shapes can be generated. In addition, manual alignment of the optical fiber with the photo diodes makes this setup more cumbersome.

5.9.1 Low-Speed Electrical Tests

5.9.1.1 Calibration and Input-output Characteristics

Before all else, the LUT needs to be calibrated, such that the small-signal gain of the datapath associated with each entry in the LUT (selected with code word **c**) approximates the ideal gain listed in Fig. 4.3(b). For the low-speed tests the high-frequency response of the datapath (overshoot, bandwidth) is not important and no attempt was made to optimize or measure these parameters.

A 1 MHz sine wave, 100 % amplitude-modulated with a 75 Hz triangle, was applied

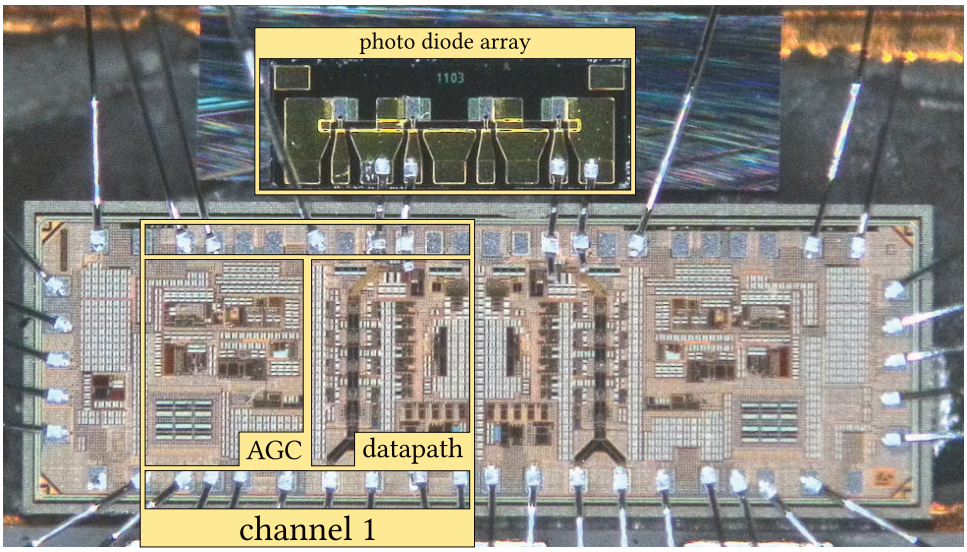


Figure 5.34: Die micrograph of the bonded linear optical receiver and photo diode array.

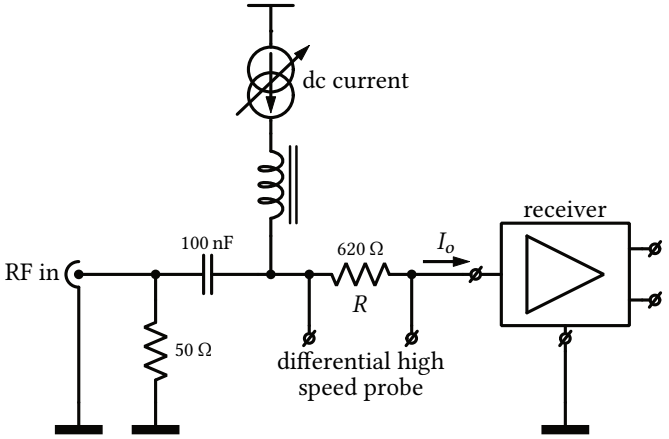
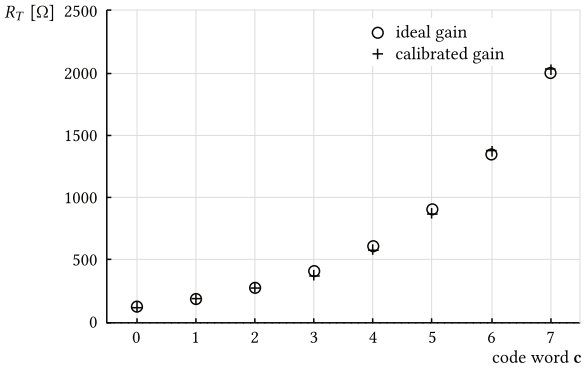


Figure 5.35: Simplified photo diode emulation circuit.

to the device under test (DUT). Both the voltage measured across the input resistor of the photo diode emulation and the DUT output voltage were applied to precision peak detectors. The chosen frequency is within the input range of the peak detectors and still sufficiently above the high-pass pole of the datapath to not introduce a significant error. Based on the measured peak values, the gain of the datapath was manually calibrated for each code word. Figure 5.36 compares the ideal gains with the calibrated gains. Clearly the required exponential shape is obtained to a good degree (Section 3.1.3). Note that it is not necessary to achieve high absolute accuracy in light of the nature of the detector in the receiver (Section 5.1).

The input-output characteristics (Fig. 5.37) were measured by plotting both outputs of the peak detectors in X-Y mode on the oscilloscope (with boxcar averaging). The slope



(a)

c	ideal $G(\mathbf{c})$ [Ω]	cal. $G(\mathbf{c})$ [Ω]
0	125	122
1	186	188
2	276	277
3	410	375
4	610	580
5	906	873
6	1346	1375
7	2000	2030

(b)

Figure 5.36: Calibrated datapath gain and ideal target gain for each code word.

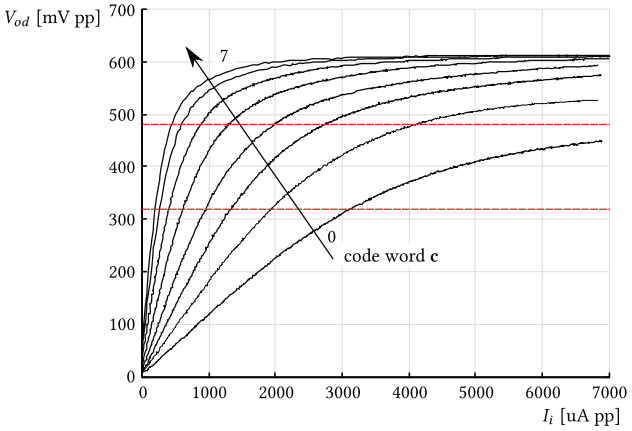


Figure 5.37: Calibrated amplitude input-output characteristic for each code word.

in the origin of each curve is the corresponding gain in Fig. 5.36(b). Although each curve is compressive, this non-linearity could be tolerated to some extent depending on the modulation scheme. In any case, it is a property of the high-speed datapath and can be revised if desired. The target output range between 320 mV to 480 mV is indicated. For peak-peak input currents higher than 200 µA, the differential peak-peak output voltage is within the target range (Table 4.1), indicating sufficient dynamic range.

5.9.1.2 Functional Test

The waveforms in Fig. 5.38 demonstrate the basic functionality of the AGC system. A 500 MHz sine, 100 % amplitude-modulated with a 14 Hz triangle, is applied to the receiver. The upper trace shows the single-ended output voltage. The middle trace is the `inh` logic signal, indicating when the detected output signal is within range. The lower trace is the amplified AGC loop error signal V_l (Section 4.1). The circuit was programmed in follower-mode, such that when `inh` is asserted, the loop filter is reconfigured to follow the midpoint voltage of the current quantization step (Section 4.1).

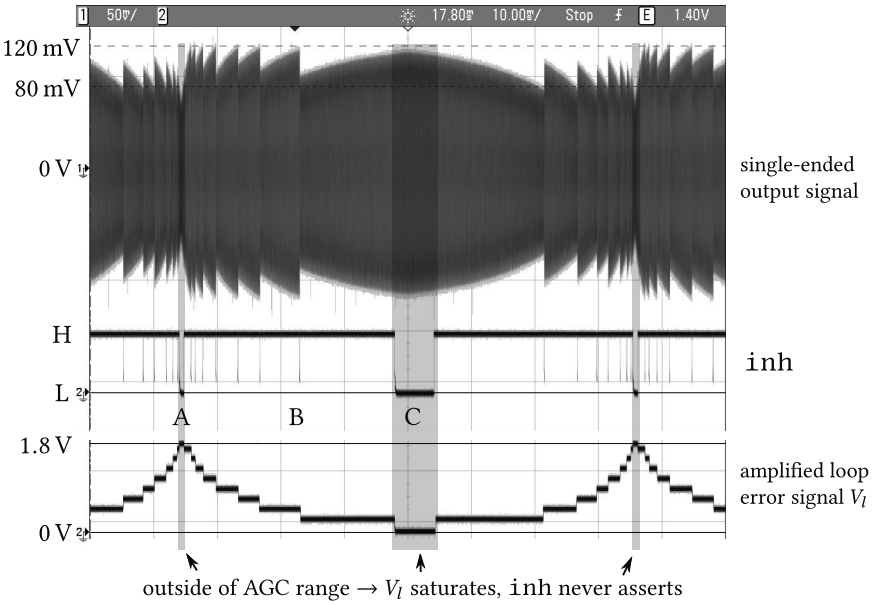


Figure 5.38: Single-ended output signal, inh and amplified loop error signal V_l with triangle-modulated 500 MHz input sine.

Very low inputs (region A) are outside of the AGC input range and inh is always low. The output of the loop filter saturates to the upper rail (clamped to 1.8 V, Section 4.1) and the highest code word is selected, corresponding to maximum gain.

As the input amplitude increases (region B), so does the output amplitude until it is within range and inh asserts. The loop filter now follows the midpoint of the highest quantization step (1.7 V). As soon as the detected output is outside of the target range, inh deasserts and the AGC loop is restored. Error signal V_l quickly decreases until the next, lower, gain is selected, at which point the output is again within range and inh is asserted again. The loop filter now follows the midpoint of the next-to-highest quantization step.

When the input signal is so large that, even at the lowest gain setting, the output signal is outside of the AGC range, inh never asserts (region C). The error signal saturates to the lower rail until the input signal decreases again.

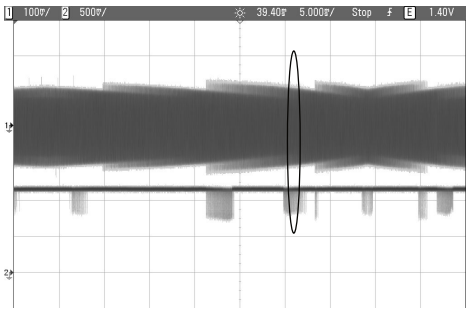
5.9.1.3 Limit Cycling

Due to the quantized nature of the AGC loop, limit cycling occurs as explained in Section 3.2.2. A system to avoid these undesired oscillations is described in Section 3.2.5.

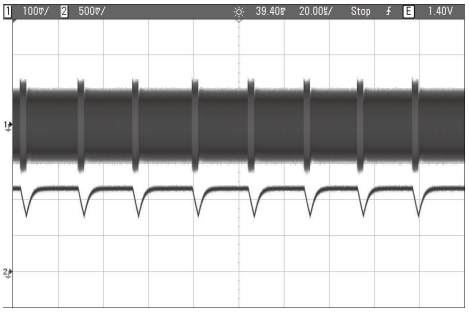
This phenomenon is illustrated in Fig. 5.39(a), with a detail shown in Fig. 5.39(b). A 500 MHz sine, 100 % amplitude-modulated with a 14 Hz triangle, is applied to the receiver. The upper waveform is the output signal, while the lower waveform is inh. The system was programmed to ignore the inh signal. The development of limit cycles is clearly visible. For a given input amplitude, the gain of the datapath is switched between two adjacent values. The detected output is alternately inside or outside the target output range, such that inh toggles at the frequency of the limit cycle. The frequency of the limit cycle is determined by the non-dominant poles of the AGC loop gain, while the duty

cycle is determined by the input and desired signal, such that the average detected output amplitude equals the desired amplitude (Sections 3.2.2.2 and 3.2.4). Note that the limited drive capability of the internal logic buffer that generates (a copy of) `inh`, in combination with internal resistance of the test tree switches and the 1× probe capacitance (~14 pF) results in the deformed shape of `inh`.

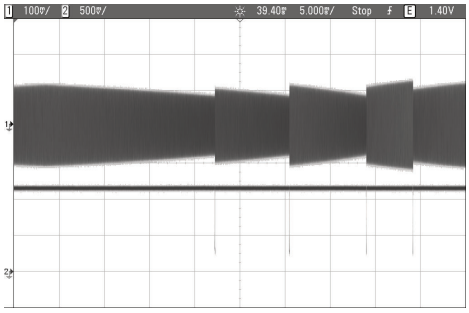
In Fig. 5.39(c), the system is configured to not ignore `inh`. Clearly the limit cycling has disappeared. Both the output signal and `inh` are well-behaved. This confirms the existence of limit cycles and the capability of the system to avoid them.



(a) Limit cycling occurs when `inh` is ignored.



(b) Detail: the gain switches between two states, such that the average detected output amplitude equals the desired amplitude.



(c) No limit cycling occurs in normal operation.

Figure 5.39: Limit cycling. Amplitude of the input signal (not shown) is modulated with a triangle. Upper trace is single-ended output, lower trace is `inh`.

5.9.1.4 Step Response

The step response was measured by applying a 500 MHz sine, 100 % amplitude-modulated with a 14 Hz square to the receiver, configured in follower mode. Due to the quantized nature of the system, the settling time cannot be defined or measured as the time it takes to reach steady state within a certain margin. Instead, the settling time is defined as the time elapsed between application of the input step and the point were the output signal enters the target range—where `inh` is asserted. This is illustrated qualitatively in Fig. 5.40 for a falling step. Before the step, in steady state, the input is outside of the AGC range, V_l is minimal and `inh` is deasserted. The datapath exhibits high-pass behavior in the

amplitude domain as predicted by Eq. (3.25). As the input step occurs, initially the loop cannot respond due to its limited GBW and the output amplitude also experiences the step—this is really nothing more than direct forward transmission from input to output. As the loop catches up, V_l increases, changing the datapath gain until the output is within range. At that moment `inh` asserts.

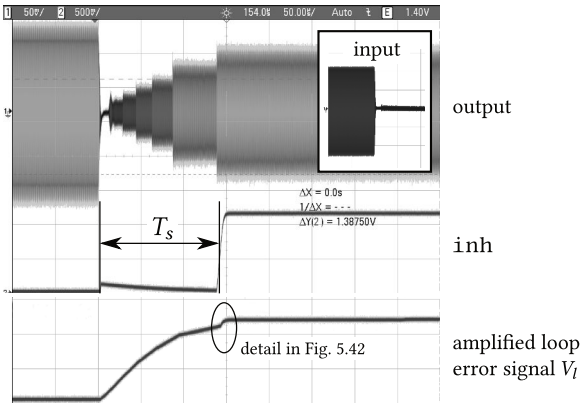


Figure 5.40: Step response illustration. Upper trace is single-ended output signal, middle trace is `inh`, lower trace is V_l.

As explained in Section 3.2.5, the settling time is not independent from the step amplitude (as for the equivalent continuous-time system) but is guaranteed to be upper bounded. Figure 5.41 shows the settling time for falling steps when the input step becomes increasingly larger. Clearly, it is never larger than 205 µs, the specified 2 % upper bound settling time (Section 4.5). The results are similar for rising steps.

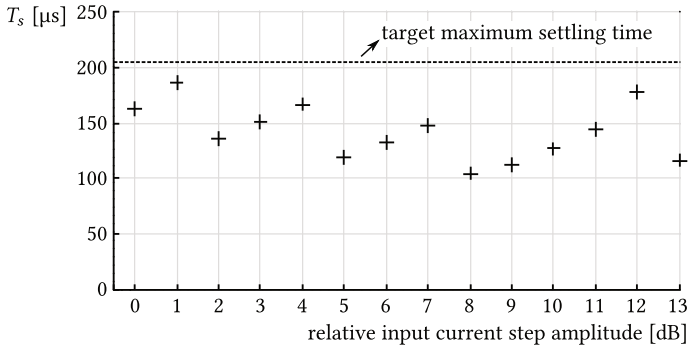


Figure 5.41: Settling time as a function of relative input current step amplitude.

5.9.1.5 Small-signal AGC Loop Filter Step Response

The step response of the AGC loop filter (which is compensated with capacitance multipliers, Section 5.4) can be measured using the same configuration as in the previous section. Indeed, as soon as `inh` asserts, the loop filter is reconfigured such that it follows

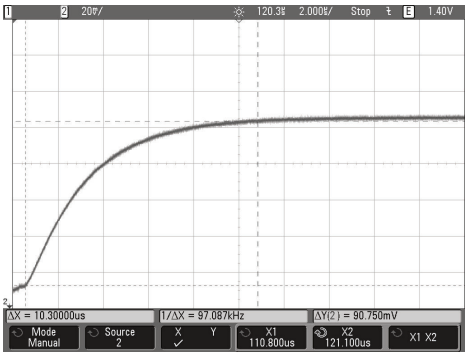


Figure 5.42: Small-signal step response of the AGC loop filter, configured as a follower. The 2 %-settling time is 10.3 µs.

the midpoint voltage between two quantization steps. This is equivalent to the application of a small step to the (now) unity-gain feedback configuration. This can be observed in Fig. 5.40 and is captured in more detail in Fig. 5.42. The response is well-behaved without under- or overshoot. The measured 2 % settling time is $T_s \approx 10.3\,\mu s$, which is equivalent to a closed-loop bandwidth of the follower (or GBW of the loop gain) of $f_{3dB} \approx 4/(2\pi T_s) \approx 61.8\,kHz$. This agrees with the design specification of 60 kHz (Section 4.5).

5.9.2 High-Speed Optical Tests

The following sections present test results at high speed. Optical PRBS NRZ data at 26 Gb/s, with high extinction ratio (ER), is applied to the photo diode bonded to the receiver. A system similar to the one presented is Section 6.3.4 is used to measure the average photo current. As mentioned in Sections 4.2 and 5.1, multilevel modulation merely requires a different setpoint of the AGC loop, as only the detector depends on the output signal shape. Hence, results with NRZ data are sufficient. In addition, only results that concern the AGC system are shown, as this is the focus of this work and not the datapath or high-speed receiver features by itself.

5.9.2.1 Detector

Input-output characteristics of the detector are shown in Fig. 5.43 for both low- and high-speed inputs. They are measured by applying a signal to the datapath and measuring the output signal amplitude A_{id} with a high-speed sampling oscilloscope, as well as the internal detector output V_o. The datapath was programmed to a fixed gain. Also shown are the theoretical curves for an ideal sine and square wave at 300 K (Eq. (5.23)).

The low-speed input is a 500 MHz sine wave generated using the photo diode emulation circuit. The offset from the theoretical curve is attributed to a combination of higher on-chip temperature (for small amplitudes) and systematic mismatch in the detector (across the entire range). As the amplitude increases, the compressive nature of the input-output characteristics (Fig. 5.37) deforms the sine wave while propagating through the datapath. This explains the larger deviation from the theoretical curve at higher amplitudes.

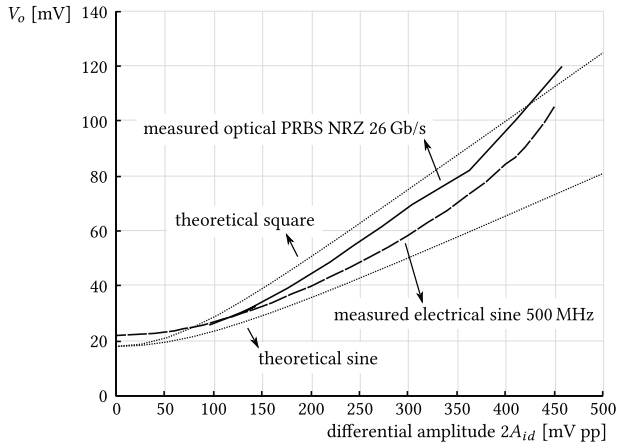


Figure 5.43: Measured and theoretical detector output, for various input signal shapes, at low and high frequency. The theoretical curves are valid at 300 K.

The curve for an optical NRZ 26 Gb/s PRBS input follows closely the theoretical curve for the square wave, although with an offset due to systematic mismatch and the non-ideal square wave shape of the input signal, which is partly restored at higher amplitudes due to the compressive behavior.

These results show that the behavior of the detector can be reasonably well predicted, both for low- and high-speed signals.

5.9.2.2 Functional Test

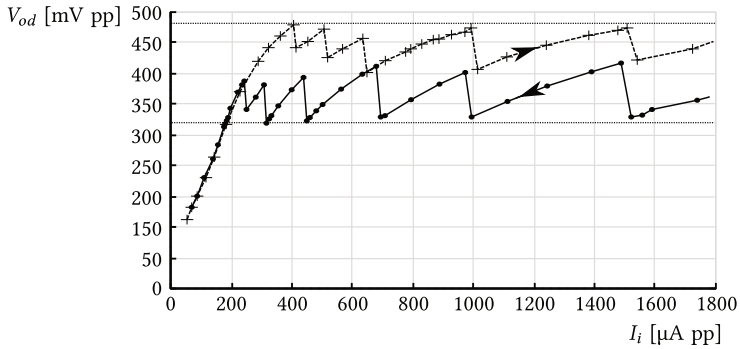


Figure 5.44: Functional test of the system as a function of input photo current for 26 Gb/s NRZ PRBS data.

Figure 5.44 demonstrates the functionality of the AGC system with full-rate NRZ 26 Gb/s PRBS optical input, in which the measured differential output amplitude is plotted versus measured input photo current. The output amplitude was measured with a high-speed sampling oscilloscope and has some measurement uncertainty due to the impact of noise and inter-symbol interference (ISI).

As the input power increases, the gain is progressively switched from high to low. For

decreasing input power, the gain is progressively switched from low to high. Ideally, these curves should follow the characteristics of Fig. 5.37 in a piecewise fashion. This is confirmed in the figure. However, it can be seen that the (large-signal) datapath gain for each code word is a bit higher than that for the 1 MHz sine wave, used for calibration. Recalibration was not performed as the larger absolute gain does not pose a problem. Indeed, as long as the intercept of the VGA remains unchanged, the settling time is unaffected (Eqs. (3.24) and (3.27)). Note that the lowest gains are not visited due to the limited optical power available in the measurement setup.

The exact transition points are difficult to establish accurately due to the measurement uncertainty and built-in hysteresis in combination with the very high-speed signals, but the allowed range of output amplitudes is in good agreement with the specification. This test confirms the validity of the event-driven AGC approach at very high speeds and the resulting massive increase in dynamic range.

5.10 Conclusion

This chapter presented the implementation, on the circuit level, of the event-driven AGC system, embedded in a high-speed optical receiver. The system comprises a mixture of analog and digital blocks, which require vastly different design approaches and implementation flows. Several key analog building block were described in detail, including discussion of topologies, circuit operation and design tradeoffs. Choices regarding device sizing were explained. The digital blocks are described in the SystemVerilog HDL and automatically converted to the transistor level.

Experimental results conducted on the realized integrated circuit chip confirm the functionality of the system, both at low and high speed.

Chapter 6

Multichannel Optical Receiver

In this chapter, the design and implementation of the $4 \times 25\,\mathrm{Gb/s}$ optical receiver array for non-return-to-zero (NRZ) modulation is discussed. Section 6.1 presents the design objectives. Next to high data rate, high sensitivity and low power consumption, the tight integration with the photo diode array is a major design challenge. The channels have to fit in 250 µm lanes, while keeping crosstalk to a minimum. In Section 6.2, the architecture of a channel is proposed. Section 6.3 elaborates in detail on one of the key building blocks of the datapath, the transimpedance amplifier (TIA) input stage. Bandwidth enhancement through inductive peaking and techniques to increase the dynamic range are discussed. Other blocks in the receiver are presented more concisely (the main amplifier (MA), Section 6.4) or are only touched upon (balancing loop and peripheral circuit, Sections 6.5 to 6.6). Finally, in Section 6.7, experimental results are presented which confirm the performance of the receiver array.

6.1 Design Objectives

The specifications of the photo diode array are listed in Table 6.1. The small-signal model is discussed in Section 2.6. Figure 6.1 shows a typical four-channel photo diode array.

Specification	Symbol	Unit	Typical value
type			p-i-n
reverse bias voltage		V	2
responsivity	$\mathcal{R}$	A/W	0.4
series resistance	R_{PD}	Ω	10
capacitance	C_{PD}	fF	115
small-signal bandwidth	$f_{3\mathrm{dB}}$	GHz	25
number of photo diodes			4
channel pitch		µm	250

Table 6.1: Photo diode array specifications. R_{PD}, C_{PD} and $f_{3\mathrm{dB}}$ at specified reverse bias voltage.

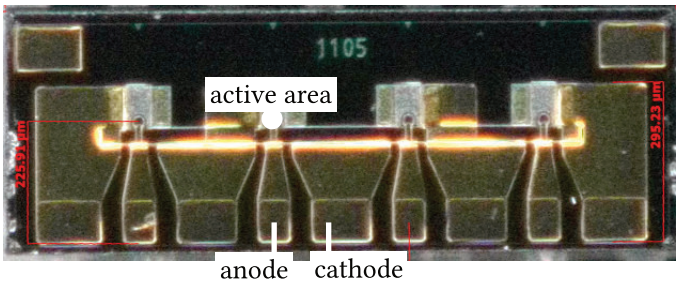


Figure 6.1: Photograph of a 4-channel photo diode array with 250 µm pitch.

Table 6.2 list the receiver array design objectives. It can be shown that for NRZ modulation, as a rule of thumb, the optimum receiver bandwidth should be 60 % to 70 % of the bit rate R_b [18]. This follows from a trade-off between inter-symbol interference (ISI) and input-referred noise: as bandwidth increases, ISI (Section 2.5.2) decreases, while input-referred rms current noise (Section 2.4) increases.

Specification	Symbol	Unit	Min.	Typ.	Max.
bit rate	R_b	Gbps		25	
small-signal bandwidth	$f_{3\text{dB}}$	GHz		16.6	
input-referred current noise	i_n^{rms}	µA		4.5	9
sensitivity (BER = 10^{-12})[1]	$i_{\text{sens}}^{\text{pp}}$	µA		63	126
	$\overline{P}_{\text{sens}}$	dBm		-11.1	-8.1
small-signal gain		dBΩ		69	
dynamic range	DR	dB	10		
differential output voltage	V_{od}^{pp}	mV	150	300	
supply voltage	V_{DD}	V		2	
power consumption per channel	P_{ch}	mW			90
number of channels				4	
channel pitch		µm		250	

Table 6.2: Optical NRZ receiver and channel design objectives.

Special attention was paid to the power consumption per channel and associated heat generation, as any active cooling is unacceptable (Section 1.1.2). Also, the channels must be tightly integrated, adjacent to each other in 250 µm wide lanes. This requirement stems from the intimate integration demands on both the receiver array and photo diode array, which have a standard pitch of 250 µm [51]. This has implications on the channel architecture (Section 6.2.1).

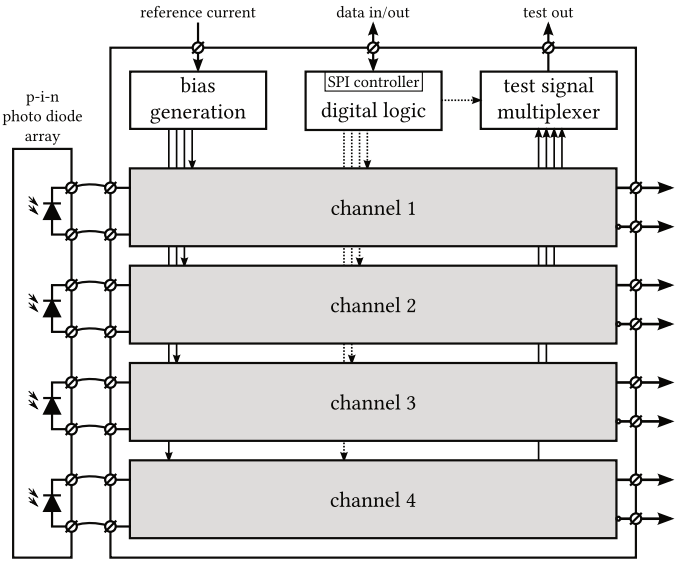


Figure 6.2: Architecture of the realized optical receiver array.

6.2 Receiver Array Architecture

Figure 6.2 shows the architecture of the realized four-channel optical receiver array. Light from the fiber array is coupled to the photo diode array. Each photo diode is connected to its dedicated channel. The channel pitch equals the photo diode array pitch of 250 µm. A common bias block generates the main bias currents for the channels, based on an external reference current. A digital serial peripheral interface (SPI) interface and accompanying logic allows tuning the gain and bandwidth of the channels to some extent and controls an analog test signal multiplexer. The latter allows to connect selected internal voltages and currents of the different channels to a dedicated analog test pin.

6.2.1 Channel Architecture

Figure 6.3 depicts the architecture of a single channel. Both anode and cathode of each photo diode are bonded to the die in order to keep loop inductance low and reduce susceptibility to interference. The p-i-n photo diodes require a reverse bias of 2 V to 2.5 V. As the TIA input generates an anode voltage of around 850 mV (Section 6.3), an extra supply rail of 3.3 V is provided for the cathode connection. This supply is shared across all channels to reduce the number of pins. Local filtering at each cathode node reduces interference. In addition, the average photo current is measured and used to adaptively bias the TIA in order to extend its dynamic range (Section 6.3.4). This also provides a convenient test signal to measure the photo current which can be used to assist during fiber alignment.

The datapath of each channel consists of a TIA input stage and a MA. The latter is composed of a single-ended to differential (S2D) converter stage, three low gain high-bandwidth gain stages and a 50 Ω output buffer. The TIA limits the total receiver band-

[1] Based on noise, NRZ-modulation, high ER. Optical sensitivity derived with typical photo diode responsivity in Table 6.1.

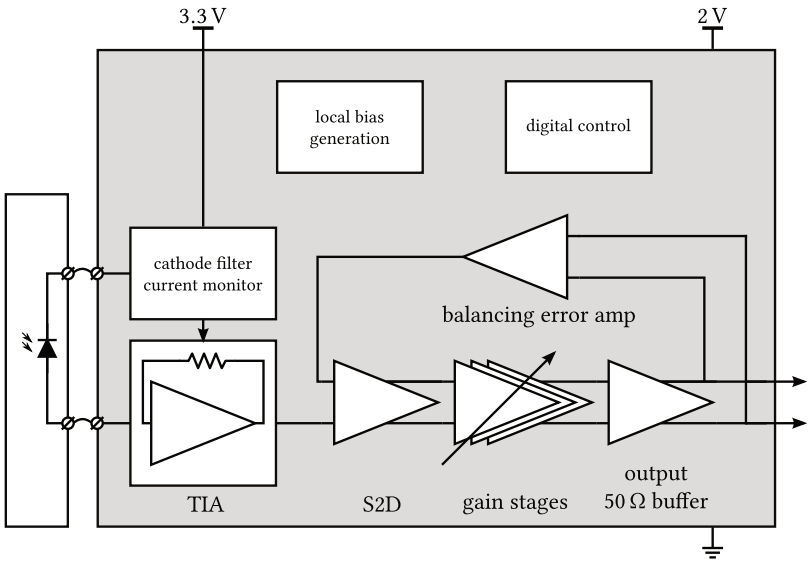


Figure 6.3: Architecture of a channel.

width. The MA is built with a larger bandwidth. This allows for higher TIA transimpedance and better noise performance.

The MA is implemented differentially, but is driven from the single-ended TIA. This is a mainly due to the limited available area and noise requirements. A true differential TIA (driven single-ended by the photo diode) is more immune to common-mode interference, but has up to $\sqrt{2}$ times higher input-referred noise (1.5 dB worse sensitivity) and is twice as big as its single-ended version [18]. The single-ended TIA is more compact, but sensitive to common-mode interference. Special measures are taken to reduce the impact. Nonetheless, now a reference voltage is needed to drive the differential MA such that the outputs of the datapath are balanced. Although a small dummy TIA could be used, in this implementation a slow control loop removes the dc offset between the differential output signals by adjusting the dc voltage at the inverting terminal of the S2D stage, thus providing a balanced differential output signal. Measuring the output signals has the advantage that offsets in the MA are also compensated, in addition to providing dc balancing. As shown in Section 6.5, a high-pass pole is created in the transfer function of the datapath. In order to limit the power penalty due to baseline wander (Section 2.5.3), this pole should be at low enough frequency. The balancing error amplifiers are physically located outside of the channel lanes, at the sides of the chip (Section 6.7). As such, area requirements are relaxed somewhat. However, this requires long wires from outside of the lanes to the inputs of the S2D stages. Noise interference through capacitive coupling is limited by the low output impedance of the error amplifier, while inductive coupling is reduced by proper shielding of the wires [47].

Each channel is supplied by its dedicated 0 V and 2 V rails to reduce interchannel crosstalk, in addition to deep trench isolation and the differential topology (Section 2.5.4).

6.3 Transimpedance Amplifier

The TIA converts photo current to a voltage. It needs to be high-bandwidth and low noise to obtain the required sensitivity (Section 6.1). As the first stage of an amplifier chain, it dominates noise behavior of the entire receiver. Further requirements are sufficient dynamic range, low power consumption and small area. As explained in Section 6.2.1, the TIA is single-ended.

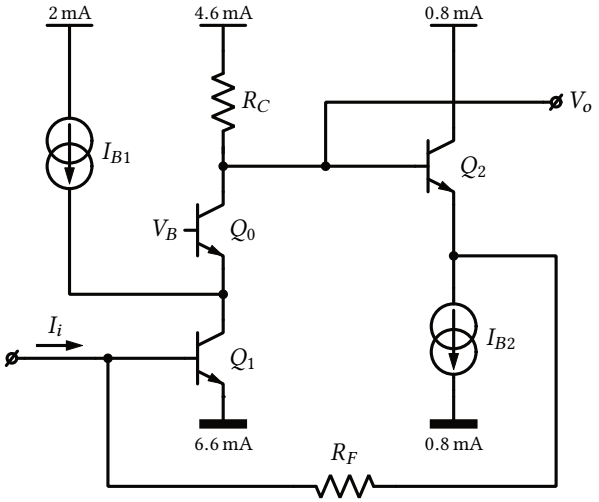


Figure 6.4: Simplified circuit diagram of the TIA.

A simplified circuit diagram is depicted in Fig. 6.4. It is a shunt-shunt feedback amplifier topology, providing both a low-impedance input and output node. Common-emitter amplifier Q_1 and emitter follower Q_2 make up the forward amplifier, while feedback resistor R_F (together with the total input capacitance) represents the feedback path. Cascode Q_0 protects Q_1 from excessive collector-emitter voltage and reduces its Miller capacitance contribution to the input capacitance. In addition, it provides a convenient low-impedance input for current source I_{B1}, which sources extra bias current to Q_1. As will be shown in Section 6.3.2.3, Q_1 has large emitter area to reduce its base resistance and associated thermal noise. This leads to increased base-emitter junction capacitance. In turn, this requires higher bias current in order to reduce the transition time through the base and improve high-frequency response. An extra current source is used as opposed to decreasing collector resistance R_C, which would lower the loop gain. Current sink I_{B2} not only biases Q_2, but also absorbs most of the input photo current I_i.

Conventionally, the output would be taken at Q_2's emitter. However, this would not leave any headroom for the tail current bias source in the subsequent stage (Section 6.4). Hence, V_o is located at the collector of Q_0. The non-dominant pole of the loop gain is located at this node (Eq. (6.16)), hence care must be taken to limit the capacitance.

6.3.1 Large-signal Behavior

The TIA is self-biased. Assuming $\beta \approx \infty$ (Section 1.3), the static large-signal behavior is described by the following expressions:

$$V_O = V_{BE0} + V_{BE2} - R_F I_I \tag{6.1}$$

$$I_{C1} = \frac{V_{DD} - V_O}{R_C} \tag{6.2}$$

$$I_{C0} = I_{C1} + I_{B1} \tag{6.3}$$

$$I_{C2} = I_{B2} - I_I \tag{6.4}$$

With $V_{BEi} \approx 850\,\text{mV}$ and the values given in Fig. 6.4, this results in $V_O \approx 1.7\,\text{V}$ when $I_I = 0$. This is the maximum output voltage under incident current, as the photo current is unipolar. Collector resistor R_C determines, via feedback, the bias current of input device Q_1, hence the term *self-biased.*

When expressing[2] $V_{BEi} \approx U_t \ln(I_{Ci}/I_{s0})$, it is possible to solve Eqs. (6.1) to (6.4) and find a closed-form solution for V_O that involves LambertW functions [92]. Figure 6.5 graphically shows the result. I_{B2} determines the maximum input current, as predicted by Eq. (6.4). In addition, V_O decreases quite rapidly with increasing input current. This poses a problem for the tail current source of the subsequent stage. The solution for these problems is discussed in Section 6.3.4. The input/output curve is reasonably linear for most of the input current range, which is due to the negative feedback. Its slope, the transresistance, is ideally $-(R_F + 1/g_{m2})$, as will be shown in the following section.

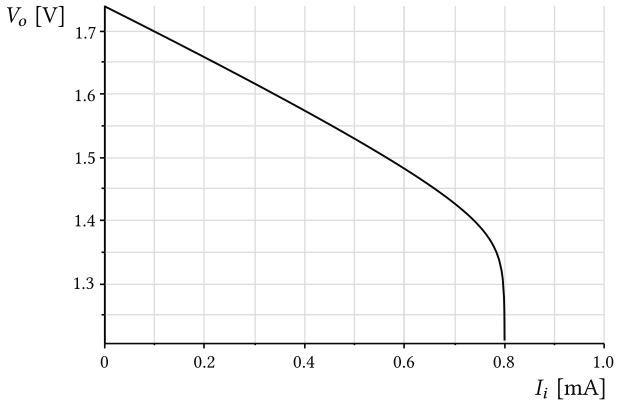


Figure 6.5: Static input/output curve of the TIA ($I_{s0} = 6.4 \times 10^{-18}$ A, $I_{B1} = 2\,\text{mA}$, $I_{B2} = 0.8\,\text{mA}$).

6.3.2 Linear Model

In order to obtain design equations, a model based on the linearized circuit shown in Fig. 6.6 is derived. The devices Q_1 and Q_2 are replaced by their transconductance. The photo diode is replaced by its linear model (Section 2.6.2), ignoring its ohmic resistance for now. C_i is the photo diode capacitance, while C_1 is the remaining capacitance at the input

[2] Ignoring the unequal emitter area of Q_1 and Q_2.

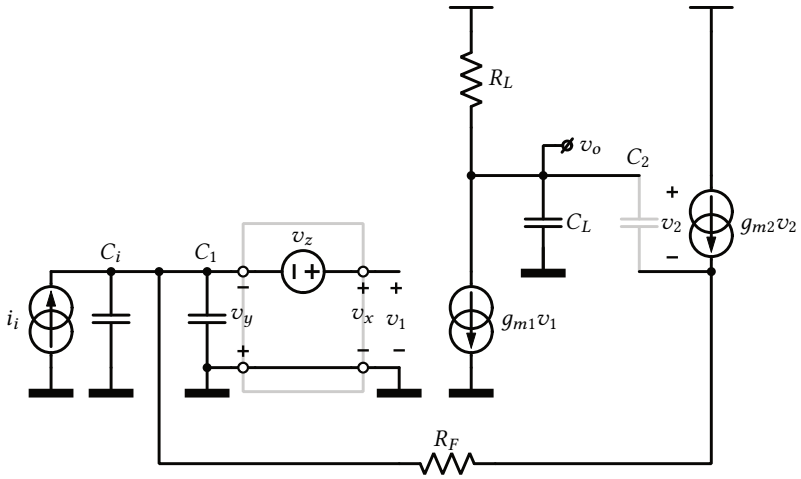


Figure 6.6: Linearized equivalent circuit of the TIA.

node (input and Miller capacitance of Q_0, ESD-protection and wire parasitics). Cascode Q_0 is ignored. C_L and R_L represent the total impedance to ground at the output. Initially, the input capacitance of Q_2, C_2, will be included.

As the circuit employs negative feedback, it is natural to apply the general feedback theorem (GFT) analysis (Appendix A). To avoid the more complex 2-GFT in this analytical derivation, the circuit is further simplified by including the Miller-multiplied collector-base capacitance of Q_1 in C_1. This ignores a feedforward zero, but—as the device is now unilateral—reverse loop gain vanishes. In effect, an ideal injection point is created intentionally such that the 1-GFT can be applied.

The input/output transfer function, input impedance and output impedance will be determined next.

6.3.2.1 Input/output Transfer Function

The input-output transfer function $H = v_o/i_i$ is decomposed as:

$$H = \frac{v_o}{i_i} = H_\infty D D_\mathrm{n} = H_\infty \frac{1 + \frac{1}{T_\mathrm{n}}}{1 + \frac{1}{T}} \tag{6.5}$$

The constituent terms are derived in the following.

Ideal Transfer Function H_∞ The ideal transfer function H_∞ is a null double injection calculation. It is the transfer function when the loop gain T becomes infinite:

$$H_\infty = \left.\frac{v_o}{i_i}\right|_{v_y=0} \tag{6.6}$$

$$= -\left(R_F + \frac{\frac{1}{g_{m2}}}{1+\frac{sC_2}{g_{m2}}}\right) \tag{6.7}$$

$$= -R_F \frac{1+\frac{\omega_z}{s}}{1+\frac{\omega_p}{s}} \tag{6.8}$$

with

$$\omega_p = \frac{g_{m2}}{C_2} \tag{6.9}$$

$$\omega_z = \omega_p \left(1 + \frac{1}{g_{m2}R_F}\right) \tag{6.10}$$

The input capacitance of Q_2 introduces a pole-zero doublet in H_∞. This has a negative impact on settling time. Consider Eq. (6.7): when a step is applied to i_i, the output rises infinitely fast to $R_F i_i$, followed by an exponential tail eventually settling on $(R_F + 1/g_{m2})\, i_i$. The pole frequency ω_p approaches the transition frequency f_{T2} of Q_2, such that, when the device is appropriately biased,

$$H_\infty \approx -\left(R_F + \frac{1}{g_{m2}}\right) \tag{6.11}$$

$$= -R'_F \tag{6.12}$$

Q_2's emitter area is kept minimal to keep its input capacitance small. A bias current of 800 µA makes f_{T2} sufficiently large. In addition, C_2 is bootstrapped to an even lower value by the local feedback loop of the emitter follower and will be ignored in the following sections.

Loop Gain T and Discrepancy Factor D As the circuit has been intentionally simplified to construct an ideal injection point, the loop gain equals the voltage loop gain. The circuit diagram reveals two poles and no zeros:

$$T = \left.\frac{v_y}{v_x}\right|_{i_i=0} = \frac{T_0}{\left(1+\frac{s}{\omega_1}\right)\left(1+\frac{s}{\omega_2}\right)} \tag{6.13}$$

with:

$$T_0 = g_{m1}R_L \tag{6.14}$$

$$\omega_1 = \frac{1}{R'_F C_T} \tag{6.15}$$

$$\omega_2 = \frac{1}{R_L C_L} \tag{6.16}$$

Input pole frequency ω_1 is determined by $C_T = C_i + C_1$, the total capacitance at the input node, and the feedback resistor. The second pole frequency ω_2 is created at the output node of the common emitter amplifier.

The discrepancy factor $D = T/(1+T)$ can be written as follows, assuming $T_0 \gg 1$:

$$D = \frac{1}{1 + \frac{s}{\omega_n Q} + \frac{s^2}{\omega_n^2}} \tag{6.17}$$

with the natural frequency ω_n the geometric mean of the unity-gain frequency and second pole of T, and the Q-factor given by:

$$\omega_n = \sqrt{T_0 \omega_1 \omega_2} = \sqrt{\frac{g_{m1} R_L}{R'_F C_T R_L C_L}} \tag{6.18}$$

$$Q = \sqrt{\frac{T_0 \omega_1}{\omega_2}} = \sqrt{\frac{g_{m1} R_L R_L C_L}{R'_F C_T}} \tag{6.19}$$

For typical element values, the input pole at ω_1 is dominant as it occurs at much lower frequency than ω_2 in the given topology. To provide sufficient stability margin, the Q-factor of D should be limited. Hence ω_2 must be placed sufficiently above the unity-gain frequency of T.

Null Loop Gain T_n, Null Discrepancy Factor D_n and Direct Transmission H_0 It is easily verified on the circuit diagram (Fig. 6.6) that, as long as $C_2 = 0$ and Q_1 is assumed unilateral, $T_\mathrm{n} = \infty$, $D_\mathrm{n} = 1$ and $H_0 = 0$.

Complete Decomposition It follows from Eqs. (6.5), (6.12) and (6.17) that H is a second order transfer function, with bandwidth and peaking determined by the input and output pole of the loop gain T (Fig. 6.7). From the figure, it is clear that the peaking is only caused by the loop gain.

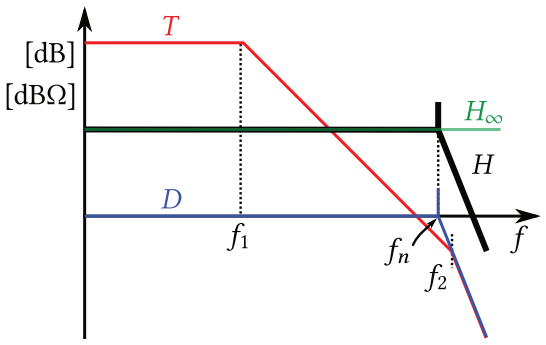


Figure 6.7: Decomposition of the input/output transfer function H.

The 3-dB bandwidth of a second order system can be written as a function of its natural

frequency and Q-factor (Fig. 6.8):

$$f_{3\text{dB}} = f_n \underbrace{\sqrt{1 - \frac{1}{2Q^2} + \sqrt{\frac{1}{4Q^4} - \frac{1}{Q^2} + 2}}}_{\alpha} \tag{6.20}$$

Equivalently, given a fixed gain-bandwidth product (GBW) f_a (*not* unity-gain frequency):

$$f_{3\text{dB}} = f_a \frac{\alpha}{Q} \tag{6.21}$$

When $Q = \sqrt{2}/2$ (Butterworth filter) is imposed, a flat frequency response is obtained with minimal time-domain overshoot (4.3 %). Equivalently, f_2 is placed at two times the GBW of T. In that case, Eq. (6.20) shows that the bandwidth of D, and hence of H, is given by the natural frequency, while Eq. (6.21) shows that the increase in bandwidth reaches a maximum value of $\sqrt{2}$, compared to a single-pole system (or f_2 at infinite frequency).

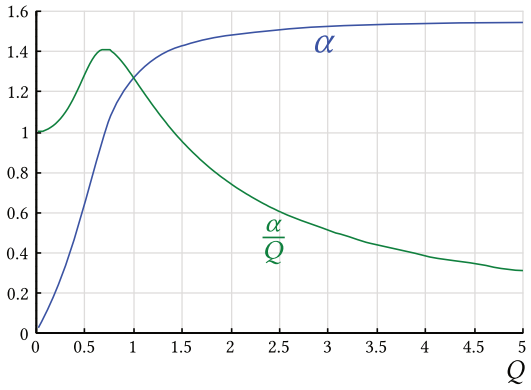


Figure 6.8: Relative 3-dB bandwidth of a second order system given a natural frequency (α) or a fixed gain-bandwidth product (α/Q), as a function of Q-factor.

Even with this bandwidth enhancement, the projected bandwidth of 16.6 GHz, in combination with a high transimpedance and low input-referred noise, could not be reached. Section 6.3.3 describes how inductive peaking is included to reach the design objective.

6.3.2.2 Terminal Impedances

For the sake of completeness, the port impedances will be derived. This is easy as the loop gain has already been established (Eq. (6.13)), and only null double injections calculations remain.

Input Impedance The input impedance Z_i is found by injecting a current in, and measuring the voltage across, the input node. Write the 1-GFT decomposition of Z_i as:

$$Z_i = \frac{v_i}{i_i} = H_\infty \frac{T}{1+T} + \frac{H_0}{1+T} \tag{6.22}$$

$$= \frac{H_0}{1+T} \tag{6.23}$$

Equation (6.23) holds as nulling v_y to find the ideal transfer function reveals $H_\infty = 0$. Indeed, ideally, the TIA is a perfect current sink. Equation (6.23) reveals the widely known fact that negative feedback (with series injection) reduces the input impedance[3] [93]. H_0 is easily found on the circuit as the parallel combination of C_T and R'_F. Hence:

$$H_0 = \left.\frac{v_i}{i_i}\right|_{v_x=0} = \frac{R'_F}{1+sC_TR'_F} \tag{6.24}$$

Output Impedance Similarly, the output impedance Z_0 is computed by injecting a current in the output node and measuring the resulting voltage. Z_0 is decomposed as:

$$Z_0 = \frac{v_o}{i_o} = H_\infty \frac{T}{1+T} + \frac{H_0}{1+T} \tag{6.25}$$

$$= \frac{H_0}{1+T} \tag{6.26}$$

Once again, $H_\infty = 0$: when v_y is nulled, all current flow in C_T, R_F and g_{m2} is impeded. Thus v_2 and v_o must be 0. Again, this confirms that negative feedback (parallel sensing) reduces output impedance. Ideally, the TIA behaves as a voltage source. H_0 is the parallel combination of C_L and R_L:

$$H_0 = \left.\frac{v_o}{i_o}\right|_{v_x=0} = \frac{R_L}{1+sC_LR_L} \tag{6.27}$$

6.3.2.3 Input-referred Noise

The sensitivity of the receiver is partly determined by the input-referred noise of the amplifier chain (Section 2.4). It is instructive to write the input-referred noise as a function of circuit noise generators. The general method of noise calculation will be applied [85]: the output contribution for each noise generator is calculated and mean-square summed to obtain the output voltage noise power spectral density (PSD). The input-referred rms current noise is obtained by integrating the output voltage noise PSD over the applicable 'system bandwidth', taking the square root (thus obtaining the output-referred rms noise voltage) and dividing by the 'midband gain' of the input/output transfer function. The noise generators are assumed statistically independent. Furthermore, only the input-referred noise current is considered as the source impedance is high (current source) such that (conceptually) input-referred voltage noise is highly attenuated. An other way of computing input-referred noise involves integration of the input-referred noise *spectrum* in which noise bandwidths are chosen such that the result is identical with the general approach [18].

It can be shown that the bulk of the circuit noise is generated by: shot noise due to Q_0's collector current and thermal noise of the resistors R_F, R_L, R_S and r_B, the feedback resistor, collector load resistor, source resistance (Section 2.6.2) and ohmic base resistance of Q_0. A number of secondary noise sources exist that can be ignored: among others the base current shot noise (high β, see Section 1.3), noise generators of cascode device Q_0 and emitter follower Q_2, parasitic ohmic resistances in bipolars, flicker noise and excess noise in resistors or active devices and drain-referred thermal noise of the current source

[3] In fact, many well-known results derive cleanly from the GNT.

I_{B1} (operates in very strong inversion for reduced g_m). Consider the PSD of the noise generators[4]:

$$\frac{\overline{i_{nR_L}^2}}{\Delta(f)} = \frac{4k\mathsf{T}}{R_L} \tag{6.28}$$

$$\frac{\overline{i_{nR_F}^2}}{\Delta(f)} = \frac{4k\mathsf{T}}{R_F} \tag{6.29}$$

$$\frac{\overline{v_{nR_S}^2}}{\Delta(f)} = 4k\mathsf{T}R_S \tag{6.30}$$

$$\frac{\overline{v_{nr_B}^2}}{\Delta(f)} = 4k\mathsf{T}r_B \tag{6.31}$$

$$\frac{\overline{i_{nI_{C1}}^2}}{\Delta(f)} = 2qI_{C1} \tag{6.32}$$

in which k (1.38×10^{-23} J/K) and q (1.6×10^{-19} C) are the Boltzmann constant and elementary charge, respectively. The thermal noise current (voltage) PSD's are inserted in parallel to (in series with) the respective resistors, while the collector shot noise current PSD is added in parallel with g_{m1}. In a similar fashion as in the previous sections, the (power) transfer function from each source to the output is computed and mean-square summed. The output voltage noise PSD is:

$$\begin{aligned}\frac{\overline{v_{on}^2}}{\Delta(f)} \approx{}& \left(\frac{4k\mathsf{T}}{R_L} + 2qI_{C1}\right) \frac{R_L}{1+(2\pi f R_L C_L)^2} \left|\frac{1}{1+T}\right|^2 + \frac{4k\mathsf{T}}{R_F}\left|\frac{R_F T}{1+T}\right|^2 \\ &+ 4k\mathsf{T}R_S \frac{R_F'}{R_S} \frac{1}{1+\frac{1}{(2\pi f C_i R_S)^2}} \left|\frac{T}{1+T}\right|^2 + 4k\mathsf{T}r_B\left(1+(2\pi f C_i R_F')^2\right)\left|\frac{T}{1+T}\right|^2 + \frac{\overline{v_{n,MA}^2}}{\Delta(f)}\end{aligned} \tag{6.33}$$

in which the last term is the input-referred voltage noise of the MA and $T = T(j\omega)$ the loop gain, evaluated on the frequency axis. The modifications to the loop gain, expressed by Eq. (6.13), by non-zero R_S and r_B are negligible. The midband gain is approximately $-R_F'$ (Eq. (6.12)). Integrating Eq. (6.33), dividing by the (squared) midband gain and taking the square root yields the input-referred rms noise current.

$$i_n^{\text{rms}} \approx \frac{1}{R_F'}\left[\int_0^\infty \frac{\overline{v_{on}^2}}{\Delta(f)}\,\mathrm{d}f\right]^{\frac{1}{2}} \tag{6.34}$$

$$\approx \left[\int_0^\infty \frac{1}{R_F'^2}\frac{\overline{v_{on}^2}}{\Delta(f)}\,\mathrm{d}f\right]^{\frac{1}{2}} \tag{6.35}$$

A number of device parameter trade-offs can be exposed by considering Eq. (6.33) for low

[4]This section uses the symbol T for absolute temperature in order to differentiate from the loop gain T.

frequencies. Then $T \approx g_{m1}R_L$ and:

$$\frac{\overline{v_{on}^2}}{\Delta(f)} \approx \left(\frac{4k\mathsf{T}}{R_L} + 2qI_{C1}\right)\frac{1}{g_{m1}^2} + 4k\mathsf{T}R_F + \frac{4k\mathsf{T}}{R_S}R_F'^2\left|\frac{1}{1+\frac{1}{(2\pi f C_i R_S)^2}}\right|^2 + 4k\mathsf{T}r_B + \frac{\overline{v_{n,MA}^2}}{\Delta(f)} \tag{6.36}$$

Equations (6.35) and (6.36) show that input-referred current noise is optimized for high R_F: low generated noise, high gain, but lower bandwidth (Eq. (6.15)); high R_L: low generated noise, but increased peaking (Eq. (6.16)); small R_S: low generated noise; small r_B: low generated noise; high I_{C1}: although generated shot noise increases, the squared transconductance induces a net noise decrease. r_B small implies a big input transistor Q_1, which leads to increased transit time through the base and base-emitter capacitance (Eq. (6.15)). This is countered by the extra current provided by I_{B1} (Fig. 6.4). The larger I_{C1} is also beneficial for noise. I_{B1} is implemented with a PMOS current mirror operating in very strong inversion (low g_m) for optimally low drain-referred noise current.

6.3.3 Bandwidth Enhancement Through Inductive Peaking

Inductive peaking is a well-known technique to enhance the bandwidth of circuits [18, 94]. One could add an inductance in series with R_C to resonate out the load capacitance or even include T-coils [95]. In this design however, on-chip inductors or transformers are precluded as channel area is limited (Section 6.1). Instead, the self-inductance of the bondwires is employed to increase the bandwidth of the TIA input stage.

In this section the implications of bondwire inductance L_{bw} on the linear model developed in Section 6.3.2 and in particular the impact on stability will be studied. The bondwire resistance and ohmic base resistance of Q_1 will be lumped together as R_{bw}. Although not rigorously correct, it does make the analysis simpler, while still obtaining a useful result. Intuitively, the bondwire inductance and input capacitance constitute a resonant tank, which is damped by R_{bw}. If the frequency, where both the inductive and capacitive reactances cancel, is well-chosen, peaking occurs which results in bandwidth enhancement.

Mutual coupling between the bondwires of the anode and cathode, which carry current in opposite direction, reduces the effective loop inductance. This effect will be ignored and only the total self-inductance of both bondwires will be taken into account.

One can try to predict the bondwire self-inductance using a 3D field solver. However, the accuracy of the result depends on the validity of the 3D structural model. Furthermore, the length of the bondwire has a major impact on the self-inductance and is unknown a priori. For these reasons, the circuit will be checked for total bondwire lengths shorter than about 1 mm and the MA stages will be made digitally tunable to deal with this uncertainty (Section 6.4). The following rule of thumb will be adopted: the self-inductance of a bondwire is approximately 0.8 nH/mm [48]. This results in $L_{\text{bw}} \leq 1\,\text{nH}$.

Let's consider the bondwire inductance and the ohmic series resistance as an extra element Z in the circuit (Fig. 6.9):

$$Z = sL_{\text{bw}} + R_{\text{bw}} \tag{6.37}$$

The extra element theorem (EET) will be applied to the GFT decomposition of the TIA, with $Z = 0$ as reference impedance.

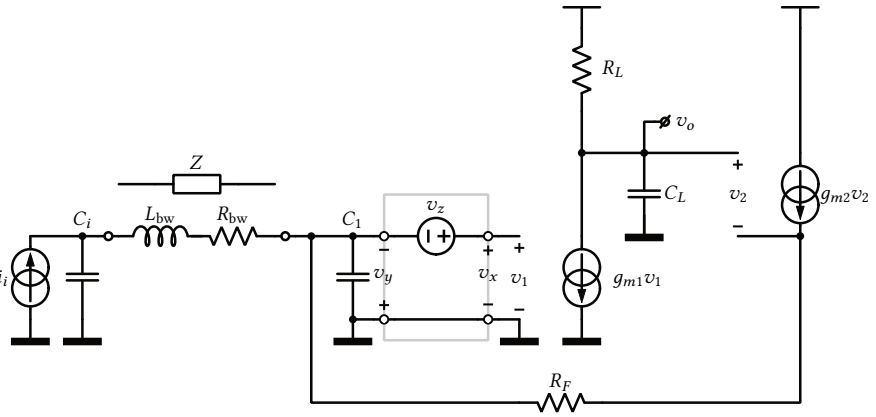


Figure 6.9: Bondwire inductance L_{bw} and resistance R_{bw} as extra element.

The following device parameters are used for the plots of the analytical expressions; they have been determined from the operating point of the complete circuit: $R_F = 390\,\Omega$, $R_C = 70\,\Omega$, $L_{\mathrm{bw}} = 600\,\mathrm{pH}$, $R_{\mathrm{bw}} = 7\,\Omega$, $C_i \approx 115\,\mathrm{fF}$, $C_1 \approx 260\,\mathrm{fF}$, $C_L \approx 40\,\mathrm{fF}$, $g_{m1} \approx 220\,\mathrm{mS}$, $g_{m2} \approx 26\,\mathrm{mS}$. C_i is the photo diode capacitance. C_1 includes the input and Miller capacitance of Q_1 and the capacitance of the bondpad ESD-protection diodes.

Impact of bondwire inductance on H_∞ The EET decomposition of H_∞ (condition $v_y = 0$) is written as:

$$H_\infty = H_{\infty,\mathrm{ref}} \frac{1 + \frac{Z}{Z_\mathrm{n}}}{1 + \frac{Z}{Z_\mathrm{d}}} \tag{6.38}$$

$H_{\infty,\mathrm{ref}}$ is already calculated (Eq. (6.12)). Setting $i_i = 0$, the driving point impedance is easily seen to be $Z_\mathrm{d} = 1/(sC_i)$. Likewise, setting $v_o = 0$ impedes any current from flowing in R_F, thus removing it from the circuit. Hence, the null driving point impedance $Z_\mathrm{n} = \infty$ and Eq. (6.38) can be written as:

$$H_\infty = \frac{H_{\infty,\mathrm{ref}}}{1 + \frac{s}{\omega_{n0}Q} + \left(\frac{s}{\omega_{n0}}\right)^2} \tag{6.39}$$

$$\omega_{n0} = \frac{1}{\sqrt{L_{\mathrm{bw}} C_i}} \tag{6.40}$$

$$Q = \frac{1}{R_{\mathrm{bw}}} \sqrt{\frac{L_{\mathrm{bw}}}{C_i}} \tag{6.41}$$

With the extra element included, H_∞ now represents a second order system. These expressions are depicted in Fig. 6.10(a).

Impact of bondwire inductance on T Likewise, the loop gain T (condition $i_i = 0$) can be decomposed as:

$$T = T_{\mathrm{ref}} \frac{1 + \frac{Z}{Z_\mathrm{n}}}{1 + \frac{Z}{Z_\mathrm{d}}} \tag{6.42}$$

T_{ref} has already been computed (Eq. (6.13)). For the driving point impedance, from $v_x = 0$ follows $v_o = 0$. Hence:

$$Z_{\text{d}} = \frac{1}{sC_i} + \left(\frac{1}{sC_1} \left\| R'_F\right.\right) \tag{6.43}$$

which can be written as:

$$Z_{\text{d}} = \frac{1}{sC_{\text{eq}}} \frac{1 + \frac{\omega_z}{s}}{1 + \frac{\omega_p}{s}} \tag{6.44}$$

with:

$$C_{\text{eq}} = \frac{C_i C_1}{C_i + C_1} \tag{6.45}$$

$$\omega_z = \frac{1}{R'_F(C_i + C_1)} \tag{6.46}$$

$$\omega_p = \frac{1}{R'_F C_1} \tag{6.47}$$

In a typical design, the pole and zero are reasonable close together and will be ignored:

$$Z_{\text{d}} \approx 1/C_{\text{eq}} \tag{6.48}$$

The null driving point impedance is calculated by nulling the output of the transfer function under consideration (T, hence $v_y = 0$). The only element left in the circuit is C_i, such that $Z_{\text{n}} = 1/(sC_i)$. In the resulting decomposition, both a complex pole-pair and a complex-zero pair appear:

$$T = T_{\text{ref}} \frac{\left(1 + \frac{s}{\omega_{nz} Q_z} + \frac{s^2}{\omega_{nz}^2}\right)}{\left(1 + \frac{s}{\omega_{np} Q_p} + \frac{s^2}{\omega_{np}^2}\right)} \tag{6.49}$$

with:

$$\omega_{nz} = \frac{1}{\sqrt{L_{\text{bw}} C_i}} \tag{6.50}$$

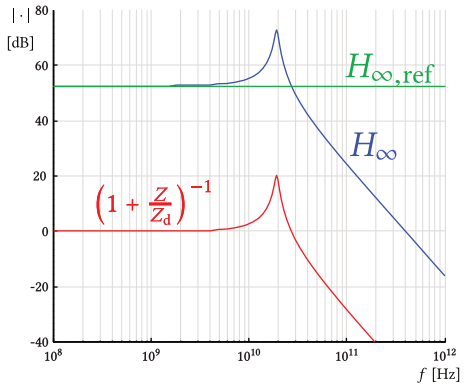
$$Q_z = \frac{1}{R_{\text{bw}}} \sqrt{\frac{L_{\text{bw}}}{C_i}} \tag{6.51}$$

$$\omega_{np} = \frac{1}{\sqrt{L_{\text{bw}} C_{\text{eq}}}} \tag{6.52}$$

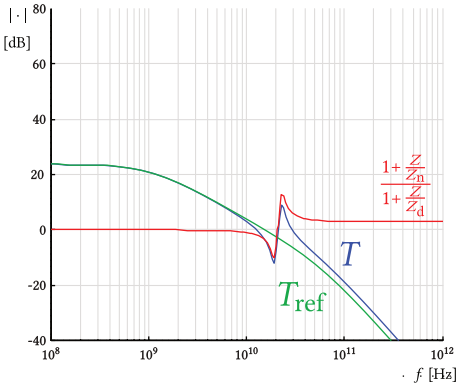
$$Q_p = \frac{1}{R_{\text{bw}}} \sqrt{\frac{L_{\text{bw}}}{C_{\text{eq}}}} \tag{6.53}$$

Two adjacent resonant peaks, one downward and one upward, are introduced in the loop gain, with $\omega_{nz} < \omega_{np}$, as shown in Fig. 6.10(b). The downward peak coincides with the upward peak in H_∞: $\omega_{nz} = \omega_{n0}$.

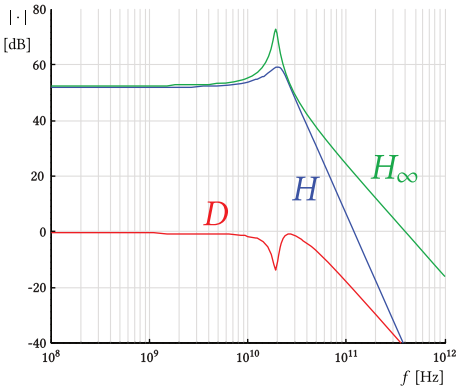
Stability Stability is checked on the Nyquist diagram of T (Fig. 6.10(d)). As the frequency increases, a severe phase lead and lag is introduced in the phase characteristic between the resonant frequencies ω_{nz} and ω_{np}. The number of encirclements around the critical point $(-1, 0)$ is zero and equals the difference between the number of zeros and number of poles of the return difference $1+T$ in the left hand-side plane (LHP) [96]. If one assumes that the circuit that remains, when the feedback loop is killed is stable, then the original circuit is stable. We conclude that the bondwire inductance does *not* significantly affect stability of the system, as no excess phase lag is introduced and the original phase margin is barely changed.



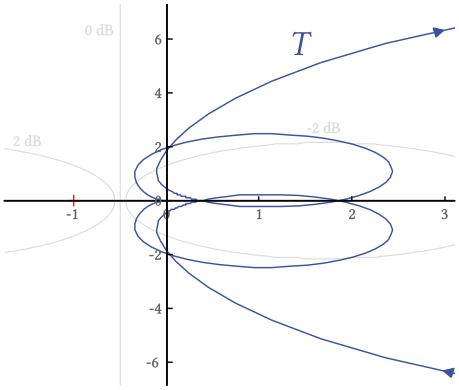
(a) Impact on $H_\infty \approx H_{\infty,\mathrm{ref}}\,(1+Z/Z_\mathrm{d})^{-1}$.



(b) Impact on $T \approx T_\mathrm{ref}\,(1+Z/Z_\mathrm{d})^{-1}(1+Z/Z_\mathrm{n})$.



(c) Impact on $H \approx H_\infty D$.



(d) Nyquist diagram of T.

Figure 6.10: Impact of bondwire inductance on the decomposition of the transfer function of the TIA and Nyquist diagram of the loop gain.

Impact of bondwire inductance on D and H The impact of L_bw on the discrepancy factor D of the GFT decomposition, through T, can also be calculated analytically. However, the resulting expressions are too complicated to be very useful for design purposes. Figure 6.10(c) shows that the downwards peak in T at ω_{nz} also appears in D and, as mentioned above, partially cancels the upwards peak in H_∞ at ω_{n0}. The resulting H exhibits

maximum peaking between ω_{nz} and ω_{np}, before rolling off with a third order slope. Note that the peaking is *not* caused by insufficient phase margin of T.

It is empirically established, that as long as $Q \gg 1$ (Eq. (6.41)), an upper bound on the 3-dB bandwidth of H is given by the 3-dB bandwidth of H_∞. From Eq. (6.20) and Fig. 6.8 follows:

$$f_{3\text{dB},H} < f_{3\text{dB},H_\infty} \approx \frac{1}{2\pi}\sqrt{\frac{1+\sqrt{2}}{L_{\text{bw}}C_i}} \tag{6.54}$$

Equation (6.54) is used as a rough estimation for the obtainable bandwidth enhancement.

Simulation results

Spectre simulations for L_{bw} between 0 nH and 1 nH confirm Eq. (6.54), for $L_{\text{bw}} \geq 400$ pH (Fig. 6.11). This minimum inductance is easily obtained in practical bondwires. For lower inductances, the peaking disappears and the bandwidth drops to approximately its single-pole value (f_a in Eq. (6.21)). For higher inductances, the bandwidth drops again as the resonant frequency decreases. The phase margin is always higher than 58°.

For the given component values, $Q \approx 10.3$ and Eq. (6.54) yields 29 GHz. Simulations without extracted layout parasitics give 25.6 GHz. With parasitics included, the bandwidth drops to 20 GHz. This illustrates that layout has a profound impact on performance of high speed circuits and should be taken into account early on in the design flow.

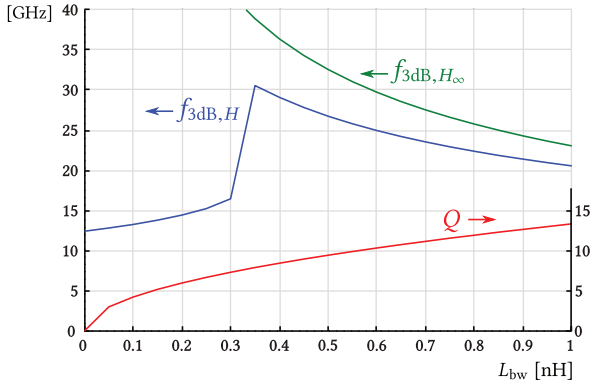


Figure 6.11: Spectre simulation results of the bandwidth (and its upper bound) of the TIA as a function of bondwire inductance.

6.3.4 Extending Dynamic Range By Adaptive Biasing

The dynamic range of the TIA is defined as the input current range where a given specified bit-error ratio (BER) is achieved [18] (Section 2.4). It is lower bounded by the sensitivity limit and upper bounded by the overload limit. The sensitivity limit is partly governed by the input-referred noise current and plays a role at low input currents. To a good degree, it can be assessed with small-signal approximations. The overload limit is affected by a multitude of non-linear effects (e.g. pulse-width distortion, jitter, asymmetric clipping) which are the dominant effects at large input currents. Often, horizontal eye closure is

observed. These effects are not captured by small-signal analysis and hard to characterize in a comprehensive analytical fashion.

The following qualitative treatment describes three phenomena that reduce the overload limit in the topology of Fig. 6.4.

1. The output voltage V_o pushes bipolar cascode device Q_0 in saturation for large negative swings. In saturation, both the base-emitter and base-collector diode are forward-biased. The base depletion region is filled with minority carriers. When the swing reverses sign, it takes a certain amount of time to remove the charge. This creates a tail in the output voltage, introducing ISI. This is resolved by choosing a suitable cascode bias voltage V_B.

2. The output voltage V_o clips on the upper supply rail for large positive swings. This cuts off cascode device Q_0. The feedback loop is effectively killed and the instantaneous transimpedance and closed-loop bandwidth collapse. When the bias voltage across R_L is adequately high, this effect occurs outside of the specified input range.

3. Emitter follower Q_2 goes into cut-off as soon as the input current exceeds the current-sinking capability of the output stage of the TIA, also killing the feedback loop. Indeed, the instantaneous collector current $I_{c2}(t) \approx I_{B2} - I_i(t)$ ($\beta \gg 1$).

The third effect can be mitigated by making the bias current of the emitter follower I_{B2} higher than the specified input current range. However, this unnecessarily increases power consumption also at low input currents. Therefore the emitter follower will be dynamically biased by adding the measured dc input current to I_{B2}.

In addition, without any measures, the unipolar nature of the input current causes the average output voltage of the TIA to decrease with increasing input current (Fig. 6.5). This is undesirable, as the tail current of the subsequent S2D-stage (Section 6.4) must remain in the active region for high output resistance and to limit base current of the current source devices, which could upset the bipolar bias current mirror. This is resolved by preventing that dc current flows in feedback resistor R_F. Ac coupling is ruled out because of area reasons. Instead the measured input dc current will be subtracted from the total input current at the TIA input. Care is required to minimize the extra parasitics. This can be regarded as a form of feedforward offset compensation [18]. Figure 6.12 illustrates the concept. $\overline{I_i}$ is the dc content or average of the input current.

Principle of Operation

The average photo current is measured at the cathode terminal of the photo diode. A typical PMOS current mirror comes to mind. However, this approach has a number of problems:

- In order to limit the voltage drop across the diode-connected transistor, it should be biased closed to weak inversion at the maximum expected photo current. This results in very big transistors.
- The cathode of the photo diode should be biased closed to 3.3 V. The breakdown voltage of the 'high-voltage' PMOS devices is 2.5 V (Section 1.3). This means that extra care should be taken to protect the transistors.
- A non-cascoded current mirror might have too much systematic offset due to different drain-source voltage. Cascoding is not possible to limit the voltage drop.

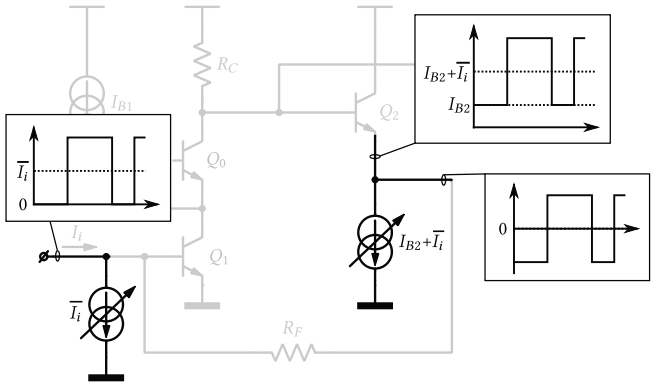


Figure 6.12: Adaptive biasing and feedforward offset compensation added to the TIA.

To avoid these issues, a solution is devised using two matched resistors across which the same voltage is imposed. When the resistors are implemented in polysilicon, which are deposed on top of thick field oxide (in contrast with the very thin field oxide under a MOS transistor gate), much higher electrical field strengths can be tolerated. This yields a much larger breakdown voltage between resistor and substrate.

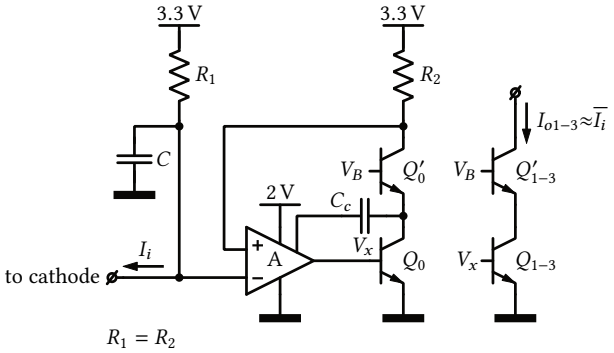


Figure 6.13: Simplified circuit to measure the average photo current.

Figure 6.13 shows the simplified circuit diagram. The cathode of the photo diode is connected to the 3.3 V supply via the decoupling filter formed by R_1 and C. The voltage drop across R_1 is imposed across R_2. If $R_1 = R_2$, equal currents flow in both resistors. The feedback loop is essentially a two-stage amplifier: single-stage opamp A drives the common-emitter amplifier Q_0, R_2. Three nominally identical output currents are provided duplicating the common-emitter stage: one to monitor the average photo current and two to adaptively bias the TIA (Fig. 6.12). The cascodes Q_0', Q_{1-3}' protect the rail devices from too high collector-emitter voltages and reduce the systematic mismatch between the output currents (as well as increase the output resistance). The feedback loop is Miller-compensated.

Photo current I_i is filtered twice to obtain $\overline{I_i}$: once by the decoupling filter and once by the closed-loop circuit. Capacitance C is mainly included to provide a low-impedance source for the cathode. It is implemented as a metal-insulator-metal (MIM) device, placed

on top of the circuit, with a shield, to limit silicon area. The value of R_{1-2} is determined as a trade-off between voltage drop and filter pole frequency on one hand, and offset of the loop on the other hand. The values were chosen as follows: $R_{1-2} = 150\,\Omega$, $C = 8\,\text{pF}$. This limits the voltage drop to 450 mV at an input current of 3 mA.

The main filtering action is provided by the limited bandwidth of the closed loop, which equals the unity-gain frequency of the loop gain, $f_{0\text{dB}} \approx g_m/C_c$, in which g_m is the transconductance of the opamp. Note that, as the current in the common-emitter stage is equal to the dc photo current, the dc loop gain is somewhat signal-dependent. It follows that the accuracy is degraded at low input currents. This is however not a problem for the intended application.

Internally, the opamp consists of a bipolar NPN differential pair with NMOS tail current source and low-swing cascoded PMOS mirror load. Bipolar transistor Q_0 is isolated from the input stage with an NMOS source follower. The high-side voltage sensing, up to 3.3 V, demands for level-shifting at the inputs of the differential pair to protect the active input devices. This is implemented with diode-connected bipolar transistor strings.

6.4 Main Amplifier and Output Buffer

The task of the main amplifier (MA) (together with the output buffer) is to amplify and limit the output voltage of the TIA such that, at the output of the receiver, a voltage with a well-defined output amplitude appears (for input signals above sensitivity levels). This allows subsequent circuitry (e.g. clock-and-data recovery (CDR)) to process the receiver's output without further amplification. The specified target is 300 mV differential peak-to-peak, but the output buffer is designed with some extra margin for 400 mV, which is a typical value in current-mode logic (CML).

The MA consists of a single-ended to differential (S2D) stage and three gain stages. The S2D stage is very similar to a gain stage and will not be discussed separately. The stages have high bandwidth, such that the performance of the entire datapath is limited by the TIA input stage. To provide a high common-mode interference rejection ratio (to limit, a.o. crosstalk, see Section 2.5.4), they are implemented fully differentially. The stages are dc coupled and programmable (see Section 6.3.3). Noise is less important, as the TIA dominates input-referred current noise (Section 6.3.2.3).

Each stage of the main amplifier is identical, but can be separately tuned using dedicated digital control signals. Usually, simple resistively loaded differential pairs with emitter followers are sufficient for this task. However, the available supply voltage of 2 V (Section 6.1) is too low to allow for an extra base-emitter drop (around 900 mV) while keeping the tail current sources of the subsequent stage inside the active region. The resulting drop in output impedance of the tail current source would harm common-mode rejection and increase mismatch in the bias current mirror. Nonetheless, it is possible to retain the beneficial effect of the emitter follower's local negative feedback—a low output impedance—by employing (a variation of) a Cherry-Hooper stage [97]. In this way, decoupling between the stages is obtained (at least for frequencies where the local loop gain is high).

The simplified circuit diagram of a MA stage is depicted in Fig. 6.14. A transimpedance amplifier Q_2, R_F, R_L is the load of a transconductor Q_1. It presents a low output impedance to the following stage. In addition, it also presents a low input impedance to the transconductor. The lack of high-impedance nodes allows for high bandwidth, while still reasonably high gain is obtained. Bias sources I_{B3} source a part of Q_1's bias current I_{B1}, limiting

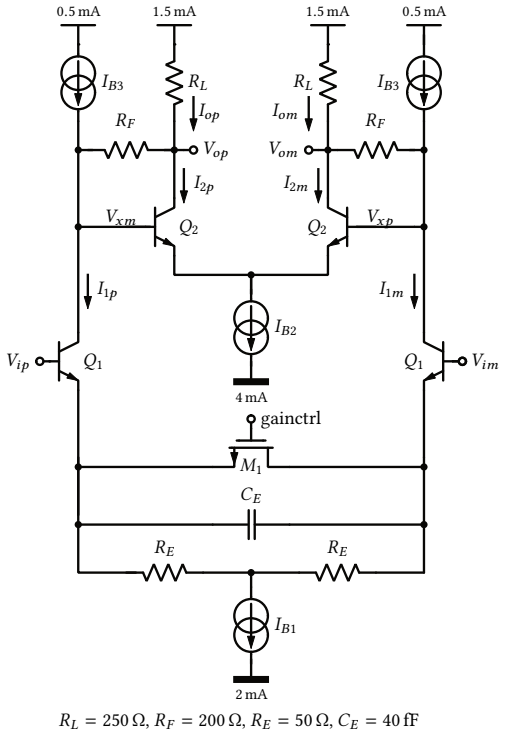


Figure 6.14: Simplified circuit diagram of a main amplifier stage.

the current through R_L. This avoids saturating Q_2 and keeps the dc level of the output nodes high enough to allow dc coupling. Fixed resistive and capacitive emitter degeneration (R_E and C_E) is used to create peaking in the input-output transfer function, increasing the bandwidth. Furthermore, NMOS transistor M_1 acts as a variable resistance that changes the low-frequency degeneration and hence programs the gain (traded with bandwidth) of the stage.

It is instructive to first isolate and analyze the TIA load. Only the low-frequency behavior will be discussed. High-frequency behavior is similar to the model of the receiver input TIA stage developed in Section 6.3.2.

Analysis of the Single-transistor Shunt-Feedback TIA

Consider a resistor in the feedback path of a single transistor resistively loaded amplifier stage. The low-frequency small-signal equivalent circuit diagram is shown in Fig. 6.15. The transistor is considered unilateral with infinitely high input impedance. This simple three-element circuit looks deceptively simple. However, the lack of any buffers introduces coupling between all elements. The GFT analyzes this circuit in an elegant fashion; the single test signal injection is indicated in the figure.

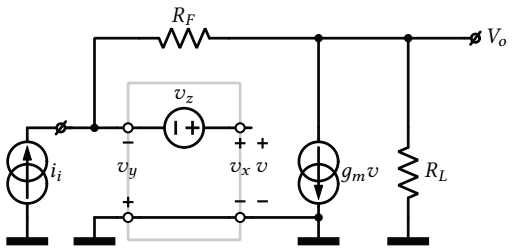


Figure 6.15: Small-signal equivalent low-frequency circuit of a single-transistor shunt-feedback TIA and 1-GFT test signal injection.

Input/Output Transfer Function The given assumptions yield an ideal injection point. The 1-GFT decomposition of the input-output transfer function $H = v_o/i_i$ is:

$$H_\infty = \left.\frac{v_o}{i_i}\right|_{v_y=0} = -R_F \tag{6.55}$$

$$T = \left.\frac{v_y}{v_x}\right|_{i_i=0} = g_m R_L \tag{6.56}$$

$$T_\mathrm{n} = \left.\frac{v_y}{v_x}\right|_{v_o=0} = -g_m R_F \tag{6.57}$$

and

$$H_0 = \frac{H_\infty T}{T_\mathrm{n}} = R_L \tag{6.58}$$

The closed-loop gain follows:

$$H = -R_F \frac{g_m R_L}{1 + g_m R_L} + \frac{R_L}{1 + g_m R_L} \tag{6.59}$$

$$\approx -R_F + \frac{1}{g_m} \tag{6.60}$$

Forward transmission through the feedback resistor produces the second term in Eq. (6.59). Even for high loop gain, the maximal transresistance is not obtained, as Eq. (6.60) shows.

Input and Output Resistance As in Section 6.3.2.2, the input and output resistances are easily calculated as:

$$R_i = \frac{H_0}{1 + T} = \frac{R_F + R_L}{1 + g_m R_L} \tag{6.61}$$

$$R_o = \frac{H_0}{1 + T} = \frac{R_L}{1 + g_m R_L} \approx \frac{1}{g_m} \tag{6.62}$$

in which T is given by Eq. (6.56) and H_0 is calculated in the context of the respective expression.

6.4.1 Large-signal and Small-signal Behavior

It is difficult to obtain an analytical solution for the input-output characteristic of a MA stage, as a set of transcendental expressions appears when writing down the nodal equations. Therefore, a more qualitative discussion will explain the large-signal behavior.

The differential output current is the sum of the (differential) output current of the transconductor and the inverting (differential) output current of the TIA[5]:

$$I_{od} = 0.5\,(I_{op} - I_{om}) \tag{6.63}$$

$$= I_{2d} - I_{1d} \tag{6.64}$$

in which $I_{1d} = 0.5\,(I_{1p} - I_{1m})$ ranges from $-I_{B1}$ to I_{B1} and similar for I_{2d}. I_{B3} is a common-mode current and does not enter I_{od}. The differential output voltage is:

$$V_{od} = 2I_{od}R_L \tag{6.65}$$

The internal differential voltage gain from the input of the transconductor to the input of the TIA load is the input transconductance times the load input impedance. Let $v_{id} = v_{ip} - v_{im}$ and $v_{xd} = v_{xp} - v_{xm}$, then with Eq. (6.61):

$$H_v = \frac{v_{xd}}{v_{id}} = G_{m1}\frac{R_F + R_L}{1 + g_{m2}R_L} \tag{6.66}$$

$G_{m1} \approx g_{m1}/(1+g_{m1}R_E)$ is the effective transconductance of the input stage. When $H_v > 1$ and for increasing input voltage, the output current of the TIA I_{2d} increases faster than I_{1d} and is completely steered to one side before the transconductor is completely switched. Hence I_{od} initially increases until I_{2d} saturates to I_{B2}, then decreases again until I_{B1} is completely steered to one side. Clearly this inversion is undesirable behavior. It can be avoided by ensuring $H_v \leq 1$ and $I_{B2} \geq 2I_{B1}$ [98]. The first condition guarantees that, for increasing input voltage, I_{1d} saturates first (to I_{B1}) before I_{2d} does. The second condition makes sure that I_{od} never changes polarity (Eq. (6.64)). This is illustrated with the dc transfer characteristics of the MA stage, shown in Fig. 6.16 for different values of the ratio I_{B2}/I_{B1}. Under those conditions, the differential peak-peak output swing is:

$$V_{od} = 2(I_{B2} - I_{B1})R_C \tag{6.67}$$

The low-frequency differential voltage gain of the stage can be written as the effective input transconductance times the transresistance of the TIA. Using Eq. (6.60):

$$H = \frac{v_{od}}{v_{id}} \approx G_{m1}\left(R_F - \frac{1}{g_{m2}}\right) \tag{6.68}$$

$$\approx G_{m1}R_F \tag{6.69}$$

in which the approximation holds for $g_{m2}^{-1} \ll R_F$. The gain is independent of I_{B3}.

Ideally, the output common-mode level can be written as:

$$V_{oc} = V_{DD} - R_L\left(\frac{I_{B2}}{2} + \frac{I_{B1}}{2} - I_{B3}\right) \tag{6.70}$$

[5]The following definitions are used [18, 86]: at a port of a differential circuit, the differential voltage and current are defined as $V_d = V_p - V_m$ and $I_d = 0.5(I_p - I_m)$. The common-mode voltage and current are given by $V_c = 0.5(V_p + V_c)$ and $I_c = I_p + I_m$.

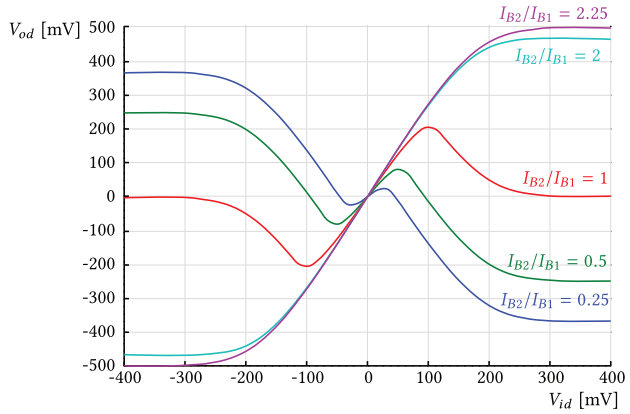


Figure 6.16: Transfer characteristic of a MA stage for $H_v \leq 1$ and different values of the ratio I_{B2}/I_{B1}.

Equations (6.69) and (6.70) show that the common mode output level can be increased independently from the gain, by adding $I_{B3} > 0$, to enable dc coupling of the stages.

In a similar fashion as in Section 6.3.2, a high-frequency model can be developed for the MA stage. This model, however, is more complex as for the receiver TIA as no buffer is present in the TIA load of the MA. In addition, the frequency dependent emitter degeneration of Q_1 introduces frequency peaking in the transfer function, extending the bandwidth. This model will not be discussed further.

6.4.2 Simulation Results

In the final design of the MA stage, $I_{B2} = 2I_{B1}$. Triode MOS transistor M_1 is sized to have a minimum on-resistance of 20 Ω. With the device parameters indicated in Fig. 6.14, H_v ranges between 0.35 and 0.64, while H ranges between 1.8 and 3.4. The gain of the complete MA, including output buffer (Section 6.4.3), is programmable between 8 and 63 times (18 dB to 36 dB), with bandwidth between 21 GHz to 40 GHz.

6.4.3 Output Buffer

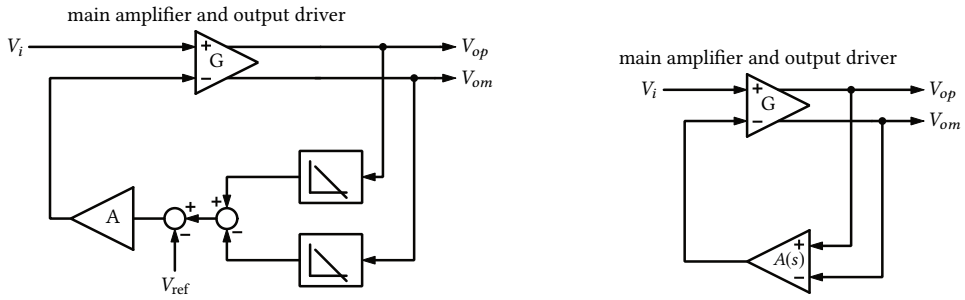
The output driver is a cascoded bipolar differential pair. Internal 50 Ω collector resistors provide termination and are matched to the characteristic impedance of the on-board transmission lines, as in CML stages [99]. The bias current can be programmed between 8 mA and 10 mA for a maximal differential peak-to-peak output swing of 400 mV to 500 mV. The cascodes protect the input devices and, although the voltage gain is not high to begin with, reduce the Miller capacitance somewhat. Frequency-dependent emitter degeneration provides a small amount of peaking to increase the bandwidth of the stage.

6.5 Balancing Loop

As the differential MA is driven from the single-ended TIA, a scheme is required to steer the unused input to the dc voltage needed to balance both output signals. In this work,

as explained in Section 6.2.1, the entire main amplifier including the output driver is enclosed in a feedback loop. This feedback approach also compensates the offset of the main amplifier, in addition to the balancing action. A drawback is that it introduces an inverting pole (high-pass action) in the closed-loop transfer function of the datapath. This poses no problem as long as the pole is placed low enough to limit the power penalty (Section 2.5.3).

6.5.1 Principle of Operation



(a) Principle. (b) Simplification for $V_{\text{ref}} = 0$.

Figure 6.17: Operation of the balancing loop.

Conceptually, the average of both outputs is subtracted and compared to a desired voltage V_{ref}, see Fig. 6.17(a). Amplifier A provides loop gain. Here $V_{\text{ref}} = 0$. As the feedback system is completely linear, it can be simplified as shown in Fig. 6.17(b). Amplifier A is now approximately an integrator $A(s)$ and carries out the comparison, averaging and loop compensation:

$$A(s) = \frac{A_0}{1 + \frac{s}{\omega_A}} \tag{6.71}$$

Consider the graphical construction of the closed-loop gain $H = (v_{op} - v_{om})/v_i = v_{od}/v_i$ in Fig. 6.18. For our intents the transfer function of the MA can be reduced to a simple gain $G_0 = G(0)$ as only low-frequencies are of concern. This is not exactly true, as will be explained in Section 6.5.1.1.

Neglecting direct forward transmission, H can be written as (Appendix A):

$$H(s) = H_\infty \frac{T(s)}{1 + T(s)} \tag{6.72}$$

with:

$$H_\infty(s) = \frac{1}{A(s)} \tag{6.73}$$

$$T(s) = G_0 A(s) \tag{6.74}$$

Clearly, H_∞ represents (almost) an ideal differentiator. This corresponds with the intuitive notion that dc is removed from the output signal v_{od}. An inverted pole is introduced in H

at the gain-bandwidth product of the loop gain via the discrepancy factor D. H becomes:

$$H = G_0 \frac{1 + \frac{\omega_z}{s}}{1 + \frac{\omega_p}{s}} \tag{6.75}$$

$$\omega_z = \omega_A \tag{6.76}$$

$$\omega_p = G_0 A_0 \omega_A \tag{6.77}$$

The inverted pole depends on the gain of the MA. Care is needed to ensure a low enough pole frequency even at maximum gain to limit the power penalty (PP) due to baseline wander (Section 2.5.3).

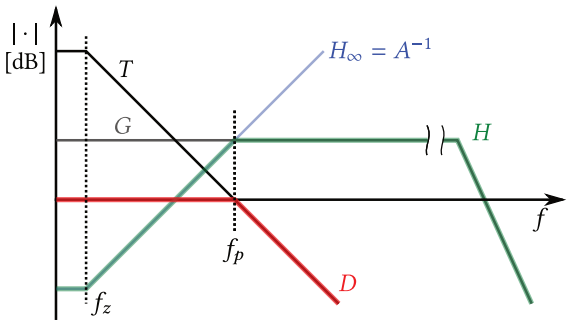


Figure 6.18: Construction of closed-loop transfer function of the balancing loop.

The finite dc loop gain determines the static error: some dc offset between both outputs will remain. In addition, the previous calculation has ignored the offset of A itself. When a small voltage offset v_{OS} is inserted in series with the non-inverting terminal of A and taken as input, the output is easily calculated to be, for infinite loop gain:

$$H'_\infty = \frac{v_{od}}{v_{OS}} = 1 \tag{6.78}$$

Hence, the residual offset between the outputs is lower-bounded by the input-referred offset of the amplifier A.

6.5.1.1 Clipping Nature of the Main Amplifier

It should be noted that the system is inherently non-linear for large enough signals because of the limiting nature of the MA. The impact could be analyzed by applying non-linear analysis techniques such as describing functions, as was done in Section 3.2.3. However, a less rigorous explanation will be given. When the MA starts to clip, the instantaneous loop gain is reduced momentarily. Averaged over time, this results in reduced GBW of the loop and hence lower effective closed-loop pole frequency than predicted by Eq. (6.77). In this application, this effect is not harmful. As long as the resulting loop gain is high enough, small offsets of the MA are still compensated.

6.5.2 Implementation

For a penalty of 0.2 dB based solely on baseline wander, a data rate of 25 Gb/s and a maximum run length of 31 bits, the pole frequency can be calculated (Section 2.5.3) as

$f_p \approx 6\,\text{MHz}$. This would be the value at maximum MA gain of $G_0 = 36\,\text{dB}$. It follows from Eq. (6.77) that the gain-bandwidth product of the amplifier A must be below 100 kHz. This requires either low input-stage transconductance or a high compensation capacitance. The former case is detrimental for input-referred offset and dc loop gain, while the latter is costly in terms of silicon area. This problem is elaborated on in Section 5.4 and Appendix B, where an even slower amplifier is required.

For this optical receiver, a single-stage amplifier with parallel compensation was chosen. The topology is a degenerated bipolar NPN differential pair with MOSFET folded cascode load. The bipolar differential pair is chosen as it has inherently less input-referred voltage offset compared to MOS transistors [29], even though much of the offset has transferred to the degeneration resistors. For this reason those are sized physically big. Additionally, bipolar transistors suffer less from flicker noise (which could be perceived as low-frequency modulated offset) [100].

Compared to a simple mirror load, the folded cascode is more symmetrical and has higher output resistance, which is beneficial for input-referred offset and gain. As is common in a dc-optimized amplifier, the rail devices operate in strong inversion with long gate length, while the cascode devices operate in moderate inversion with smaller gate length as they do not contribute significantly to input-referred offset and gain is still sufficiently high.

An output PMOS follower brings the output range inside the input range of the first stage of the MA (Section 6.4). The follower's bias current source provides sufficient current sourcing capability, even in case the bipolar input transistor of the first stage of the MA would enter saturation. This is needed because as soon as the follower would go into cut-off, the loop would be killed.

Simulation Results

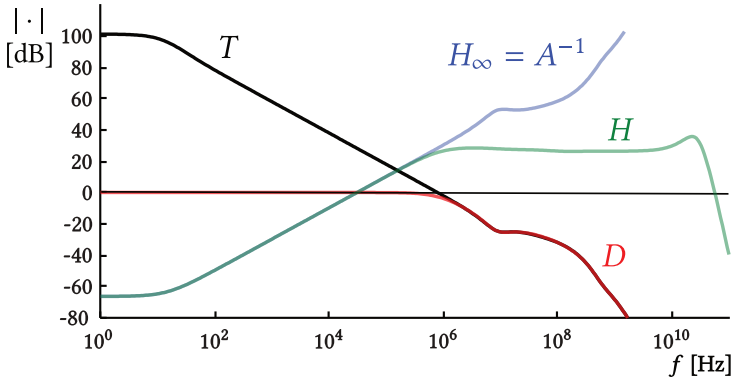


Figure 6.19: Simulated GFT decomposition of the closed-loop transfer function of the MA including balancing loop.

A simulated GFT decomposition of the closed-loop transfer function of the MA, programmed at medium gain, is depicted in Fig. 6.19. This validates the model developed in the previous section. The inverted pole is at approximately 1.2 MHz, while the dc loop gain is over 100 dB.

The 3-sigma input-referred offset voltage (random and systematic) of the amplifier is smaller than 10 mV, a value deemed sufficient for this application.

6.6 Peripheral Circuits

Several important building blocks such as a bandgap reference voltage generator and bias current distribution tree are included in the final chip. In addition, test structures and test signal multiplexers are provided to examine certain internal signals. Finally, a digital SPI interface and several registers allow to program the channels and test functions. These blocks will not be discussed further.

6.7 Experimental Results

The receiver was fabricated in the 130 nm SiGe BiCMOS technology described in Section 1.3. Figure 6.20 shows the die micrograph. Each channel occupies 250 µm × 800 µm (including bond pads), with an additional 210 µm × 310 µm for each balancing loop error amplifier (Section 6.5), located at the left at and right edges of the chip. The compensation capacitors of the balancing loop are clearly visible. The total die size measures 2400 µm × 800 µm. The channel pitch is 250 µm, equal to the photo diode array pitch. Channels 1 and 4 are in the outer lanes, while channels 2 and 3 occupy the middle ones.

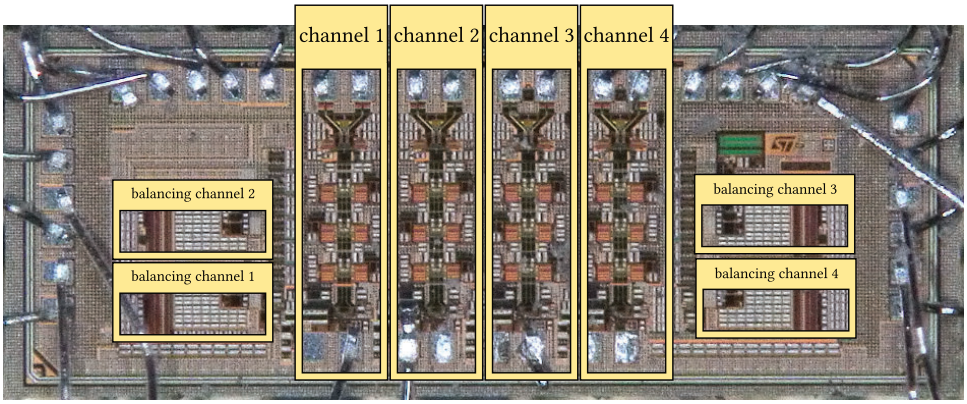


Figure 6.20: Photograph of the optical receiver array.

A detailed photograph of a channel is depicted in Fig. 6.21. Each channel is isolated by deep trenches (Section 2.5.4). The TIA and adaptive biasing circuits are located close to the input bond pads. The gain stages and output driver make up the bulk of the channel and have a highly symmetrical layout to minimize systematic offset and retain high common-mode rejection. Supply decoupling capacitors delineate the datapath.

The naked receiver die was not packaged, but bonded directly to the test board. For the various tests, two types of assemblies were built, with different photo diodes. This was due to delays in manufacturing of the intended photo diode. All outputs are ac coupled to the subsequent instruments.

On average, the receiver consumes 77 mW per channel. On all samples, the dc offset between the outputs of a channel is smaller than 10 mV.

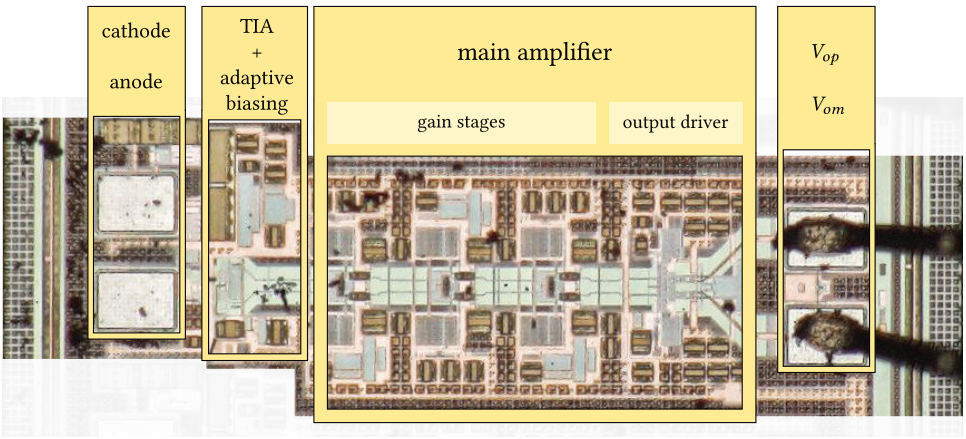


Figure 6.21: Detailed photograph of a channel.

6.7.1 Impact of Bondwire Inductance

An assembly was built without attached fiber array (Fig. 6.22, inset). Manual alignment with a positioner is required to couple incident light. The die is placed in a cavity in order to reduce bondwire inductance, in particular of the ground bondwires. Transmission lines fan out radially from the chip to SMP connectors (Fig. 6.24). The photo diode array is placed at an angle compared to the die. Hence the channels have increasingly longer input bondwires and higher associated bondwire inductance. Channel 1 has the shortest bondwire, while channel 4 has the longest. Note that the bondwires of channel 3 and 4 are crossed, to make the photo diode array compatible with the receiver layout.

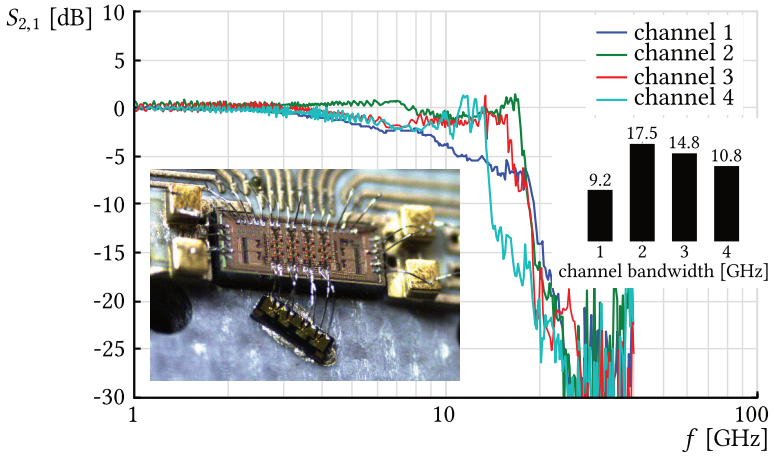


Figure 6.22: Small-signal optical-electrical measurements as a function of bondwire length, normalized to the magnitude at 1 GHz.

Optical-electrical S-parameters measurements were performed on each channel using an Agilent PNA-X network analyzer with a lightwave component analyzer (LCA) module. Figure 6.22 depicts $S_{2,1}$, transmission from optical input to electrical output, for each

channel, normalized to the magnitude at 1 GHz. Although the 3-dB bandwidth is hard to determine exactly due to nature of the measured data, the trends predicted in Section 6.3.3 can be observed, albeit that the absolute value of the bandwidth is lower. This is attributed to: the simple model used for the analysis, impact of output bondwires (which create a low-pass filter), influence of assembly, photo diodes and connectors. For the lowest inductance (channel 1), peaking is not sufficient to obtain sufficient bandwidth. As bondwire inductance increases, the natural frequency shifts down until a sweet spot is hit were maximum bandwidth extension occurs (channel 2). Further increase of bondwire inductance reduces the bandwidth again (channels 3 and 4).

6.7.2 Adaptive Biasing

The adaptive biasing of the TIA (Section 6.3.4) was tested on a similar board as used in Section 6.7.1. PRBS $2^{31} - 1$ NRZ data was applied at increasingly higher optical powers using an erbium doped fiber amplifier (EDFA). The bit rate was limited to 22.5 Gb/s due to limitations of the pattern generator in the test setup at the time of measurement. Figure 6.23 shows the measured differential output eye diagrams without and with adaptive biasing. The differential output amplitude is 400 mV peak-peak. As explained in Section 6.3.4, hard non-linearities in the TIA cause horizontal eye closure and increased jitter. This is clearly observed for the cases $i_{\mathrm{PD}}^{\mathrm{pp}} = 600\,\mu\mathrm{A}$ and $i_{\mathrm{PD}}^{\mathrm{pp}} = 1000\,\mu\mathrm{A}$. For even higher input currents, $i_{\mathrm{PD}}^{\mathrm{pp}} = 2040\,\mu\mathrm{A}$ and $i_{\mathrm{PD}}^{\mathrm{pp}} = 2940\,\mu\mathrm{A}$, the receiver fails to produce an output. This is partly caused by limiting effects in the TIA and partly due to the lower dc voltage at the TIA output, which decreases beyond the output range of the balancing loop error amplifier (Section 6.5). As a result, the loop gain of the balancing control loop drops and the datapath becomes permanently unbalanced.

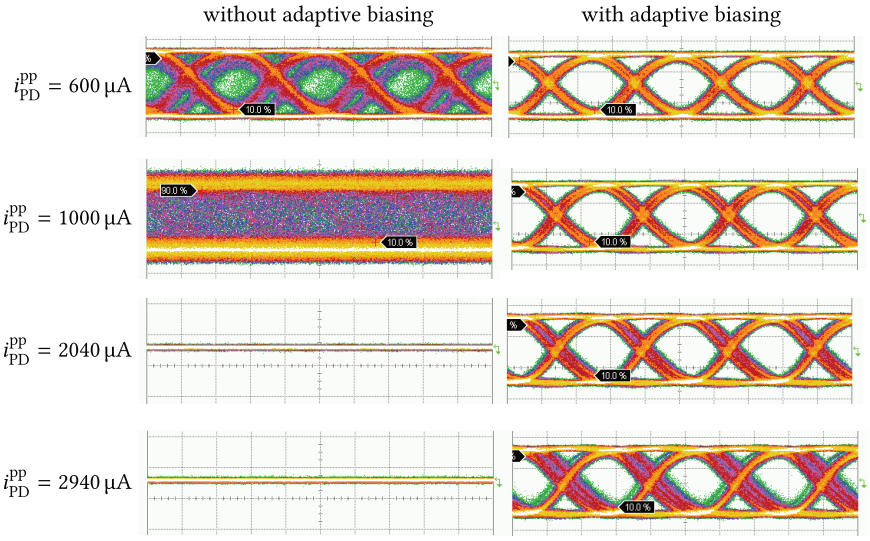


Figure 6.23: Eye diagrams at various input currents without (left) and with (right) adaptive biasing at 22.5 Gb/s.

With adaptive biasing enabled, the eye is restored in all cases, as the TIA is now able to sink all input current without cutting off its output emitter follower. In addition, the dc

voltage at the TIA output is kept reasonable fixed. However, the effect of clipping to the supply rail and saturation of the cascode device now become apparent, in particular for the highest input current. Eventually the eye will close again. This could not be measured as no sufficient optical power could be generated. From the eye diagrams, it can be inferred that the dynamic range is sufficiently high.

6.7.3 Sensitivity

To measure the BER performance, an assembly including an attached fiber array was built (Fig. 6.24, detail in Fig. 6.25). Another type of photo diode array was used, who's pinout corresponds with the chip layout, such that no cross-bonding is required. The photo diode responsivity is 0.41 A/W, while its capacitance and series resistance is 115 fF an 10 Ω, respectively. Alignment is fixed and cannot be changed. The fiber array is held in place with an aluminium structure. Incident light couples to the photo diode array perpendicularly. Next to the die, discrete decoupling capacitors are visible.

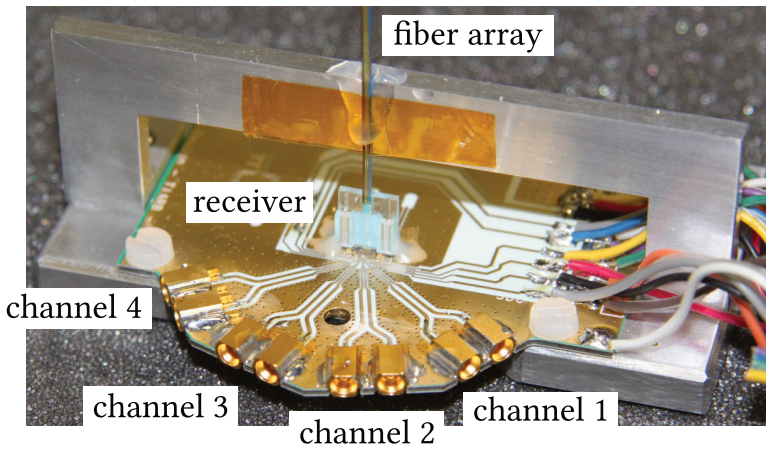


Figure 6.24: Photograph of the assembly with fixed fiber array.

Isolated Channel Performance Measured differential output eye diagrams are shown for all channels in Fig. 6.26. All but the measured channel are dark. Input data is PRBS $2^{31} - 1$ NRZ, with extinction ratio and rms jitter of 14 dB and 950 fs, respectively. Average input photo current is 100 µA. The equipment used was an Agilent 86117A sampling oscilloscope with an Agilent 86107A precision time base module to reduce jitter. All eyes are clearly open. The differential output amplitude is 400 mV peak-peak, while rms jitter amounts to 1.8 ps for channels 2–4. Channel 1, however, is somewhat noisier. The reason is a lower amount of power supply decoupling capacitance on the die for channel 1, as some of its area has been sacrificed for on-chip test structures. The eye diagrams suggest a high signal-to-noise ratio (SNR) for channel 4, followed by channel 2 with channel 3 marginally worse. Unfortunately, after the eye diagram measurements, the fiber and photo diode of channel 4 misaligned, reducing the optical coupling to virtually nothing. As a fix would have required dismantling the assembly, further measurements on channel 4 were not pursued.

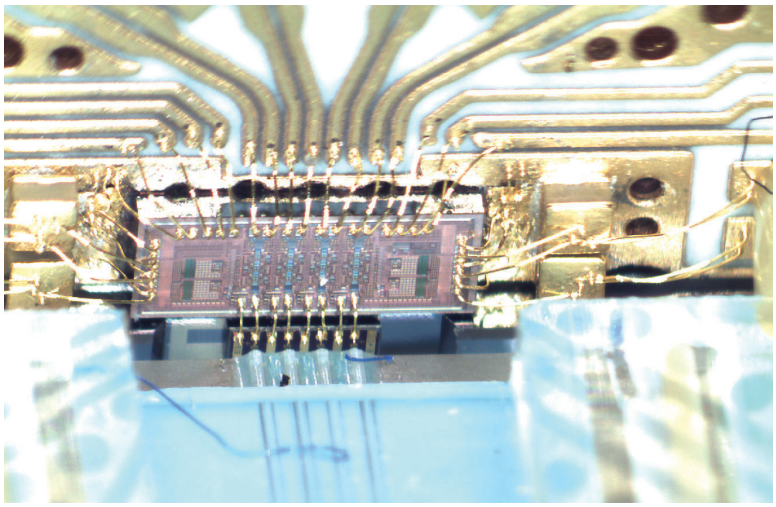


Figure 6.25: Detail of the receiver and photo diode array. Fiber and photo diode array are at the bottom of the picture, output traces at the top.

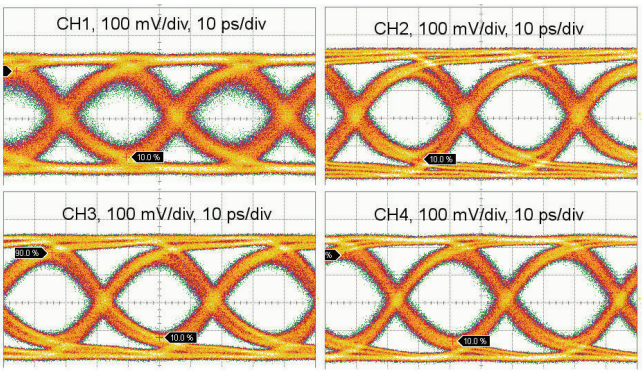
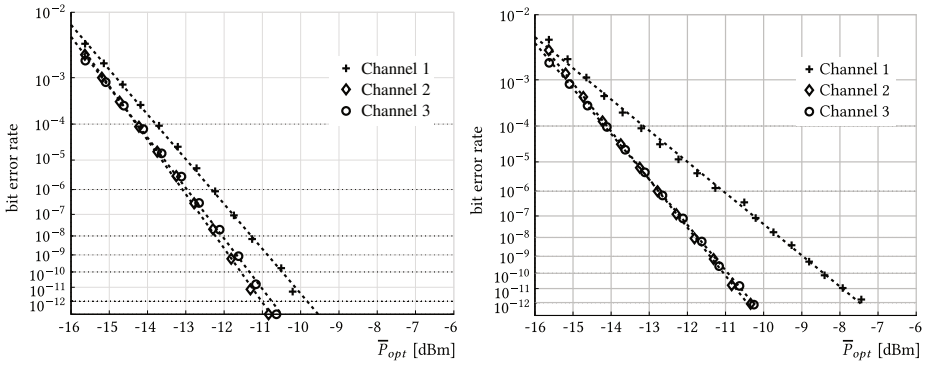


Figure 6.26: Measured eye diagrams of channel 1–4, average photo input current 100 µA.

The BER has been measured using an SHF 12100B pattern generator and SHF 11100B error analyzer using an off-the-shelf transmitter at 25 Gb/s. The extinction ratio and rms jitter of the optical input signal was 14 dB and 950 fs, respectively. Figure 6.27(a) shows the BER curves of channel 1–3 for a NRZ PRBS $2^7 - 1$ input signal at 25 Gb/s. Channel 2 and 3 show an optical sensitivity of −11 dBm and −10.9 dBm (electrical sensitivity 65.1 µA pp and 66.7 µA pp) at a BER of 10^{-12}, respectively. In line with expectations, channel 1 infers an extra penalty of 1 dB due to lower power supply decoupling.

Figure 6.27(b) shows the BER for a NRZ PRBS $2^{31} - 1$ data pattern at 25 Gb/s data rate. Compared to PRBS $2^7 - 1$, channel 2 and 3 infer a power penalty of 0.7 dB (electrical sensitivity 76.5 µA pp and 78.3 µA pp) while the penalty for channel 1 is 2.4 dB (electrical sensitivity 143 µA pp). Also measured, but not shown, is the BER for a PRBS $2^{15} - 1$ pattern. The results are similar to the results of PRBS $2^{31} - 1$. The penalty for longer pattern lengths is caused by the low-frequency high-pass pole in the datapath, introduced by the balancing control loop (Section 6.5). The bigger penalty for channel 1 indicates increased

influence of the reduced supply decoupling at longer pattern lengths. Hereafter, further measurements on channel 1 are not presented.



(a) PRBS $2^7 - 1$ NRZ. (b) PRBS $2^{31} - 1$ NRZ.

Figure 6.27: BER performance of channel 2–4 at 25 Gb/s NRZ.

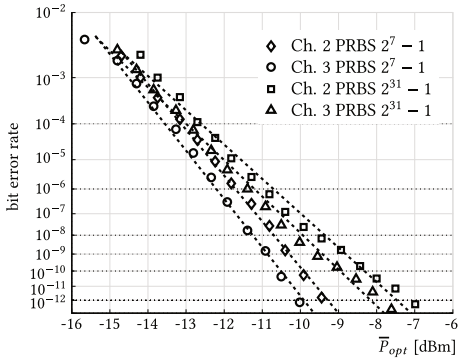


Figure 6.28: BER performance of channel 2 and 3 at 25 Gb/s PRBS $2^7 - 1$ and PRBS $2^{31} - 1$ optical duobinary (ODB).

Duobinary Performance In addition to NRZ modulation, an optical duobinary (ODB) (Section 2.3) back-to-back link has been measured, with the transmitter described in [101]. The photo diode acts as an intensity detector. The extinction ratio and peak-to-peak jitter were 9 dB and 13 ps, respectively. Figure 6.28 depicts the performance of channel 2 and 3 for 25 Gb/s PRBS $2^7 - 1$ and PRBS $2^{31} - 1$. For PRBS $2^7 - 1$, sensitivity is −10 dBm and −9.5 dBm for channel 2 and 3, respectively, indicating a power penalty of 1 dB to 1.4 dB as compared to NRZ signaling. An extra penalty of 2 dB is measured for PRBS $2^{31} - 1$. The degradation, particularly pronunciated for the longer data patterns, is caused by the V-shaped eye opening typical for duobinary encoding and non-perfect timing in the transmitter, in combination with limited total system bandwidth.

Crosstalk Crosstalk measurements are presented in Table 6.3 for channel 2 attacked by channel 3 and vice versa, for a 25 Gb/s PRBS $2^{31} - 1$ data pattern. The BER performance

degradation is shown for an aggressor input power 5 dB and 8 dB higher than the sensitivity of the victim channel. In spite of the small channel pitch of 250 µm, only a penalty of 0.5 dB is observed for the +5 dB attacker. This is mainly attributed to inductive coupling between the bondwires of the adjacent channels, as the die substrate is high-ohmic and various on-chip isolation measures have been taken (Section 2.5.4). The extra penalty for the +8 dB attacker is a mere 0.1 dB. This can be explained by recognizing that, even though there is a twofold difference in input powers, in both cases the back-end stages of the attacking channel are limiting. Hence, the extra degradation must be caused by the front-end stages, in which the signals are smaller to begin with. Due to limitations of the test setup, no measurements could be performed with two adjacent aggressor channels.

victim	aggressor	+5 dB	+8 dB
ch. 2	ch. 3	0.5 dB	0.6 dB
ch. 3	ch. 2	0.5 dB	0.6 dB

Table 6.3: Power penalty due to crosstalk, 25 Gb/s PRBS $2^{31} - 1$.

6.8 Conclusion

A 4×25 Gb/s optical receiver array with small area footprint has been presented. The design and implementation has been discussed in detail, with emphasis on the TIA input stage and MA stages. Bandwidth enhancement through inductive peaking, dynamic range extension with adaptive biasing and the balancing control loop have been elaborated on. The measurements show that the tight integration pitch of 250 µm is feasable. The receiver shows a good sensitivity of −10.3 dBm for PRBS $2^{31} - 1$ NRZ at low power consumption of 77 mW per channel. Furthermore, the power penalty due to crosstalk has been presented, as well as results for optical duobinary modulation.

The experience gathered during the design of this receiver has been used to maximum extent during the design of the linear datapath for the automatic gain control (AGC) system described in Chapter 5, as well as in other follow-up projects within the group.

Chapter 7

Conclusion

The research described in this dissertation is divided into two parts, both of which are related to high-speed optical receivers. This final chapter highlights the important results of the research. Suggestions for future work are given where possible.

7.1 Event-driven AGC for Linear Optical Receiver

A first, important part of the research focused on the conception, modeling and implementation of an event-driven automatic gain control (AGC) system in a high-speed optical linear receiver for multilevel modulation formats (Chapters 3 to 5). A general feedback AGC system in continuous-time was analyzed. It was shown that in order to obtain constant settling time of the closed-loop system with a fixed integrator in the loop, the variable gain amplifier (VGA) transfer curve must be exponentially dependent on its control signal and a logarithmic amplifier must be included in the loop. As the VGA constitutes the multistage datapath of a linear optical receiver, it requires controlled frequency response and in particular limited time-domain overshoot across the gain range, to enable reliable reception of multilevel modulation formats. It has been argued that this control is hard to achieve with fully analog building blocks. Therefore, an event-driven digital approach was proposed as an extension of the continuous-time system. A quantizer and look-up table (LUT) were inserted in the loop and the frequency response of the VGA was made digitally programmable. The quantizer converts the amplified loop error signal to a digital code word, which serves as an index in the LUT. The latter subsequently changes the frequency response of the VGA for each code word. While this approach considerably simplifies the design of the datapath, it also introduces limit-cycling in the closed-loop system. This phenomenon has been analyzed using describing functions, a non-linear analysis technique. In order to avoid these limit cycles, a window comparator is introduced to detect whether the detected output amplitude is within an allowed range.

A system-level model of the proposed approach was developed in Chapter 3, while design on the system level, based on the top-level specifications of the receiver, has been discussed in Chapter 4. In Chapter 5, the implementation of the various building blocks is presented in detail along with experimental results which confirm the validity of the idea. The implemented functionality is not limited to the AGC system only: the realized chip is highly configurable and can be reprogrammed (on-the-fly) for different scenarios.

The performed research suggests multiple possible topics for further study:

- Development of a high-speed accurate peak detector. The output voltage of the implemented detector equals the mean absolute value of its differential input. While this function is performed adequately up to very high speed, the detector output depends on the signal shape. An ideal peak detector is both fast and accurate and detects the true peak-to-peak amplitude, regardless of signal shape.
- Reduction of area footprint by simplification on the system level. It should be possible to combine certain blocks, e.g. the detector, transconductor and logarithmic amplifier using a clever topology, hence reducing silicon area.
- Automatic calibration of the datapath and loading of the LUT, going a step further with the digitally-assisted approach. At start-up, or periodically, an intelligent digital algorithm could perform calibration of the datapath by injecting a known signal and measuring the detector output, thereby automatically filling the LUT with the most appropriate settings (gain, bandwidth, peaking). This is facilitated by a better detector.
- Application of the event-driven approach to a feedforward AGC architecture or burst-mode receiver.

The topology is already being used in new receiver designs in the context of follow-up European FP7 projects in the group, such as "Distributed Core for Unlimited Bandwidth Supply for all Users and Services" (Discus) and "Photonics for High-performance, Low-Cost & Low-energy Data Centers, High Performance Computing Systems" (PhoxTrot).

7.2 Multichannel Optical Receiver

A second part of the research involved the development of a four-channel 4×25 Gb/s limiting optical receiver for non-return-to-zero (NRZ) modulation. The main design challenges were the high data rate, low power consumption and small channel area footprint due to tight integration demands with the photo diode array. Chapter 6 presented the basic architecture and several key building blocks were discussed in detail. Emphasis was put on the transimpedance amplifier (TIA) input stage and the main amplifier (MA) stages, which were discussed in detail. Bandwidth extension through inductive peaking with the bondwire inductance has been presented, as well as dynamic range extension with adaptive biasing. Furthermore, the balancing control loop has been elaborated on. Finally, experimental results confirm the feasibility of the approach.

The vast experience and knowledge gathered during the design of this multichannel receiver has impacted all subsequent optical receiver designs in the group. A particular example is the datapath of the linear optical receiver, used in this work as the VGA in the AGC system.

Appendices

Appendix A

Summary of the General Network Theorem

> The organization of the problem comes from within our minds and feedback is present only if we perceive a closed chain of dependency.
>
> *Samuel J. Mason, 1953*

During the analysis and design of a circuit, various small-signal analyses are used to study and verify linear circuit behavior, next to other types of analyses which expose non-linearities. Rather than examining full-blown transfer functions from input to output, additional insight can be gained by using several simpler lower-level transfer functions, each portraying a subset of the circuit properties—a divide-and-conquer approach. The general network theorem (GNT) or dissection theorem (DT), having a solid theoretical foundation, provides exact solutions for these lower-level transfer functions for any linear circuit, simple or complex.

The key to developing a useful hand model of a circuit is in simplifying the circuit to such an extent that the expressions involved become easy enough to calculate, yet still adequately expose the circuit innards while presenting this information in a manner useful for design. This is the principle of design-oriented analysis (D-OA), in which expressions appear in low-entropy form, not merely as a convoluted collection of symbols [32, 102, 103]. Conform to this mindset, the GNT, as an alternative to brute-force nodal or mesh analysis, decomposes a potentially complex transfer function into simpler parts. This is highly desirable, as it often leads to additional insight into circuit behavior and provides better design guidance. This section provides a short overview of the GNT and its derived theorems [104–107].

During the course of this work, the theorems were numerically implemented in the Cadence Virtuoso EDA suite as a new analysis for the Spectre/APS circuit simulator. It allows validating hand analysis results or finding out which lower-level transfer function dominate circuit behavior, prior to hand analysis. The integration is concisely presented [101, 108].

A.1 The General Network Theorem

Consider a linear time-invariant (LTI) system. The GNT [104] states that any first-level transfer function H, i.e. the ratio of a response u_o to an excitation u_i, can be factored into a combination of lower-level transfer functions, whether or not feedback is involved:

$$H = \frac{u_o}{u_i} = H_y D D_\mathrm{n} \tag{A.1}$$

$$= H_y \frac{1 + \frac{1}{T_\mathrm{n}}}{1 + \frac{1}{T}} \tag{A.2}$$

in which $D = 1/(1 + 1/T)$ and $D_\mathrm{n} = 1 + 1/T_\mathrm{n}$. The lower-level transfer functions are calculated using test signal injection and nulling techniques. The test signal configuration—i.e. the *number* of signal injections and their *location*—determines the final value of the set H_y, T and T_n and their interpretation. It follows that the same H can be decomposed into different mutually consistent sets. Each set, bearing its own interpretation, potentially exposes different circuit properties. The key is to choose an appropriate test signal injection configuration such that the lower-level transfer functions have a desired physical interpretation in terms of the circuit elements.

The origin of the GNT dates back to Bode, who showed that any transfer function can be expressed as a bilinear transformation in one of its circuit elements [96].

Let's determine the general form of the GNT under a given number of test signal injections, disregarding the location of injection. The latter will be dealt with in the sections below and will establish the final interpretation.

1-GNT Consider the system depicted in Fig. A.1. Assume a single test signal injection u_z, which is either voltage or current in the context of circuits. Note that $u_z = u_x + u_y$. The decomposition result is Eq. (A.2), with:

$$H_y = \left.\frac{u_o}{u_i}\right|_{u_y=0} \tag{A.3}$$

$$T = \left.\frac{u_y}{u_x}\right|_{u_i=0} \tag{A.4}$$

$$T_\mathrm{n} = \left.\frac{u_y}{u_x}\right|_{u_o=0} \tag{A.5}$$

Both H_y and T_n are null double-injection (ndi) calculations, while T is a single-injection (si) calculation. An ndi calculation is conceptually performed by mutually adjusting two signal sources, in casu the system input u_i and injected signal u_z, such that another signal (u_y or u_o) is nulled[1]. This kind of calculation is generally simpler than a si calculation as the mere knowledge of the existence of a null allows a high degree of circuit simplification, as the null propagates through the system. T and T_n are called *return ratio* and *null return ratio*, respectively, while D and D_n are called the discrepancy factor and null discrepancy factor, respectively.

[1] Nulling a voltage (current) is not the same as shorting of a node pair (or cutting a branch). The latter changes the circuit determinant, while the former does not [105].

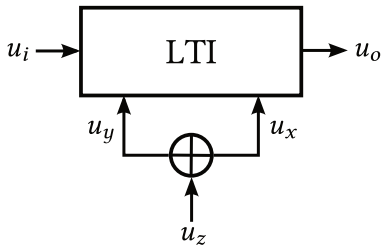


Figure A.1: Single injection in a LTI system with excitation u_i and response u_o.

2-GNT Let's inject two test signals u_{z1} and u_{z2}, simultaneously—each can be either voltage or current—but not necessarily at the same location. The 2-GNT decomposition adopts the following form:

$$H = H_y \frac{1 + \frac{1}{T_{\mathrm{n}1}} + \frac{1}{T_{\mathrm{n}2}} + \frac{1}{T_{\mathrm{n}1}} \frac{1}{T_{\mathrm{n}2}^{(1)}}}{1 + \frac{1}{T_1} + \frac{1}{T_2} + \frac{1}{T_1} \frac{1}{T_2^{(1)}}} \tag{A.6}$$

with:

$$H_y = \left. \frac{u_o}{u_i} \right|_{u_{y1}=0,\, u_{y2}=0} \tag{A.7}$$

$$T_1 = \left. \frac{u_{y1}}{u_{x1}} \right|_{u_i=0,\, u_{y2}=0} \tag{A.8}$$

$$T_2 = \left. \frac{u_{y2}}{u_{x2}} \right|_{u_i=0,\, u_{y1}=0} \tag{A.9}$$

$$T_2^{(1)} = \left. \frac{u_{y2}}{u_{x2}} \right|_{u_i=0,\, u_{x1}=0} \tag{A.10}$$

$$T_1^{(2)} = \left. \frac{u_{y1}}{u_{x1}} \right|_{u_i=0,\, u_{x2}=0} \tag{A.11}$$

and a redundancy relation:

$$T_1 T_2^{(1)} = T_2 T_1^{(2)} \tag{A.12}$$

Similar definitions hold for $T_{\mathrm{n}i}$ and $T_{\mathrm{n}i}^{(j)}$ as for T_i and $T_i^{(j)}$, but with the output u_o nulled instead of the input set to zero. H_y, $T_{\mathrm{n}i}$ and $T_{\mathrm{n}i}^{(j)}$ are double-null triple-injection (dnti) calculation, while T_i and $T_i^{(j)}$ are ndi calculations.

N-GNT It is possible to formulate the general decomposition of H under N injections—the N-GNT or to even generalize the theory to multiple-input multiple-output linear systems, as shown in [109]. Moreover, a N-GNT with N simultaneous injections is equivalent to i successive applications of a GNT with K_i ($K_i < N$) injections in a nested fashion. The results appear in yet a different format, exposing other circuit features. However, this is beyond the scope of this work.

The test signal injection configuration determines the interpretation of the lower-level transfer functions. The GNT morphs into the general feedback theorem (GFT), extra element theorem (EET) and the chain theorem (CT).

A.1.1 The General Feedback Theorem

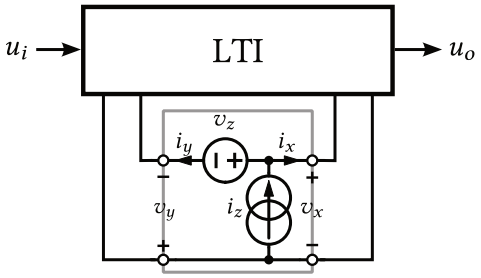


Figure A.2: Simultaneous test voltage and current injection (at the same location) in a LTI circuit with excitation u_i and response u_o.

Consider the major loop of a feedback system. By simultaneously injecting both a voltage and current (Fig. A.2) such that the total feedback error signal—both voltage and current—of the major loop can be nulled, the 2-GNT morphs into the 2-GFT. Typically, the voltage and current are injected at the same location. Equation (A.6) can be rewritten as follows:

$$H = H_\infty D D_\mathrm{n} \tag{A.13}$$

$$= H_\infty \frac{1 + \frac{1}{T_\mathrm{n}}}{1 + \frac{1}{T}} \tag{A.14}$$

with

$$T = \frac{T_\mathrm{fwd}}{1 + T_\mathrm{rev}} \tag{A.15}$$

$$T_\mathrm{fwd} = T_\mathrm{v,fwd} \,||\, T_\mathrm{i,fwd} \tag{A.16}$$

$$T_\mathrm{rev} = T_\mathrm{v,rev} \,||\, T_\mathrm{i,rev} \tag{A.17}$$

and

$$T_\mathrm{n} = \frac{T_\mathrm{n,fwd}}{1 + T_\mathrm{n,rev}} \tag{A.18}$$

$$T_\mathrm{n,fwd} = T_\mathrm{nv,fwd} \,||\, T_\mathrm{ni,fwd} \tag{A.19}$$

$$T_\mathrm{n,rev} = T_\mathrm{nv,rev} \,||\, T_\mathrm{ni,rev} \tag{A.20}$$

in which '||' denotes the parallel operator. There are three redundancy relations:

$$T_\mathrm{v,fwd} T_\mathrm{v,rev} = T_\mathrm{i,fwd} T_\mathrm{i,rev} \tag{A.21}$$

$$T_\mathrm{nv,fwd} T_\mathrm{nv,rev} = T_\mathrm{ni,fwd} T_\mathrm{ni,rev} \tag{A.22}$$

$$H_\infty T = H_0 T_\mathrm{n} \tag{A.23}$$

Equation (A.14) can be written in multiple ways. Another useful form is given, with Eq. (A.23), by:

$$H = H_\infty \frac{T}{1 + T} + \frac{H_0}{1 + T} \tag{A.24}$$

The physical interpretation of the lower-level transfer functions, under the stated test signal configuration, is as follows:

- H_∞: the 'ideal' transfer function, when loop gain is infinite ($T \to \infty$)
- T: the loop gain[2]
- T_n: null loop gain, calculated under output null conditions
- H_0: direct forward transmission (and common-mode gain), when loop gain vanishes ($T \to 0$)

At the risk of stating the obvious, remember that H can be *any* transfer function, including voltage (current) gain, transimpedance (transconductance), input or output impedance, (null) loop gain…

A.1.1.1 Calculation

The following formulas for the 2-GFT decomposition result from applying the 2-GNT:

H_∞	$\left.\frac{u_o}{u_i}\right\|_{i_y=0,\, v_y=0}$	dnti	'ideal' transfer function
$T_\mathrm{v,fwd}$	$\left.\frac{v_y}{v_x}\right\|_{u_i=0,\, i_y=0}$	ndi	forward open-circuit voltage loop gain
$T_\mathrm{i,fwd}$	$\left.\frac{i_y}{i_x}\right\|_{u_i=0,\, v_y=0}$	ndi	forward short-circuit current loop gain
$T_\mathrm{v,rev}$	$\left.\frac{v_x}{v_y}\right\|_{u_i=0,\, i_x=0}$	ndi	reverse open-circuit voltage loop gain
$T_\mathrm{i,rev}$	$\left.\frac{i_x}{i_y}\right\|_{u_i=0,\, v_x=0}$	ndi	reverse short-circuit current loop gain
$T_\mathrm{nv,fwd}$	$\left.\frac{v_y}{v_x}\right\|_{u_o=0,\, i_y=0}$	dnti	forward open-circuit voltage null loop gain
$T_\mathrm{ni,fwd}$	$\left.\frac{i_y}{i_x}\right\|_{u_o=0,\, v_y=0}$	dnti	forward short-circuit current null loop gain
$T_\mathrm{nv,rev}$	$\left.\frac{v_x}{v_y}\right\|_{u_o=0,\, i_x=0}$	dnti	reverse open-circuit voltage null loop gain
$T_\mathrm{ni,rev}$	$\left.\frac{i_x}{i_y}\right\|_{u_o=0,\, v_x=0}$	dnti	reverse short-circuit current null loop gain

Equation (A.23) represents the easiest way to manually calculate H_0. Another expression is [110]:

$$H_0 = \frac{1}{1+T_\mathrm{rev}}\left(H\Big|_{i_y=0,\, v_x=0}\frac{T_\mathrm{fwd}}{T_\mathrm{v,fwd}} + H\Big|_{i_x=0,\, v_y=0}\frac{T_\mathrm{fwd}}{T_\mathrm{i,fwd}} + H\Big|_{i_x=0,\, v_x=0}T_\mathrm{rev}\right) \tag{A.25}$$

which is a weighted sum of three dnti calculations. Note that H_0 is *not* equal to $H|_{i_x=0,\, v_x=0}$.

A.1.1.2 Ideal injection points

The 2-GFT simplifies to the 1-GFT when the test signal injection point is ideal. In an ideal injection point the error current i_y is automatically nulled when the error voltage v_y is nulled, or vice versa. It follows that only a single voltage or current injection is required (1-GNT).

Under ideal voltage injection, $T_\mathrm{i,fwd}$ and $T_\mathrm{ni,fwd}$ are infinite. The calculations simplify to:

[2] In a generalized sense of Bode's return ratio [96]. There are an infinite number of loop gains (in that sense) in a circuit, hence a better name is the 'principal' loop gain, being derived under the GFT decomposition.

H_∞	$\left.\frac{u_o}{u_i}\right\|_{v_y=0}$	ndi	'ideal' transfer function
$T = T_{\text{v,fwd}}$	$\left.\frac{v_y}{v_x}\right\|_{u_i=0}$	si	(forward open-circuit voltage) loop gain
$T_\text{n} = T_{\text{nv,fwd}}$	$\left.\frac{v_y}{v_x}\right\|_{u_o=0}$	ndi	(forward open-circuit voltage) null loop gain
H_0	$\left.\frac{u_o}{u_i}\right\|_{v_x=0}$	ndi	direct forward transmission

Under ideal current injection, $T_{\text{v,fwd}}$ and $T_{\text{nv,fwd}}$ are infinite. It follows:

H_∞	$\left.\frac{u_o}{u_i}\right\|_{i_y=0}$	ndi	'ideal' transfer function
$T = T_{\text{i,fwd}}$	$\left.\frac{i_y}{i_x}\right\|_{u_i=0}$	si	(forward open-circuit voltage) loop gain
$T_\text{n} = T_{\text{ni,fwd}}$	$\left.\frac{i_y}{i_x}\right\|_{u_o=0}$	ndi	(forward open-circuit voltage) null loop gain
H_0	$\left.\frac{u_o}{u_i}\right\|_{i_x=0}$	ndi	direct forward transmission

It can be shown that if the desired H_∞ is obtained with nonzero reverse loop gain T_rev, no ideal injection point can be found that produces the same H_∞. In other words, T_rev must vanish when an ideal injection point can be found (that produces the desired H_∞).

It is pointless to inject in an ideal injection point when H_∞ does not result in the desired interpretation. In some circuits, there exist multiple injection points that result in the same set H_∞, T and T_n (but different third-level transfer functions). If one of those injection points is ideal, the 1-GFT can be used for hand analysis. Furthermore, a circuit can be deliberately simplified to create an ideal injection point to facilitate analysis.

A.1.1.3 Discussion

Although these decompositions may appear difficult and complex, they are not. The formulas involved are highly structured, have a physical interpretation and are generally easier to calculate than the first-level transfer function. Typically, only a few of them are required to obtain a useful hand model of the circuit. The loop gain is the most complex one. Fortunately, when determined for one transfer function (e.g. current gain), it can be reused in others[3] (e.g. input impedance) such that only null injection calculations remain. In addition, D-OA tools such as doing the algebra on the graph and inverted pole/zero notation should be used to quickly write down simplified transfer functions, without resorting to convoluted algebra. When in doubt, cross-check with the simulator.

None of the lower-level transfer functions are 'true' transfer functions of the circuit, in the sense that their denominator equals the circuit determinant. Rather, they are ratios of 'true' transfer functions. Hence, a right-hand plane pole appearing in e.g. H_0 bears no meaning with respect to stability of the considered circuit.

The principal loop gain T (Eq. (A.15)) is a return ratio with respect to a parameter k in Bode's sense and can be used for stability assertion. That is, $F = 1 + T = \Delta/\Delta_0$ is a return difference, with Δ the circuit determinant and Δ_0 the circuit determinant with $k = 0$. The Nyquist criterion gives the difference between the poles of the numerator and poles of the denominator of F in the right hand-side plane (RHP). If the circuit associated with Δ_0 is assumed or known to be stable (as is the case single-loop feedback systems, by definition),

[3] As long as the circuit determinant remains the same.

stability is readily evaluated. If not, multiple-loop feedback theory should be applied[4] [96]. Note that none of constituting parts of T are return ratios. By all means, however, they can be used as approximations for the principal loop gain in hand calculations.

Theoretically, an infinite number of loop gains can be found in a system, each with a gain margin and phase margin. The margins are different, in general, yet they all converge to zero as the system approaches instability [111, 112]. It can be shown that, a.o., the following loop gain definitions are return ratios in Bode's sense [113]:

- 2-GFT loop gain.
- Tian's loop gain, under conditions given in [114] (implemented in Spectre's *stb*).
- Middlebrook's old method [115] (MB75). Although derived with disregard to reverse loop gain, the result is generally applicable.
- Result from single voltage or single current injection.

None of these methods require an ideal injection point. The 2-GFT, MB75 and Tian's loop gain are invariant with respect to the injection point, as long as the same loop is considered (but 2-GFT H_∞ and T_n are not, obviously). Moreover, Tian's is symmetrical[5] as it considers both the forward and reverse loop transmission at once. *Unique to the 2-GFT loop gain is the clear relation to the closed-loop transfer function and hence closed-loop behavior, as it emerges as the result of the decomposition of the closed-loop transfer function.*

Equation (A.24) looks similar to the asymptotic gain model (AGM) [31, 116–119]. The AGM however, is equivalent to the 1-GFT and has the same restrictions: it requires an ideal injection point and hence necessarily omits reverse loop gain and common-mode gain. For hand calculation, AGM, not being derived from the GNT, does not invoke the null injection technique. The AGM derives the lower-level transfer functions by taking limits of one specific dependent generator. In general, this results in a H_∞ that does not assume the desired form. In contrast, the starting point of the GFT *is* a desired H_∞, and the lower-level transfer functions, with their associated interpretation, follow.

Circuits with multiple feedback loops, nested or otherwise (e.g. nested Miller compensation or common-mode/differential-mode decomposition), can be analyzed by either nested application of the GFT or by extra injections and applying multiple-loop feedback theory to the resulting loop gains. As stated in the introduction, nested and flat application of the GNT are essentially equivalent and it is up to the designer to choose the most appropriate form for the given circuit.

Finally, the presence of a non-zero lower-level transfer function in a given decomposition establishes no intrinsic property of the circuit. It is merely a way of organizing the behavior of the circuit in an understandable fashion. For instance, reverse loop gain vanishes under ideal injection, while it could be nonzero under the desired injection. Still, the same circuit is considered. It all comes from within our minds. " 'Feedback loops' are conveniences in modeling a system identified so that constituents of various functions can be associated with certain physical properties" [112, 120].

A.1.2 The Extra Element Theorem

The N-EET allows to decompose a transfer function in terms of its value when N extra elements (EEs) Z_i are absent, and a correction factor expressing the modifications due

[4] Incidentally, this is the reason why T of the EET, although a return ratio, is useless for stability assertion in general.

[5] Tian's loop gain equals $T_\mathrm{fwd} + T_\mathrm{rev}$ (Eqs. (A.16) and (A.17)) when the same reference node is used.

the EEs. An element is absent if it is assigned either a zero or infinite value, selected by appropriate injection.

The 1-EET considers only one extra element W and a single injection suffices:

$$H = H_{\text{ref}} DD_{\text{n}} \tag{A.26}$$

$$= H_{\text{ref}} \frac{1 + \frac{1}{T_{\text{n}}}}{1 + \frac{1}{T}} \tag{A.27}$$

When W is an impedance or admittance, either voltage or current injection can be used. In case W is a voltage gain or a current gain, respectively voltage or current injection *must* be used to avoid indeterminacies. Depending on the injection signal type and direction, Eq. (A.27) morphs into multiple equivalent formulations. To avoid confusion, the following practical guideline is adopted: *in order to obtain H_{ref} (by nulling u_y) equal to H under $W = \infty$ condition when u_y faces W, or to obtain H_{ref} equal to H under $W = 0$ condition when u_x faces W, choose the type (voltage or current) of the injected signal u_z equal to the type of the numerator of W.*

For example, in case W is an impedance Z, choose voltage injection. When v_y faces Z, H_{ref} is calculated with reference value $Z = \infty$.

Under the given guideline, the return ratio and null return ratio (Eqs. (A.4) and (A.5)) can be expressed as ratios of the EE and the driving point immitances W_{d} and null driving point immitances W_{n} as follows. When u_y faces W, H_{ref} is calculated with reference value $W = \infty$ and:

$$T = \frac{W}{W_{\text{d}}} \tag{A.28}$$

$$T_{\text{n}} = \frac{W}{W_{\text{n}}} \tag{A.29}$$

such that:

$$H = H|_{W=\infty} \frac{1 + \frac{W_{\text{n}}}{W}}{1 + \frac{W_{\text{d}}}{W}} \tag{A.30}$$

When u_x faces W, H_{ref} is calculated with reference value $W = 0$. In this dual case, the return ratio and null return ratio (Eqs. (A.4) and (A.5)) can be expressed as:

$$T = \frac{W_{\text{d}}}{W} \tag{A.31}$$

$$T_{\text{n}} = \frac{W_{\text{n}}}{W} \tag{A.32}$$

such that:

$$H = H|_{W=0} \frac{1 + \frac{W}{W_{\text{n}}}}{1 + \frac{W}{W_{\text{d}}}} \tag{A.33}$$

Whichever form is used to calculate the 1-EET decomposition is up to the designer. Equation (A.27) provides a general computation method, however occasionally Eqs. (A.30) and (A.33) are more convenient as W is already known.

The N-GNT can be rewritten as the N-EET, for multiple extra elements. This will not be considered here.

A.1.3 The Chain Theorem

Consider two cascaded amplifier stages. A single test signal u_z is injected in a wire that carries the entire signal from the output of the first stage to input of the second stage. Moreover, u_y faces output of the first stage. As the signal cannot bypass the injection point, T_n is infinite. For current injection, the GNT morphs into:

$$H = H_\mathrm{ref} \frac{1}{1 + \frac{1}{T}} \tag{A.34}$$

$$= H_\mathrm{ref} \frac{Z_{i2}}{Z_{o1} + Z_{i2}} \tag{A.35}$$

With H_ref the voltage-buffered gain of the two stages, i.e. the gain of the cascade when an ideal voltage buffer would be inserted between the stages. The discrepancy factor represents the loading that the second stage imposes upon the first one. Z_{i2} and Z_{o1} are the input impedance of the second stage and the output impedance of the first stage, respectively.

Similarly, for voltage injection:

$$H = H'_\mathrm{ref} \frac{1}{1 + \frac{1}{T'}} \tag{A.36}$$

$$= H'_\mathrm{ref} \frac{Z_{o1}}{Z_{i2} + Z_{o1}} \tag{A.37}$$

With H'_ref the current-buffered gain of the two stages, i.e. the gain of the cascade when an ideal current buffer would be inserted between the stages. The discrepancy factor represents the loading that the second stage imposes upon the first one. In addition, H can be expressed as the parallel combination of the voltage-buffered and current-buffered gains:

$$H = H_\mathrm{ref} \,||\, H'_\mathrm{ref} \tag{A.38}$$

Extension of the CT to three or more stages uses multiple current and voltage injections or nested application of the GNT and is beyond the scope of this overview.

A.2 Implementation in Cadence Virtuoso

The GNT states how to decompose a transfer function, using injection and nulling techniques. Analytically this often results in easier calculations and leads to insightful results. During the course of this work, the various GNT theorems were numerically integrated in the Cadence Virtuoso software [101, 108] to allow direct application of the theory to real-world designs. It allows validating hand analysis results or finding out which lower-level transfer functions dominate circuit behavior.

The integration aims to be fully transparent, with use model identical to the other analyses types. It integrates as a new analysis type, *gnt*, with the familiar Choosing Analyses form for Spectre/APS and presents options similar to ac, stb and the like (Fig. A.3). This implementation supports nested GNT application, such as doing an EET analysis on the lower-level transfer functions of a GFT factorization. The decomposition results are included in the psf data and can be accessed with ViVa's Results Browser. The specification-driven simulation environment ADE XL is also supported.

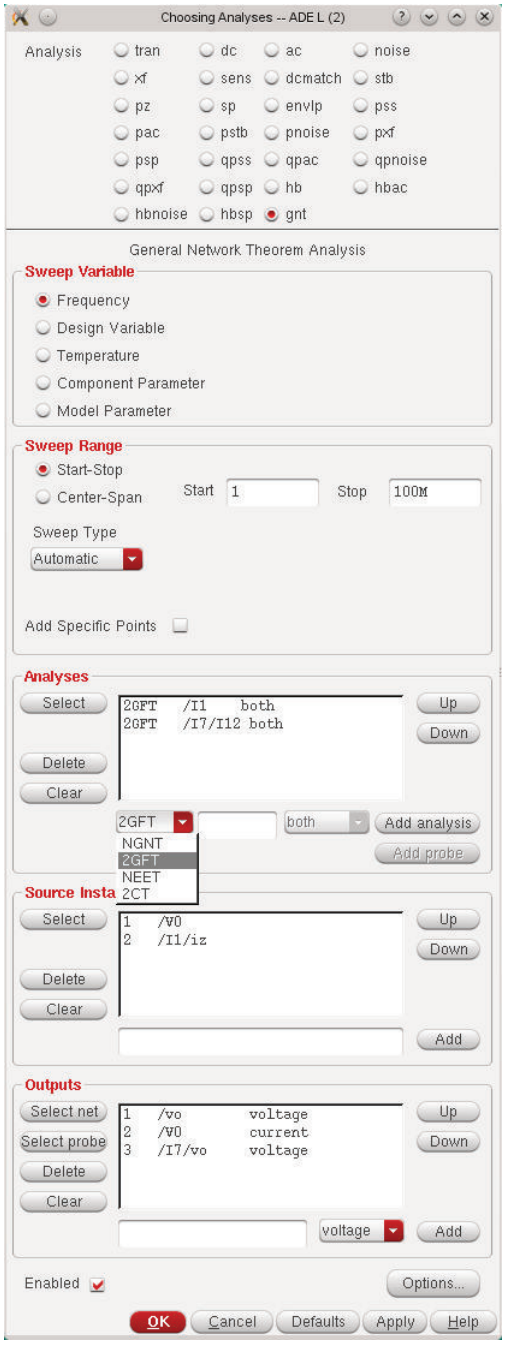


Figure A.3: Options for *gnt* in ADE's Choosing Analyses form.

Appendix B

Miller Compensation with Capacitance Multipliers

Slow control loops, like the automatic gain control (AGC) system in this work, require low gain-bandwidth product (GBW) products. This typically requires a big compensation capacitor or small input stage transconductance of the loop filter. In the first case precious silicon area is wasted, while in the second case gain and offset suffers. In this appendix, Miller compensation with capacitance multipliers is examined. It will be shown that for equal GBW, a smaller compensation capacitance can be used at the expense of increased power consumption, compared to straightforward Miller compensation.

Miller compensation with capacitance multipliers has been studied before [82,121,122] and an analysis is presented in [123] for big load capacitances. In the current context, the load capacitance of the loop filter is small, a few hundred femtofarad. Therefore, a design-oriented analysis is presented for conditions in this work. Classic Miller compensation of two-stage amplifiers will be reviewed. This technique has been described extensively in literature [31, 78, 93, 124, 125]. Barring measures, the stabilizing effect of pole-splitting is partially undone by feedforward transmission that results in a right hand-side plane (RHP) zero. Inserting a current buffer breaks the forward transmission through the feedback path and eliminates the zero. If additionally a current gain is employed, either the GBW decreases accordingly or, for the same GBW, a smaller compensation capacitance is needed. This can result in huge area savings and hence potential silicon cost reduction. As the dynamic behavior of the amplifier is now determined by an internal loop, a model of the system, including the internal loop and the input resistance of the current buffer, will be developed. Compensation of this internal loop will be discussed.

B.1 Miller Compensation Revisited

Consider the linear model of a two-stage amplifier with Miller compensation depicted in Fig. B.1. A compensation capacitance C_c is connected across both stages, which are modeled as a transconductance, output resistance and output capacitance. The input-output transfer function $H = v_o/v_i$ is frequently analyzed using nodal or mesh analysis [31, 78]. An alternative method uses the 1-extra element theorem (EET), with C_c designated as the extra element [33, 108]. In any case, the outcome is the same and shown in Fig. B.2: the

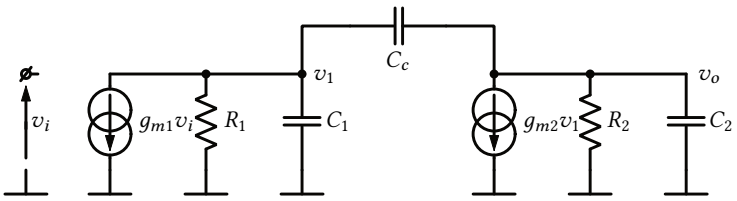


Figure B.1: Model of a two-stage Miller-compensated amplifier.

initial poles are split and a RHP zero appears.

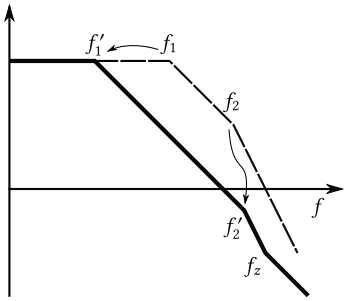


Figure B.2: Classical pole-splitting compensation result.

Here, those results will be reproduced from a feedback point of view, for which the general feedback theorem (GFT) is a natural analysis tool. H will be decomposed as:

$$H = H_\infty \frac{1 + \frac{1}{T_n}}{1 + \frac{1}{T}} \tag{B.1}$$

$$= H_\infty D D_n \tag{B.2}$$

For a short introduction and the definitions of the second-level transfer functions, see Appendix A. Most compensated opamps can be approximated by an ideal integrator, hence it makes sense to choose the injection location such that H_∞ represents an ideal integrator. Multiple locations satisfy this condition, however, in this simple model, an ideal voltage injection point exists at the input of controlled source g_{m2}, shown in Fig. B.3. This simplifies the analysis as the 2-GFT now reduces to the 1-GFT.

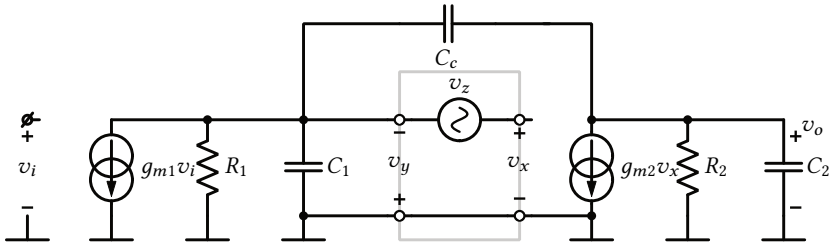


Figure B.3: Injection location to perform 1-GFT analysis on the two-stage Miller compensated amplifier.

The components of H are calculated next. For H_∞:

$$H_\infty = \left.\frac{v_o}{v_i}\right|_{v_y=0} = \frac{g_{m1}}{sC_c} \tag{B.3}$$

Indeed, all current generated by g_{m1} must flow in C_c when v_y is nulled. H_∞ is an ideal integrator, with Bode magnitude plot depicted in Fig. B.4.

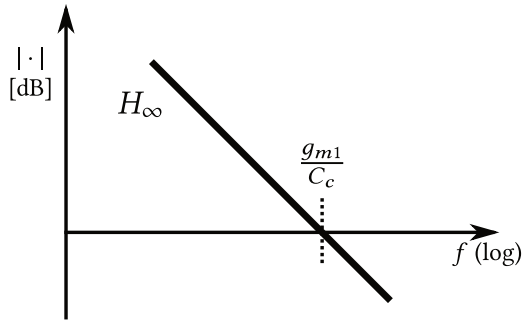


Figure B.4: Desired transfer function H_∞: ideal integrator.

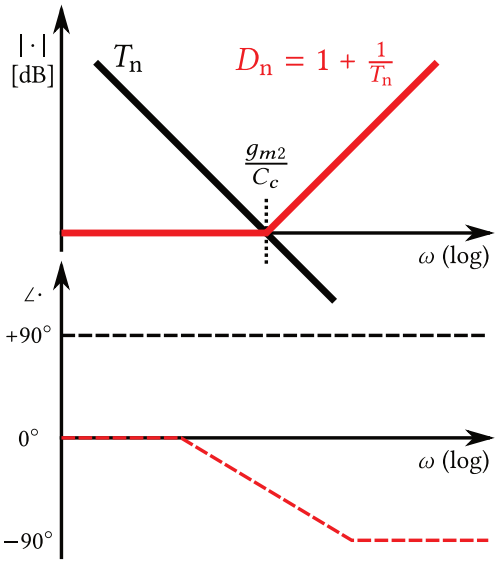


Figure B.5: Null loop gain T_n and null discrepancy factor D_n.

The null loop gain is a null double injection calculation:

$$T_\mathrm{n} = \left.\frac{v_y}{v_x}\right|_{v_o=0} = -\frac{g_{m2}}{sC_c} \tag{B.4}$$

and represents and inverted ideal integrator. The null discrepancy factor follows from

Fig. B.5:

$$D_{\mathrm{n}} = 1 - \frac{s}{g_{m2}/C_c} \tag{B.5}$$

which is a RHP zero, as indicated in the Bode phase diagram.

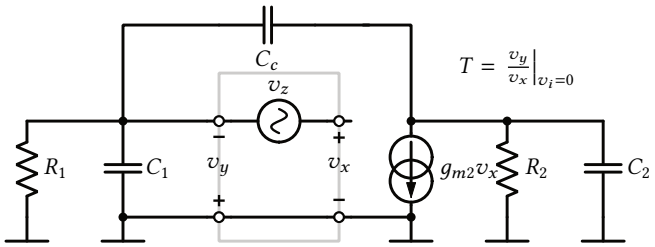


Figure B.6: Circuit for determination of the loop gain T.

To calculate the loop gain $T = v_y/v_x$, v_i is set to zero as shown in Fig. B.6. It is easily inferred from the circuit diagram that T has two poles and one zero[1]. In addition, T must vanish when the frequency tends to zero or infinity. Hence the shape of T is known and depicted in Fig. B.7. Denote the midband gain T_m. Now assume $C_c \gg C_1, C_2$. For midband frequencies C_c is a short while C_1 and C_2 are open, approximately. It follows that:

$$T_m = g_{m2}\,(R_1 \,||\, R_2) \tag{B.6}$$

It is reasonable to assume that $|T_m| \gg 1$. The discrepancy factor D is graphically derived

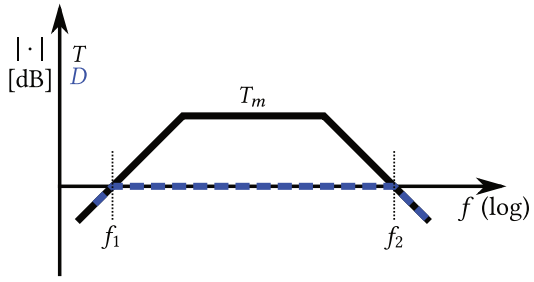


Figure B.7: Loop gain T and discrepancy factor D.

and can be written as:

$$D = \frac{1}{\left(1 + \frac{\omega_1}{s}\right)\left(1 + \frac{s}{\omega_2}\right)} \tag{B.7}$$

The corner frequencies f_1 and f_2 of D directly affect the complete transfer function H.

[1]The number of poles (of a circuit) is equal to the maximum number of independent initial conditions on the energy-storing (reactive) elements. The number of zeros (of a network function) is the maximum number of energy-storing (reactive elements) that can be simultaneously infinite while still producing a non-zero output [126, 127].

They are approximated by neglecting C_1, C_2 and R_2, R_2, respectively:

$$\omega_1 \approx \frac{1}{g_{m2}R_1R_2C_c} \tag{B.8}$$

$$\omega_2 \approx \frac{g_{m2}(C_c + C_1)}{C_1C_2 + C_c(C_1 + C_2)} \tag{B.9}$$

$$\approx \frac{g_{m2}}{(C_1 + C_2)} \tag{B.10}$$

The inverted pole ω_1 is determined by the output resistance of the first stage and the Miller-multiplied compensation capacitance. As C_c tends to infinity, ω_1 tends to zero. However, ω_2 is restricted by the output capacitances as g_{m2} becomes diode-connected. Furthermore, stability should be ensured by placing any other parasitic pole in the amplifier at a decent distance from f_2.

Figure B.8 shows the complete GFT decomposition of H, including the important frequencies. The discrepancy factor and null discrepancy factor introduce two poles and a RHP zero into the ideal integrator. For illustrative purposes, the direct forward transmission H_0 is calculated for high frequencies, using the redundancy relation Eq. (A.23):

$$H_0 \approx -\frac{g_{m1}(C_c + C_1)}{C_1C_2 + C_c(C_1 + C_2)} \tag{B.11}$$

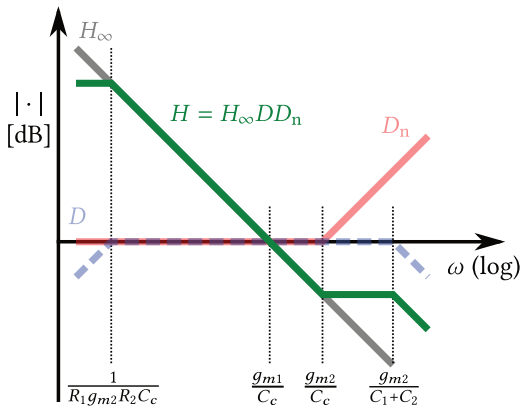


Figure B.8: Summary of the 1-GFT decomposition of a two-stage Miller compensated amplifier.

The RHP zero is undesired as the extra phase lag reduces stability margin. It can be moved to the left hand-side plane (LHP) or to infinity by adding a well-chosen resistance in series with C_c. It can also be completely avoided by breaking the direct forward transmission. This is achieved by including a voltage buffer or a current buffer in the feedback path. Which method is chosen is subject to several trade-offs, that will not be discussed here [31]. As our ultimate target is a model of Miller compensation with current multipliers, the inclusion of a current buffer is analyzed next.

B.2 Avoiding the RHP Zero: Current Buffer

The analysis will be conducted in three steps. First, an ideal current buffer is added in the feedback path. Next, a gain factor is included to model an ideal current amplifier. Third, the finite input resistance of the buffer is incorporated to obtain a realistic model. This last step is necessary as comparison to a real circuit revealed that the model was too incomplete.

B.2.1 Ideal Current Buffer

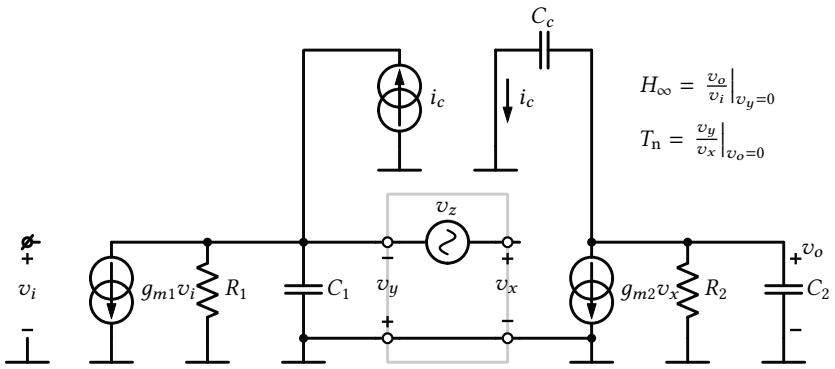


Figure B.9: Introduction of current buffer to avoid RHP zero.

As depicted in Fig. B.9, an ideal current buffer measures the current flowing into the reference node through C_c and injects a copy in the output of the first stage. The second-level transfer functions of the decomposition are modified as follows:

$$H_\infty = \frac{g_{m1}}{sC_c} \tag{B.12}$$

$$T_\mathrm{n} = \infty \tag{B.13}$$

H_∞ has not changed. The null loop gain is infinite as nulling v_o requires that $v_x = 0$ for any v_y. It follows that $D_\mathrm{n} = 1$. Clearly no zero is introduced (compare to Fig. B.5). For the loop gain and discrepancy factor, consider Fig. B.10. D can still be written in the form of Eq. (B.7). For midband frequencies C_c is a short. It follows that the midband gain has increased to $T_m = g_{m2}R_1$. In addition, remark that as the first stage is completely decoupled from the second stage, the first pole of T *must* be determined by R_2 and C_c. Hence the lower unity-gain frequency of D is:

$$\omega_1 \approx \frac{1}{g_{m2}R_1R_2C_c} \tag{B.14}$$

which is equal to the case without buffer. It is easy to show that the higher unity-gain frequency of D is given by:

$$\omega_2 \approx \frac{g_{m2}}{C_2 + C_c}\frac{C_c}{C_1} \tag{B.15}$$

$$\approx \frac{g_{m2}}{C_1} \tag{B.16}$$

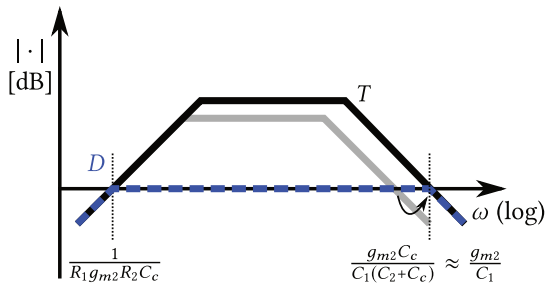


Figure B.10: Changes to the loop gain T when a current buffer is used.

Figure B.11(b) shows the decomposition of H with a current buffer, with the important frequencies indicated. Compared to normal Miller compensation (Fig. B.11(a)), the zero has disappeared and the second pole frequency is higher, while the GBW is the same. This means that, for equal GBW, this configuration supports a bigger load capacitance. However, parasitic poles now become relatively more important and stability must be carefully assessed. The redundancy relation (Eq. (A.23)) shows that the direct forward transmission $H_0 = 0$. This confirms the previous findings.

B.2.2 Ideal Current Amplifier

It will now be shown that gain in the current buffer will allow the use of a smaller compensation capacitance while maintaining the same GBW product, at the expense of increased power consumption.

Let's introduce a current gain $k > 1$, as shown in Fig. B.12. The shape of the second-level transfer functions remains unchanged. Only a change in magnitude is introduced.

$$H_\infty = \frac{g_{m1}}{skC_c} \tag{B.17}$$

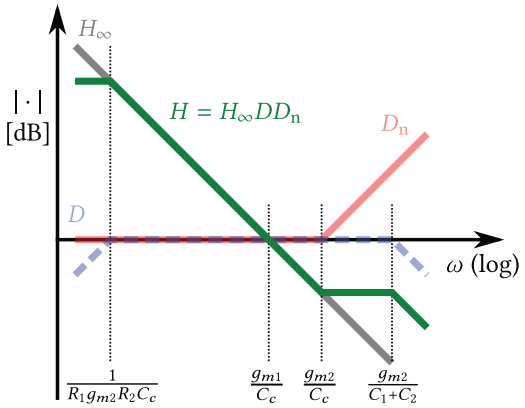
The pole frequencies of D are easily derived from Fig. B.13 as:

$$\omega_1 \approx \frac{1}{kg_{m2}R_1R_2C_c} \tag{B.18}$$

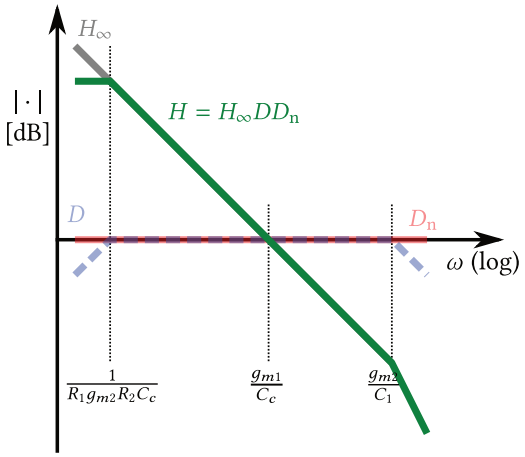
$$\omega_2 \approx \frac{kg_{m2}}{C_2 + C_c}\frac{C_c}{C_1} \tag{B.19}$$

$$\approx \frac{kg_{m2}}{C_1} \tag{B.20}$$

This shows that the GBW of H is k times lower while its second pole of H is at a k times higher frequency. Considering T, any parasitic pole from the current buffer will degrade the stability of the loop, hence they must be placed far enough from the loop GBW, i.e. f_2. The significance of the configuration with current gain k is, that for equal GBW, a k times smaller capacitance can be used. In other words: the equivalent compensation capacitance is kC_c.



(a) Without current buffer.



(b) With current buffer.

Figure B.11: Comparison of the 1-GFT decomposition without and with current buffer.

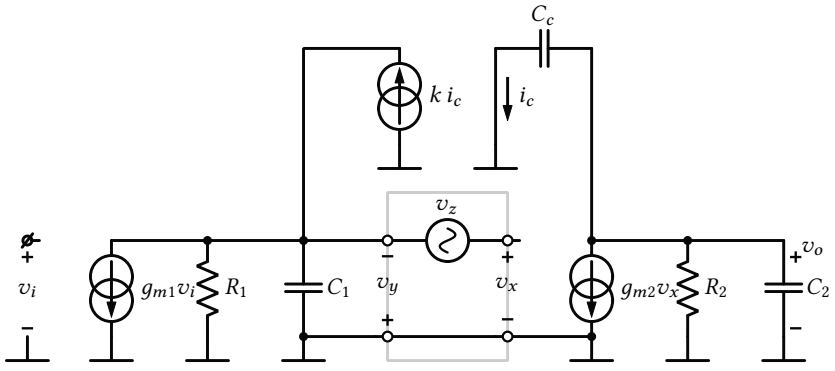


Figure B.12: Introduction of current gain.

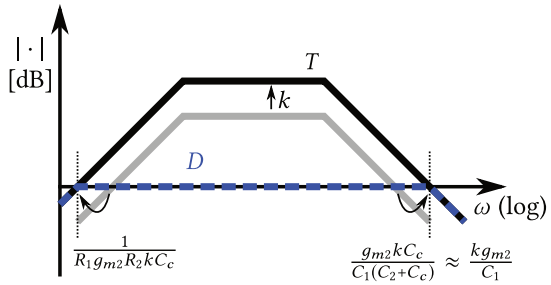


Figure B.13: Changes to the loop gain T when a current gain k is introduced.

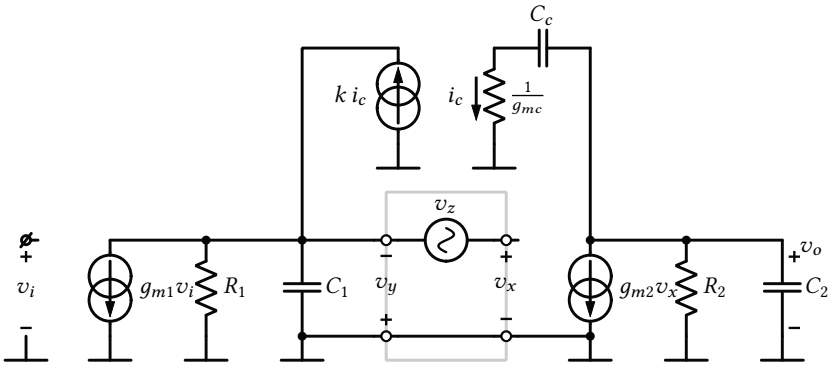


Figure B.14: More realistic current buffer with non-zero input impedance.

B.3 Current Amplifier with Non-zero Input Resistance

The model developed so far proves to be too simple in practice. In order to obtain a more realistic model, the non-zero input resistance $1/g_{mc}$ of the current amplifier is now included (see Fig. B.14). This makes the system more complex and the analysis more elaborate, not in the the least because the system now has three poles. Indeed, each capacitance can sustain an independent initial condition. However, a lot of algebra will be avoided by applying the 1-EET (with zero reference impedance), in a nested fashion, to the 1-GFT decomposition. In other words: the effect of the extra element $Z = 1/g_{mc}$ on the second level transfer functions is to be computed. One can calculate the 1-EET by using either (null) return ratios or (null) driving point impedances [105]. Both methods are equivalent. The latter approach will be taken.

It will be shown that non-zero input resistance will introduce a LHP zero in H_∞ and an additional pole in T. The latter which will lead to a complex pole pair in D and stability concerns of the internal loop.

Clearly, the null loop gain T_n of the 1-GFT decomposition of H remains infinite. The changes to H_∞, T and D are considered next. Remark that for each lower-level transfer function, the respective original injection conditions must be kept in place when applying the EET.

EET Decomposition of H_∞

The 1-EET decomposition of H_∞ (condition $v_y = 0$) is:

$$H_\infty = H_\infty|_{Z=0} \frac{1 + \frac{Z}{Z_\mathrm{n}}}{1 + \frac{Z}{Z_\mathrm{d}}} \tag{B.21}$$

in which Z_d and Z_n are the driving point impedance and null driving point impedance across Z, respectively. Obviously, $H_\infty|_{Z=0}$ is given by Eq. (B.17).

For the driving point impedance Z_d, consider Fig. B.15, in which the circuit diagram is redrawn under 1-GFT H_∞ conditions ($v_y = 0$). The circuit input must be set to zero. From $v_i = 0$ follows that $i_c = 0$, which is the excitation current to calculate Z_d. This holds for any possible voltage developed across Z, hence Z_d must be infinite.

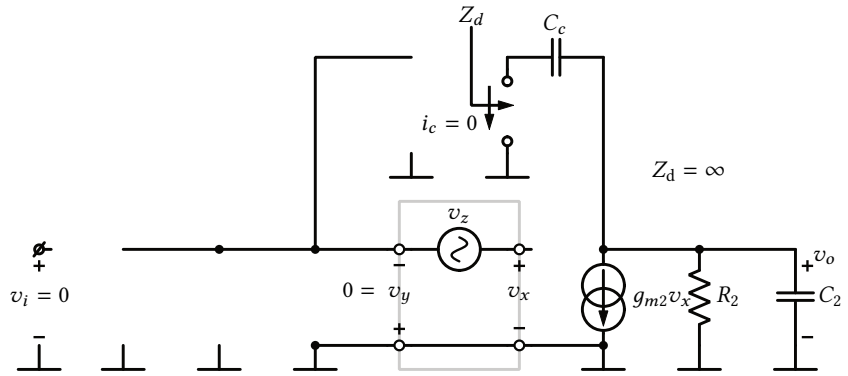


Figure B.15: Circuit diagram to calculate the driving point impedance Z_d of the 1-EET decomposition of the 1-GFT H_∞.

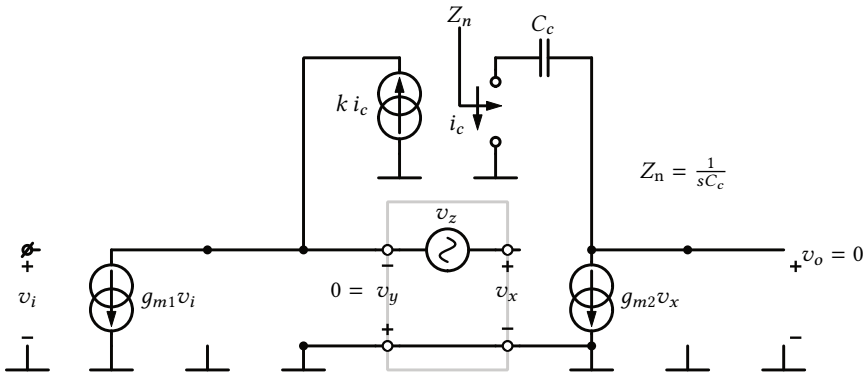


Figure B.16: Circuit diagram to calculate the null driving point impedance Z_n of the 1-EET decomposition of the 1-GFT H_∞.

The circuit for the null driving point impedance Z_n, the impedance across Z when the ouput v_o is nulled, is depicted in Fig. B.16. This is easily seen to be the capacitance C_c. In

summary:

$$H_\infty|_{Z=0} = \frac{g_{m1}}{skC_c} \tag{B.22}$$

$$Z_\mathrm{d} = \infty \tag{B.23}$$

$$Z_\mathrm{n} = \frac{1}{sC_c} \tag{B.24}$$

$$H_\infty = \frac{g_{m1}}{skC_c}\left(1 + \frac{sC_c}{g_{mc}}\right) \tag{B.25}$$

A LHP zero is introduced by the non-zero input resistance, as graphically depicted in Fig. B.17.

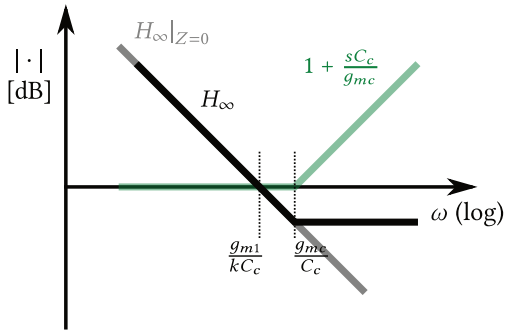


Figure B.17: 1-EET decomposition of 1-GFT H_∞. The input resistance of the current amplifier introduces a LHP zero.

EET Decomposition of T

The 1-EET decomposition of T (condition $v_i = 0$) is:

$$T = T|_{Z=0}\,\frac{1 + \frac{Z}{Z_\mathrm{n}}}{1 + \frac{Z}{Z_\mathrm{d}}} \tag{B.26}$$

Once again, $T|_{Z=0}$ is already computed and shown in Fig. B.13. The null driving point impedance Z_n, is calculated with $v_i = 0$ (condition for T) and $v_y = 0$ (output nulled). It follows that $i_c = 0$, which is the excitation current of Z_n, for any possible voltage. Hence $Z_\mathrm{n} = \infty$.

For the driving point impedance, the excitation source for T is set to zero: $v_x = 0$. Together with $v_i = 0$ (condition for T), the circuit of Fig. B.18 remains. Z_d is now easily derived:

$$Z_\mathrm{d} = \frac{1 + sR_2(C_c + C_2)}{sC_c(1 + sR_2C_2)} \tag{B.27}$$

$$\approx \frac{1 + sR_2C_c}{sC_c(1 + sR_2C_2)} \tag{B.28}$$

in which the approximation holds for $C_c \gg C_2$. This yields:

$$\frac{Z}{Z_\mathrm{d}} \approx \frac{sC_c(1 + sR_2C_2)}{g_{mc}(1 + sR_2C_c)} \tag{B.29}$$

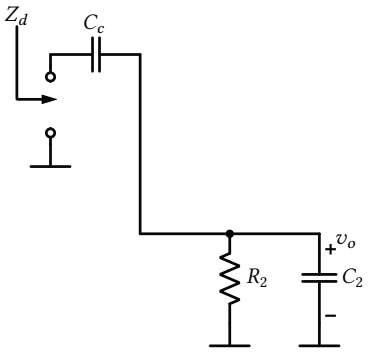


Figure B.18: Circuit diagram to calculate the driving point impedance Z_d of the 1-EET decomposition of the 1-GFT loop gain T.

The factor $(1 + Z/Z_\mathrm{d})^{-1}$ will be approximated graphically. Consider Fig. B.19. In Eq. (B.29), as $C_c \gg C_2$, the pole is lower than zero. Also, assume $g_{mc}R_2 \gg 1$, then the pole is also lower than the unity-gain frequency of the ideal differentiator. This means that Eq. (B.29) can be drawn as the black line in the figure. The location of the pole and zero can be interchanged. It follows that the factor is approximated as a single-pole low-pass filter:

$$\left(1 + \frac{Z}{Z_\mathrm{d}}\right)^{-1} \approx \frac{1}{1 + \frac{sC_c}{g_{mc}}} \tag{B.30}$$

Hence, compared to the configuration with zero input resistance, this analysis shows that an extra pole is introduced in the loop gain, as shown in Fig. B.20.

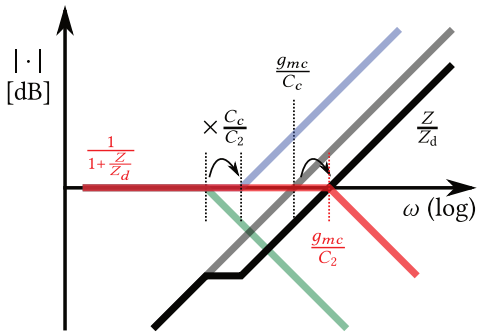


Figure B.19: Graphical approximation of the factor $(1 + Z/Z_\mathrm{d})^{-1}$ of the 1-EET decomposition of the 1-GFT loop gain T.

Discrepancy Factor D of H

After this preparatory work, we can now calculate the discrepancy factor D of the 1-GFT decomposition *of the input-output transfer function* H, using the T, derived when input non-zero input resistance is included in the model. This is shown in Fig. B.21. The two normal poles of T result in a complex pole pair in D, which can be specified by the natural

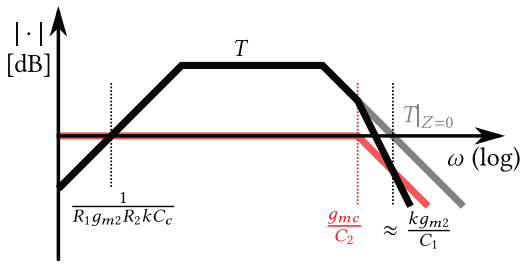


Figure B.20: 1-EET decomposition of 1-GFT T. An extra pole is introduced in the loop gain.

frequency and Q-factor:

$$\omega_n = \sqrt{\frac{kg_{m2}}{C_1}\frac{g_{mc}}{C_2}} \tag{B.31}$$

$$Q = \sqrt{\frac{kg_{m2}}{g_{mc}}\frac{C_2}{C_1}} \tag{B.32}$$

A design constraint on the stability of the internal loop can be expressed as a maximum Q-factor.

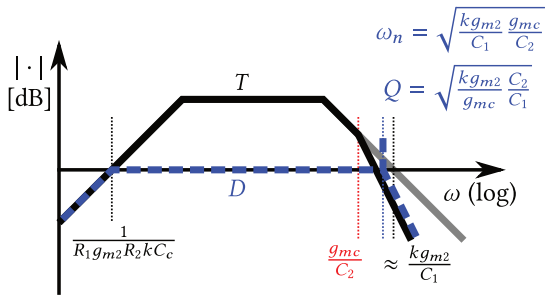


Figure B.21: Discrepancy factor of the 1-GFT decomposition of the input-output transfer function of a two-stage opamp with multiplied Miller capacitance. The two poles in the loop gain can introduce peaking.

B.3.1 Complete Model

Figure B.22 shows how the complete transfer function is assembled from its components. This Bode magnitude plot with its associated frequencies represents a usable model of a two-stage opamp with multiplied Miller capacitance given the following assumptions:

$$C_c \gg C_1, C_2 \tag{B.33}$$

$$g_{m2}(R_1 + R_2) \gg 1 \tag{B.34}$$

$$g_{mc}R_2 \gg 1 \tag{B.35}$$

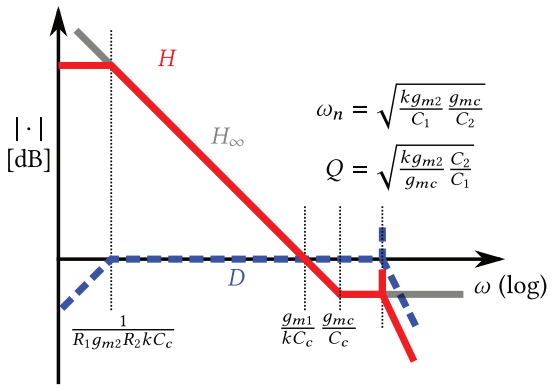


Figure B.22: 1-GFT decomposition of the input-output transfer function of a two-stage opamp with multiplied Miller capacitance. This model can be used for design.

In case this opamp is used in a feedback system, the resulting system is a multiple-loop feedback system[2]. In general, stability analysis of such kind of systems is not trivial as the detailed treatment given in [96] reveals. For this particular system, the procedure can be summarized as follows. First, check stability of the internal loop with the outer loop disabled ($g_{m1} = 0$). This is exactly what has been done during the loop gain calculation in the previous section. Second, restore the outer loop and perform stability analysis on *its* loop gain, which is wholly of partly given by H. In the simulator, a nested GFT can perform these steps numerically. It is interesting that the order of which the loops are checked can be reversed: first check the outer loop with the interal one disabled, then the internal loop (with the outer loop still active). This, however, is not in line with the analytical treatment given above.

B.3.2 Internal Loop Compensation

The analysis of the loop gain (Fig. B.20) gives hints as how to compensate the internal loop, as shown in Fig. B.23. A first option is to move low-frequency pole down by decreasing g_{m2} or increasing C_1. This will reduce the bandwidth of the closed loop (Eq. (B.31)). A second option is to increase g_{mc} and decrease C_2 in order to speed up the high-frequency pole. However, when $C_1 \approx C_2$, g_{mc} must be higher than kg_{m2}, which might be difficult. In both cases the Q-factor of D (Eq. (B.32)) will decrease, indicating the stabilizing effect of pole separation.

B.4 Conclusion

A model for a two-stage Miller-compensated amplifier with capacitance multipliers is derived. Classic Miller compensation was analyzed from a feedback perspective with the GFT. A current buffer was inserted to avoid the RHP zero. Gradually, the model was improved using techniques such as the EET and graphical approximations. In Chapter 5,

[2] In a single-loop feedback system the return difference $(1+T)$ with respect to the controlling parameter of any active device is equal to unity if the controlling parameter of any other active device vanishes [128].

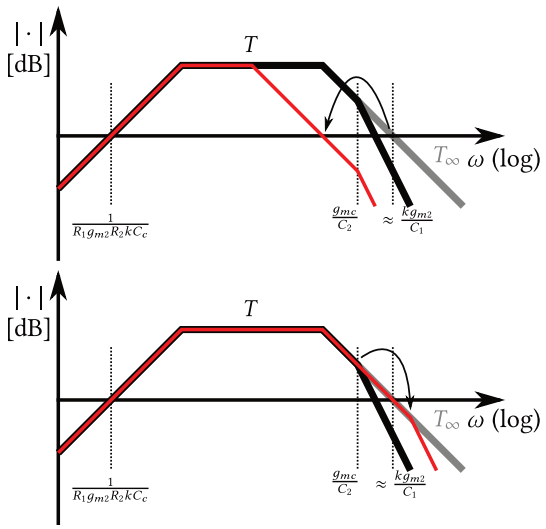


Figure B.23: Possible scenarios for compensation of the internal loop.

the result was used to design the loop filter of the AGC system: with a current gain $k = 8$, the compensation capacitance was reduced from 240 pF to 30 pF.

List of Publications

Publications in International Journals

- J. Put, X. Yin, X.-Z. Qiu, J. Gillis, **J. Verbrugghe**, J. Bauwelinck, J. Vandewege, F. Blache, D. Lanteri, M. Achouche, H.-G. Krimmel, D. van Veen, and P. Vetter, "DC-coupled burst-mode receiver with high sensitivity, wide dynamic range and short settling time for symmetric 10G-GPONs," *Optics Express*, vol. 19, no. 26, p. B594, Dec. 2011.

- X. Yin, J. Put, **J. Verbrugghe**, J. Gillis, X.-Z. Qiu, J. Bauwelinck, J. Vandewege, H.-G. Krimmel, and M. Achouche, "A 10 Gb/s burst-mode TIA with on-chip reset/lock CM signaling detection and limiting amplifier with a 75 ns settling time," in *2012 IEEE International Solid-State Circuits Conference*. IEEE, Feb. 2012, pp. 416–418.

- X. Yin, B. Moeneclaey, X.-Z. Qiu, **J. Verbrugghe**, K. Verheyen, J. Bauwelinck, J. Vandewege, M. Achouche, and Y. Chang, "A 10Gb/s APD-based linear burst-mode receiver with 31dB dynamic range for reach-extended PON systems." *Optics express*, vol. 20, no. 26, pp. B462–9, Dec. 2012.

- X. Yin, X.-Z. Qiu, J. Gillis, J. Put, **J. Verbrugghe**, J. Bauwelinck, J. Vandewege, H. Krimmel, D. van Veen, P. Vetter, and F. Chang, "Experiments on a 10 Gb/s Fast-Settling High-Sensitivity Burst-Mode Receiver With On-Chip Auto-Reset for 10G-GPONs [Invited]," *Journal of Optical Communications and Networking*, vol. 4, no. 11, p. B68, Oct. 2012.

- **J. Verbrugghe**, B. Schrenk, J. Bauwelinck, X. Yin, S. Dris, J. A. Lazaro, V. Katopodis, P. Bakopoulos, and H. Avramopoulos, "Quaternary TDM-PAM as upgrade path of access PON beyond 10Gb/s." *Optics express*, vol. 20, no. 26, pp. B15–20, Dec. 2012.

- R. Vaernewyck, J. Bauwelinck, X. Yin, R. Pierco, **J. Verbrugghe**, G. Torfs, Z. Li, X.-Z. Qiu, J. Vandewege, R. Cronin, A. Borghesani, and D. Moodie, "113Gb/s (10 x 11.3Gb/s) ultra-low power EAM driver array," *Optics express*, vol. 21, no. 1, pp. 256–62, Jan. 2013.

- R. Vaernewyck, X. Yin, **J. Verbrugghe**, G. Torfs, X.-Z. Qiu, E. Kehayas, and J. Bauwelinck, "A Low Power 2x28Gb/s Electroabsorption Modulator Driver Array with On-Chip Duobinary Encoding," *IEICE Transactions on Communications*, vol. E97.B, no. 8, pp. 1623–1629, 2014.

- X.-Z. Qiu, X. Yin, **J. Verbrugghe**, B. Moeneclaey, A. Vyncke, C. Van Praet, G. Torfs, J. Bauwelinck, and J. Vandewege, "Fast Synchronization 3R Burst-Mode Receivers for Passive Optical Networks," *Journal of Lightwave Technology*, vol. 32, no. 4, pp. 644–659, Feb. 2014.
- **J. Verbrugghe**, R. Vaernewyck, B. Moeneclaey, X. Yin, G. Maxwell, R. Cronin, G. Torfs, X.-Z. Qiu, C. P. Lai, P. Townsend, and J. Bauwelinck, "Multichannel 25 Gb/s Low-Power Driver and Transimpedance Amplifier Integrated Circuits for 100 Gb/s Optical Links," *Journal of Lightwave Technology*, vol. 32, no. 16, pp. 2877–2885, Aug. 2014.

Publications in International Conferences

- **J. Verbrugghe**, M. De Bock, and P. Rombouts, "Design of an integrated analog controller for a class-D audio amplifier," in *2009 16th IEEE International Conference on Electronics, Circuits and Systems - (ICECS 2009)*. IEEE, Dec. 2009, pp. 727–730.
- X. Yin, X. Qiu, J. Gillis, J. Put, **J. Verbrugghe**, J. Bauwelinck, J. Vandewege, F. Blache, D. Lanteri, M. Achouche, H. Krimmel, D. van Veen, and P. Vetter, "DC-coupled burst-mode receiver with high sensitivity, wide dynamic range and short settling time for symmetric 10G-GPONs," in *Optical Communication (ECOC), 2011 37th European Conference and Exhibition on*, 2011.
- X. Yin, B. Moeneclaey, X.-Z. Qiu, **J. Verbrugghe**, K. Verheyen, J. Bauwelinck, J. Vandewege, M. Achouche, and Y. Chang, "A 10Gb/s APD-based linear burst-mode receiver with 31dB dynamic range for reach-extended PON systems," in *38th European Conference and Exhibition on Optical Communication (ECOC - 2012)*, Amsterdam, 2012.
- R. Vaernewyck, J. Bauwelinck, X. Yin, R. Pierco, **J. Verbrugghe**, G. Torfs, Z. Li, X. Qiu, J. Vandewege, R. Cronin, A. Borghesani, and D. Moodie, "A 113 Gb/s (10 x 11.3 Gb/s) ultra-low power EAM driver array," in *38th European Conference and Exhibition on Optical Communication, Proceedings*. Amsterdam: IEEE, 2012, pp. 1–3.
- C. Lai, A. Naughton, P. Ossieur, P. Townsend, D. Smith, A. Borghesani, D. Moodie, G. Maxwell, J. Bauwelinck, R. Vaernewyck, **J. Verbrugghe**, X. Yin, X. Qiu, M. Eiselt, K. Grobe, N. Parsons, R. Jensen, and E. Kehayas, "Energy-efficient colourless photonic technologies for next-generation DWDM metro and access networks," in *Photonics in Switching (PS), 2012 International Conference on*. IEEE, 2012, pp. 1–3.
- X. Yin, X.-Z. Qiu, J. Gillis, J. Put, **J. Verbrugghe**, J. Bauwelinck, J. Vandewege, H. Krimmel, D. van Veen, P. Vetter, and Y. F. Chang, "Experiments on 10Gb/s fast settling high sensitivity burst-mode receiver with on-chip auto-reset for 10G-GPONs," in *2012 Conference on optical fiber communication, collocated National Fiber Optic Engineers conference (OFC/NFOEC 2012)*. IEEE, 2012, p. 3.
- **J. Verbrugghe**, B. Moeneclaey, and J. Bauwelinck, "Implementation of the dissection theorem in cadence virtuoso," in *2012 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*. IEEE, Sep. 2012, pp. 145–148.

- D. Chiaroni, G. de Valicourt, M. Achouche, F. Blache, X. Yin, X.-Z. Qiu, **J. Verbrugghe**, and J. Bauwelinck, "Power dynamics control in optical packet rings," in *Photonics in Switching 2012, Proceedings*, 2012, p. 3.

- S. Dris, J. Bauwelinck, X. Yin, B. Schrenk, J. Lazaro, V. Katopodis, P. Bakopoulos, **J. Verbrugghe**, and H. Avramopoulos, "Quaternary TDM-PAM and its Implications for TDMA Equipment," in *European Conference and Exhibition on Optical Communication*. Washington, D.C.: OSA, 2012, p. Tu.3.B.3.

- J. Bauwelinck, R. Vaernewyck, **J. Verbrugghe**, W. Soenen, B. Moeneclaey, C. Van Praet, A. Vyncke, G. Torfs, X. Yin, X. Qiu, J. Vandewege, N. Sotiropoulos, H. de Waardt, R. Cronin, G. Maxwell, T. Tekin, P. Bakopoulos, C. Lai, and P. Townsend, "High-Speed Electronics for Short-Link Communication," in *39th European Conference and Exhibition on Optical Communication (ECOC 2013)*. Institution of Engineering and Technology, 2013, pp. 164–166.

- R. Vaernewyck, **J. Verbrugghe**, W. Soenen, B. Moeneclaey, G. Torfs, X. Yin, and J. Bauwelinck, "High-speed electronic integrated circuits for metro, access and data center networks," in *Workshop Optical Interconnect in Data Centers, Abstracts*, Berlin, 2014, p. 1.

- **J. Verbrugghe**, B. Moeneclaey, and J. Bauwelinck, "Integration of the General Network Theorem in ADE and ADE XL: Toward a Deeper Insight Into Circuit Behavior," in *Proceedings of CDNLive! EMEA*, 2014.

- X. Yin, G. Torfs, R. Vaernewyck, A. Vyncke, **J. Verbrugghe**, B. Moeneclaey, M. Ruffini, D. B. Payne, and J. Bauwelinck, "Performance evaluation of single carrier 40-Gbit/s downstream for long-reach passive optical network," in *18th International Conference on Optical Networking Design and Modeling, Proceedings*. IEEE, 2014, pp. 162–167.

Bibliography

[1] D. Deckmyn, "Netflix start vandaag in Vlaanderen," Sep. 2014. [Online]. Available: http://www.standaard.be/cnt/dmf20140918_01276061

[2] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.

[3] J. Baliga, R. Ayre, W. V. Sorin, K. Hinton, and R. S. Tucker, "Energy Consumption in Access Networks," in *OFC/NFOEC 2008 - 2008 Conference on Optical Fiber Communication/National Fiber Optic Engineers Conference*. IEEE, Feb. 2008, pp. 1–3.

[4] J. Baliga, R. W. a. Ayre, K. Hinton, and R. S. Tucker, "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 149–167, Jan. 2011.

[5] D. Coleman, "10G and Beyond in the Data Center with OM3 Fiber," *Cabling Business Magazine*, pp. 32–35, Apr. 2008.

[6] B. Hayes, "Cloud computing," *Communications of the ACM*, vol. 51, no. 7, p. 9, Jul. 2008.

[7] B. Raghavan and J. Ma, "The energy and emergy of the internet," in *Proceedings of the 10th ACM Workshop on Hot Topics in Networks - HotNets '11*. New York, New York, USA: ACM Press, 2011, pp. 1–6.

[8] L. Valcarenghi, I. Cerutti, and P. Castoldi, "Energy efficient optical access and metro networks," in *2010 12th International Conference on Transparent Optical Networks*. IEEE, Jun. 2010, pp. 1–4.

[9] J. Whitney and P. Delforge, "Data Center Efficiency Assessment," National Resources Defence Council, Tech. Rep. August, 2014.

[10] G. E.-s. Initiative, "GeSI SMARTer 2020: The Role of ICT in Driving a Sustainable Future," GeSi, Tech. Rep., 2012.

[11] M. F. Bari, R. Boutaba, R. Esteves, L. Z. Granville, M. Podlesny, M. G. Rabbani, Q. Zhang, and M. F. Zhani, "Data Center Network Virtualization: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 909–928, Jan. 2013.

[12] A. Greenberg, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta, "Towards a next generation data center architecture," in *Proceedings of the ACM workshop on Programmable routers for extensible services of tomorrow - PRESTO '08*. New York, New York, USA: ACM Press, 2008, p. 57.

[13] Cisco, "Cisco Global Cloud Index : Forecast and Methodology, 2013-2018," Cisco, Tech. Rep., 2013.

[14] ——, "Cisco Visual Networking Index: Forecast and Methodology, 2013-2018," Cisco, Tech. Rep., 2014.

[15] J. G. Koomey, *Estimating total power consumption by servers in the US and the world*. February, 2007, vol. 15.

[16] "Green Grid Metrics: Describing Datacenter Power Efficiency," The Green Grid, Tech. Rep., 2007.

[17] J. G. Koomey, "Growth in data center electricity use 2005 to 2010," Analytics Press, Oakland, CA, Tech. Rep., 2011.

[18] E. Säckinger, *Broadband Circuits for Optical Fiber Communication*. John Wiley & Sons, 2005.

[19] H. Liu, C. F. Lam, and C. Johnson, "Scaling Optical Interconnects in Datacenter Networks Opportunities and Challenges for WDM," in *2010 18th IEEE Symposium on High Performance Interconnects*. IEEE, Aug. 2010, pp. 113–116.

[20] J. E. Cunningham, D. Beckman, D. Huang, T. Sze, K. Cai, and A. V. Krishnamoorthy, "PAM-4 signaling over VCSELs using 0.13 um CMOS," in *Information Photonics, 2005. IP 2005. OSA Topical Meeting on*, Jun. 2005, pp. 1–3.

[21] M. Maier, *Optical Switching Networks*. Cambridge: Cambridge University Press, 2008.

[22] ITU-T, "ITU-T G.694.1: Spectral grids for WDM applications: DWDM frequency grid," ITU-T, Tech. Rep., 2012.

[23] T. McConaghy, K. Breen, J. Dyck, and A. Gupta, *Variation-Aware Design of Custom Integrated Circuits: A Hands-on Field Guide*. Springer Science & Business Media, 2012.

[24] R. Middlebrook, "Low-entropy expressions: The key to design-oriented analysis," *Frontiers in Education Conference*, pp. 399–403, 1991.

[25] D. Binkley, M. Bucher, and D. Foty, "Design-oriented characterization of CMOS over the continuum of inversion level and channel length," in *ICECS 2000. 7th IEEE International Conference on Electronics, Circuits and Systems (Cat. No.00EX445)*, vol. 1. IEEE, 2000, pp. 161–164.

[26] D. Binkley, *Tradeoffs and Optimization in Analog CMOS Design*. John Wiley & Sons, 2008.

[27] C. Enz, F. Krummenacher, and E. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-current applications," *Analog Integrated Circuits and Signal Processing Journal on Low-Voltage and Low-Power Design*, vol. 8, no. July, pp. 83–114, 1995.

[28] S. C. Yogesh, A. Niknejad, and C. Hu, "BSIM Compact MOSFET Models," in *Nano-Terra Workshop*, 2011.

[29] W. M. C. Sansen, *Analog Design Essentials*. Springer, 2006.

[30] W. Sansen, "Analog design procedures for channel lengths down to 20 nm," in *2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS)*, no. 2. IEEE, Dec. 2013, pp. 337–340.

[31] P. R. Gray, P. J. Hurst, S. H. Lewis, and R. G. Meyer, *Analysis and design of analog integrated circuits*, 5th ed. John Wiley & Sons, Incorporated, 2010.

[32] R. Middlebrook, "Teaching design-oriented analysis at the first level," in *Proceedings of IEEE Frontiers in Education Conference - FIE '93*. IEEE, 1993, pp. 20–21.

[33] ——, "Structured Analog Design," 2004. [Online]. Available: http://rdmiddlebrook.com/D_OA_Rules&Tools/index.asp

[34] F. Langford-Smith, *Radiotron Designer's Handbook*, 4th ed. Amalgamated Wireless Valve Company Pty., 1953.

[35] J. Millman and C. C. Halkias, *Integrated Electronics*. McGraw-Hill, 1972.

[36] G. Avenier, M. Diop, P. Chevalier, G. Troillard, N. Loubet, J. Bouvier, L. Depoyan, N. Derrier, M. Buczko, C. Leyris, S. Boret, S. Montusclat, A. Margain, S. Pruvost, S. T. Nicolson, K. H. K. Yau, N. Revil, D. Gloria, D. Dutartre, S. P. Voinigescu, and A. Chantre, "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, Sep. 2009.

[37] A. Pawlikiewicz and S. Elrai, "RF CMOS or SiGE BiCMOS in RF and Mixed Signal Circuit Design," in *2007 14th International Conference on Mixed Design of Integrated Circuits and Systems*, no. March. IEEE, Jun. 2007, pp. 333–338.

[38] Y. Ota, R. Swartz, V. Archer, S. Korotky, M. Banu, and A. Dunlop, "High-speed, burst-mode, packet-capable optical receiver and instantaneous clock recovery for optical bus operation," *Journal of Lightwave Technology*, vol. 12, no. 2, pp. 325–331, 1994.

[39] P. Ossieur, T. De Ridder, X. Qiu, and J. Vandewege, "Influence of random DC offsets on burst-mode receiver sensitivity," *Journal of Lightwave Technology*, vol. 24, no. 3, pp. 1543–1550, Mar. 2006.

[40] J. Redd and C. Lyon, "Spectral content of NRZ test patterns," *Maxim Application Note 3455*, pp. 10–14, 2004.

[41] T. Ono, Y. Yano, K. Fukuchi, T. Ito, H. Yamazaki, M. Yamaguchi, and K. Emura, "Characteristics of optical duobinary signals in terabit/s capacity, high-spectral efficiency WDM systems," *Journal of Lightwave Technology*, vol. 16, no. 5, pp. 788–797, May 1998.

[42] H. Shankar, "Duobinary modulation for optical systems," 2002. [Online]. Available: https://www.inphi.com/products/whitepapers/DuobinaryModulationForOpticalSystems.pdf

[43] K. Szczerba, P. Westbergh, E. Agrell, M. Karlsson, P. a. Andrekson, and A. Larsson, "Comparison of Intersymbol Interference Power Penalties for OOK and 4-PAM in Short-Range Optical Links," *Journal of Lightwave Technology*, vol. 31, no. 22, pp. 3525–3534, Nov. 2013.

[44] P. Winzer and R.-J. Essiambre, "Advanced Optical Modulation Formats," *Proceedings of the IEEE*, vol. 94, no. 5, pp. 952–985, May 2006.

[45] A. X. Widmer and P. A. Franaszek, "A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code," *IBM Journal of Research and Development*, vol. 27, no. 5, pp. 440–451, Sep. 1983.

[46] S. D. Personick, "Receiver Design for Digital Fiber Optic Communication Systems, I," *Bell System Technical Journal*, vol. 52, no. 6, pp. 843–874, Jul. 1973.

[47] G. Vasilescu, *Electronic noise and interfering signals: principles and applications*. Berlin: Springer-Verlag, 2005.

[48] T. Schmerbeck, *A survey of isolation techniques for mixed-signal ICs*. MEAD Education SA, 2012.

[49] A. Hastings, *The Art of Analog Layout*, 2006.

[50] G. Agrawal, *Fiber-optic communication systems*, 2nd ed. New York: John Wiley & Sons, 1997.

[51] P. Muller and Y. Leblebici, *CMOS Multichannel Single-Chip Receivers for Multi-Gigabit Optical Data Communications*, ser. Analog Circuits and Signal Processing. Dordrecht: Springer Netherlands, 2007.

[52] J. P. Alegre Pérez, S. C. Pueyo, and B. C. López, *Automatic Gain Control*. New York, NY: Springer New York, 2011.

[53] W. Victor and M. Brockman, "The Application of Linear Servo Theory to the Design of AGC Loops," *Proceedings of the IRE*, vol. 48, no. 2, pp. 234–238, Feb. 1960.

[54] E. Bertran and J. Palacin, "Control theory applied to the design of AGC circuits," in *[1991 Proceedings] 6th Mediterranean Electrotechnical Conference*. IEEE, 1991, pp. 66–70.

[55] D. Mercy, "A review of automatic gain control theory," *Radio and Electronic Engineer*, vol. 51, no. 11-12, p. 579, 1981.

[56] T. B. Mills and H. S. Suzuki, "Design Concepts for Low-Cost Transistor AGC Systems," *IEEE Transactions on Broadcast and Television Receivers*, vol. 14, no. 3, pp. 38–43, 1968.

[57] A. Isidori, *Nonlinear Control Systems, Volume 1*, 1995.

[58] A. Gelb and W. E. V. Velde, *Multiple-input describing functions and nonlinear system design*. McGraw-Hill, 1968.

[59] J. Hauptmann, F. Dielacher, R. Steiner, C. Enz, and F. Krummenacher, "A low-noise amplifier with automatic gain control and anticlipping control in CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 27, no. 7, pp. 974–981, Jul. 1992.

[60] H. B. Jr, J. Khoury, and T. Viswanathan, "Variable gain voltage signal amplifier," May 1995. [Online]. Available: http://www.google.com/patents/US5412346

[61] R. Harjani, "A low-power CMOS VGA for 50 Mb/s disk drive read channels," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 42, no. 6, pp. 370–376, Jun. 1995.

[62] D. Welland, S. Phillip, K. Leung, G. Tuttle, S. Dupuie, D. Holberg, R. Jack, N. Sooch, K. Anderson, A. Armstrong, R. Behrens, W. Bliss, T. Dudley, W. Foland, N. Glover, and L. King, "A digital read/write channel with EEPR4 detection," in *Proceedings of IEEE International Solid-State Circuits Conference - ISSCC '94*. IEEE, 1994, pp. 276–277.

[63] J. Khoury, "On the design of constant settling time AGC circuits," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 3, pp. 283–294, Mar. 1998.

[64] B. Murmann and B. Boser, "Digitally Assisted Analog Integrated Circuits," *Queue*, vol. 2, no. 1, p. 64, Mar. 2004.

[65] B. Murmann, C. Vogel, and H. Koeppl, "Digitally enhanced analog circuits: System aspects," in *2008 IEEE International Symposium on Circuits and Systems*. IEEE, May 2008, pp. 560–563.

[66] J. Mark and T. Todd, "A Nonuniform Sampling Approach to Data Compression," *IEEE Transactions on Communications*, vol. 29, no. 1, pp. 24–32, Jan. 1981.

[67] C. Weltin-Wu and Y. Tsividis, "An Event-driven Clockless Level-Crossing ADC With Signal-Dependent Adaptive Resolution," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 9, pp. 2180–2190, Sep. 2013.

[68] Y. Tsividis, "Event-driven data acquisition and continuous-time digital signal processing," in *IEEE Custom Integrated Circuits Conference 2010*. IEEE, Sep. 2010, pp. 1–8.

[69] ——, "Continuous-time digital signal processing," *Electronics Letters*, vol. 39, no. 21, p. 1551, Oct. 2003.

[70] M. Abramowitz and I. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, 1972.

[71] Y. Zhou and G. Huang, "A novel wide-band envelope detector," in *2008 IEEE Radio Frequency Integrated Circuits Symposium*. IEEE, Jun. 2008, pp. 219–222.

[72] C. D. Presti, F. Carrara, A. Scuderi, and G. Palmisano, "Fast Peak Detector with Improved Accuracy and Linearity for High-Frequency Waveform Processing," in *2007 IEEE International Symposium on Circuits and Systems*. IEEE, May 2007, pp. 3884–3887.

[73] J. Alegre, S. Celma, M. Sanz, and P. Martinez, "A Low-Ripple Fast-Settling CMOS Envelope Detector," in *2006 Ph.D. Research in Microelectronics and Electronics*. IEEE, 2006, pp. 9–12.

[74] J.-H. Kim, J.-B. Shin, J.-Y. Sim, and H.-J. Park, "5-Gb/s Peak Detector Using a Current Comparator and a Three-State Charge Pump," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 58, no. 5, pp. 269–273, May 2011.

[75] R. Meyer, "Low-power monolithic RF peak detector analysis," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 1, pp. 65–67, 1995.

[76] D. Sheingold, *Nonlinear circuits handbook: Designing with analog function modules and IC's*. Analog Devices, 1976.

[77] W. R. Bennett, "Response of a Linear Rectifier to Signal and Noise*," *Bell System Technical Journal*, vol. 23, no. 1, pp. 97–113, Jan. 1944.

[78] J. Huijsing, *Operational amplifiers: theory and design*, 1st ed. Springer Publishing Company, Incorporated, May 2001.

[79] M. Pilotte, "AN-691 Operation of RF Detector Products at Low Frequency," pp. 1–16, 2005.

[80] Texas Instruments, "AN-311 Theory and Applications of Logarithmic Amplifiers," no. April, pp. 1–10, 1991.

[81] ——, "LOG114: single-supply, high-speed, precision logarithmic amplifier," 2007.

[82] G. Rincon-Mora, "Active capacitor multiplier in Miller-compensated circuits," *IEEE Journal of Solid-State Circuits*, vol. 35, no. 1, pp. 26–32, Jan. 2000.

[83] H. Inose, T. Aoki, and K. Watanabe, "Asynchronous delta-modulation system," *Electronics Letters*, vol. 2, no. 3, p. 95, 1966.

[84] B. Schell and Y. Tsividis, "A Continuous-Time ADC/DSP/DAC System With No Clock and With Activity-Dependent Power Dissipation," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 11, pp. 2472–2481, Nov. 2008.

[85] R. J. Baker, *CMOS Circuit Design, Layout, and Simulation, Second Edition*. Wiley-IEEE Press, Oct. 2004.

[86] K. Kundert, "A Test Bench for Differential Circuits," *Available from www.designers-guide.org*, pp. 1–7, 2002.

[87] IEEE Computer Society, "IEEE Standard for SystemVerilog - Unified Hardware Design, Specification, and Verification Language," Tech. Rep., 2013.

[88] S. Sutherland, S. Davidmann, and P. Flake, *SystemVerilog for Design Second Edition: A Guide to Using SystemVerilog for Hardware Design and Modeling*, 2006.

[89] C. Spear and G. Tumbush, *SystemVerilog for Verification: A Guide to Learning the Testbench Language Features*. Springer Science & Business Media, 2012.

[90] C. Chong and K. Smith, "A CMOS current comparator with well-controlled hysteresis," in *Proceedings of the 33rd Midwest Symposium on Circuits and Systems*. IEEE, 1990, pp. 269–272.

[91] K. Smith and A. Sedra, "The current conveyor-A new circuit building block," *Proceedings of the IEEE*, vol. 56, no. 8, pp. 1368–1369, 1968.

[92] T. Banwell, "Bipolar transistor circuit analysis using the Lambert W-function," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 11, pp. 1621–1633, 2000.

[93] B. Razavi, *Design of analog CMOS integrated circuits*. McGraw-Hill, 2002.

[94] F. Langford-Smith, *Radio Designer's Handbook*. Iliffe, 1967.

[95] J. Proesel, C. Schow, and A. Rylyakov, "25Gb/s 3.6pJ/b and 15Gb/s 1.37pJ/b VCSEL-based optical links in 90nm CMOS," in *2012 IEEE International Solid-State Circuits Conference*. IEEE, Feb. 2012, pp. 418–420.

[96] H. Bode, *Network analysis and feedback amplifier design*. D. Van Nostrand Co., Inc., 1945.

[97] E. Cherry and D. Hooper, "The design of wide-band transistor feedback amplifiers," *Proceedings of the Institution of Electrical Engineers*, vol. 110, no. 2, p. 375, 1963.

[98] W. Pohlmann, "A silicon-bipolar amplifier for 10 Gbit/s with 45 dB gain," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 5, pp. 551–556, May 1994.

[99] M. Alioto and G. Palumbo, *Model and design of bipolar and MOS current-mode logic*, 2005.

[100] J. Babcock, B. Loftin, P. Madhani, A. Pinto, and D. Schroder, "Comparative low frequency noise analysis of bipolar and MOS transistors using an advanced complementary BiCMOS technology," in *Proceedings of the IEEE 2001 Custom Integrated Circuits Conference (Cat. No.01CH37169)*. IEEE, 2001, pp. 385–388.

[101] J. Verbrugghe, B. Moeneclaey, and J. Bauwelinck, "Integration of the General Network Theorem in ADE and ADE XL: Toward a Deeper Insight Into Circuit Behavior," in *Proceedings of CDNLive! EMEA*, 2014.

[102] "website of R. D. Middlebrook." [Online]. Available: http://www.rdmiddlebrook.com/

[103] D. Peter, "We Can Do Better: A Proven, Intuitive, Efficient And Practical Design Oriented Circuit Analysis Paradigm Is Available, So Why Aren't We Using It To Teach Our Students?" in *ASEE Annual Conference & Exposition*, 2007.

[104] R. Middlebrook, "Applications notes - The general feedback theorem: a final solution for feedback systems," *IEEE Microwave Magazine*, vol. 7, no. 2, pp. 50–63, Apr. 2006.

[105] ——, "Null double injection and the extra element theorem," *IEEE Transactions on Education*, vol. 32, no. 3, pp. 167–180, 1989.

[106] R. Middlebrook, V. Vorperian, and J. Lindal, "The N Extra Element Theorem," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 45, no. 9, pp. 919–935, 1998.

[107] R. Middlebrook, "The DT and the CT: The Dissection Theorem and the Chain Theorem."

[108] J. Verbrugghe, B. Moeneclaey, and J. Bauwelinck, "Implementation of the dissection theorem in cadence virtuoso," in *2012 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*. IEEE, Sep. 2012, pp. 145–148.

[109] B. Moeneclaey and J. Verbrugghe, "On The General Network Theorem," in *Internal Communication*, 2014.

[110] R. D. Middlebrook, "GFT Asymptotic Loop Gain H0," 2008.

[111] D. Maclean, "Improved methods of assessing feedback in wide-band multiloop amplifiers," *IEEE Transactions on Circuits and Systems*, vol. 27, no. 9, pp. 779–792, Sep. 1980.

[112] R. D. Middlebrook, "Topics in Multiple-Loop Regulators and Current-Mode Programming," *IEEE Transactions on Power Electronics*, vol. PE-2, no. 2, pp. 109–124, Apr. 1987.

[113] F. Wiedmann, "Loop gain simulation." [Online]. Available: http://sites.google.com/site/frankwiedmann/loopgain

[114] 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.

[115] R. Middlebrook, "Measurement of loop gain in feedback systems," *International Journal of Electronics*, vol. 38, no. 4, pp. 485–512, Apr. 1975.

[116] ——, "Design-oriented analysis of feedback amplifiers," *Proceedings National Electronics Conference*, vol. 20, pp. 234–238, 1964.

[117] C. Toumazou, *Trade-Offs in Analog Circuit Design: The Designer's Companion*. Springer, 2004.

[118] S. Rosenstark, *Feedback amplifier principles*. New Jersey: MacMillan Publishing Company, 1986.

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

[120] S. Mason, "Feedback Theory-Some Properties of Signal Flow Graphs," *Proceedings of the IRE*, vol. 41, no. 9, pp. 1144–1156, Sep. 1953.

[121] S. Pennisi, "CMOS multiplier for grounded capacitors," *Electronics Letters*, vol. 38, no. 15, p. 765, 2002.

[122] I. Hwang, "Area-efficient and self-biased capacitor multiplier for on-chip loop filter," *Electronics Letters*, vol. 42, no. 24, p. 1392, 2006.

[123] W. Aloisi, G. Palumbo, and S. Pennisi, "Design methodology of Miller frequency compensation with current buffer/amplifier," *IET Circuits, Devices & Systems*, vol. 2, no. 2, p. 227, 2008.

[124] W.-k. Chen, *Circuit Analysis and Feedback Amplifier Theory*, 2006.

[125] A. B. Grebene, *Bipolar and MOS analog integrated circuit design.* Wiley-Interscience, 2002, vol. 2002.

[126] A. Hajimiri, "Generalized Time- and Transfer-Constant Circuit Analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 6, pp. 1105–1121, Jun. 2010.

[127] V. Vorpérian, *Fast analytical techniques for electrical and electronic circuits*, 2002.

[128] W.-K. Chen, *Active network analysis.* World Scientific Publishing Co., Inc., Jan. 1991.