5 GHz Optical Front End in $0.35\mu m$ CMOS

By Mengxiong Li, MSc.

Thesis submitted to The University of Nottingham for the degree of
Doctor of Philosophy

Oct 2007

ABSTRACT

With the advantages of low cost, low power consumption, high reliability and potential for large scale integration, CMOS monolithically integrated active pixel chips have significant application in optical sensing systems. The optical front end presented in this thesis will have application in Optical Scanning Acoustic Microscope System (O-SAM), which involves a totally non-contact method of acquiring images of the interaction between surface acoustic waves (SAWs) and a solid material to be characterized.

In this work, an ultra fast optical front-end using improved regulated cascade scheme is developed based on AMS $0.35\mu m$ CMOS technology. The receiver consists of an integrated photodiode, a transimpedance amplifier, a mixer, an IF amplifier and an output buffer. By treating the n-well in standard CMOS technology as a screening terminal to block the slow photo-generated bulk carriers and interdigitizing shallow $p+$ junctions as the active region, the integrated photodiode operates up to 4.9 GHz with no process modification. Its responsivity was measured to be 0.016 A/W. With multi-inductive-series peaking technique, the improved ReGulated-Cascade (RGC) transimpedance amplifier achieves an experimentally measured -3dB bandwidth of more than 6 GHz and a transimpedance gain of 51 dBΩ, which is the fastest reported TIA in CMOS $0.35\mu m$ technology. The 5 GHz Gilbert cell mixer produces a conversion gain of 11 dB, which greatly minimized the noise contribution from the IF stage. The noise figure and input IIP_3 of the mixer were measured to be 15.7 dB and 1.5 dBm, respectively. The IF amplifier and output buffer pick up and further amplify the signal for post processing. The optical front end demonstrates a typical equivalent input noise current of 35 $pA/\sqrt{Hz}$ at 5 GHz, and a total transimpedance gain of 83 dBΩ while consuming a total current of 40 mA from 3.3 V power supply. The -3 dB bandwidth for the optical front end was measured to be 4.9 GHz. All the prototype chips, including the optical front end, and the individual circuits including the photodiode, TIA, mixer were probe-tested and all the measurements were taken with Anritsu VNA 37397D and Anritsu spectrum analyser MS2721A.

Acknowledgments

I would like to thank Ian Harrison and Barrie Hayes Gill, my supervisors, for their guidance throughout my research work. I am very grateful to Barrie who, over the time of the research work, has constantly provided ideas, suggestions and much inspiration. For Ian, it is your brilliant insights and support that make this work possible, I owe you so much not only for your support on the research, but also for your support in my life, without reservation, like a friend.

I would like to express my thanks to the research staff in applied optics group in Electrical and Electronic Engineering Department, University of Nottingham for their support and cooperation on the work. Thanks also go to the many colleagues and technicians who have in one way or another been involved in my work.

Amongst my friends, I would like to thank Frank, Vinoth, Kudip and Fen for the happy time in the office. Especially I would like to thank Sheng, my deep friend in UK, for all you did for me I will always be grateful. For Ting, thank you so much for the happy time and your patience on the word processing.

Finally, my sincere thanks go to my family, and my deepest gratitude and love to my parents, my brother and my dear sister for their belief in me.

Acronyms and Abbreviations

ADS	Advanced Design System
AMS C35	Austria Micro Systems CMOS $0.35\mu m$ process
APD	Avalanche Photodiode
APS	Active Pixel Sensor
BER	Bit Error Rate
CCD	Charge Coupled Device
CB	Common Base
CG	Common Gate
CGH	Computer Generated Hologram
CMOS	Complementary Metal Oxide Semiconductor
CS	Common Source
D2S	Differential to Single
DCR	Direct Conversion Receiver
DFB	Distributed Feedback
DSB	Double Side Band
DUT	Device Under Test
EHP	Electron Hole Pair
ENR	Excessive Noise Ratio
FET	Field Effect Transistor
ICP	Input 1-dB Compression Point
IF	Intermediate Frequency
I/O	Input Output
G-S-G	Ground Signal Ground
LO	Local Oscillator

LRM	Line Reflection Matched
LA	Limiting Amplifier
MSM	Metal Semiconductor Metal
NA	Numerical Angle
NDT	Non Destructive Test
NF	Noise Factor (Noise Figure)
O-SAM	Optical scanning Acoustic Microscopy
PCB	Printed Circuit Board
PD	Photodiode
PIN	P+ Insulator N+
RBW	Resolution Bandwidth
RGC	ReGulated Cascode
RF	Radio Frequency
SAW	Surface Acoustic Wave
SiGe	Silicon Germanium
SML	Spatial Modulated Light
SOI	Silicon On Insulator
SOLT	Short Open Load Through
SSB	Single Side Band
TIA	Transimpedance Amplifier
VBW	Video Bandwidth
VCSEL	Vertical Cavity Surface Emitting Laser
VLSI	Very Large Scale Integrated
WG	Wave Guide

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The work presented in this thesis is concerned with the realization of a high speed optical front end in CMOS process which senses and demodulates the light signal in an optical sensing system. The system finds applications in solid material non-destructive testing, i. e., flaw detection and material characterization by acquiring images of the interaction between surface acoustic waves (SAWs) and a solid material. It uses lasers to generate and detect the SAWs, and has been developed as an accurate non-contacting laser based system for high resolution surface waves measuring and imaging. The generated surface acoustic waves will be reflected and changed when they propagates along the surface of the solid material, which can be imaged by an optical sensor. Basically, the surface acoustic waves contain a fundamental frequency which is determined by the laser generation system, but they are also rich in high frequency harmonics, which can be up to tens GHz. For the purpose of improving the spatial resolution of the imaging and thus the better accuracy of the non-destructive testing, a high speed back end optical sensor plays an important role in the system for flaw detection or material characterization.

Much of this work is focused on the design, analysis and implementation of the devices and circuits for the high speed optical front end in a low cost CMOS $0.35\mu m$ process. Prior to the main part of the thesis, chapter 1 gives a brief description about the applications of non-destructive testing which addresses the principles of photoacoustic waves and laser generation and detection systems.

1.1 Photoacoustic History

The photoacoustic effect was discovered by Alexander Graham Bell in 1880 [1, 2], when he observed how focusing a ″rapidly-interrupted″ beam of sunlight onto a thin diaphragm of material resulted in a ″clear, musical tone[2].″ Postulating the ″sonorousness″ under these conditions was a fundamental property of matter, Bell devised a method for studying the effect in solids, liquids and gases. He tested a variety of substances including potash, selenium, tobacco smoke, cork, platinum, ink, wool, and even a cigar. Bell, who had just made his famous telephone call to Watson in the spring of 1876, hoped to use the photoacoustic effect to create a ″photophone″, a device like the telephone, but with the signal transmitted as light rather than electricity. Perhaps Bell anticipated the advantages of optical networks. These discoveries were then further investigated and published in 1881 [3].

Similar experiments on gases were performed by John Tyndall and Wilhelm Rntgen in 1881 after hearing about Bell's discovery. Then the field lay dormant for 50 years until the discovery of the microphone made it possible to enhance the measurements. In 1938 Viengrov at the State Optical Institute of Leningrad used the method to study infrared absorption in gases and the gas content in gas mixtures. Pfund developed a gas analyser in 1939 which was used in John Hopkins Hospital in Baltimore to measure CO and CO_2. Luft developed a commercial gas analyser which became available in 1946. The interest in photoacoustic grew in this period and it was only used to monitor gases. When the infrared spectrometer was invented, a more accurate method for monitoring gases existed so the field again lay dormant until 1970[4, 5].

It was the invention of the laser that gave the field of photo acoustics new possibilities. Consequently, the photoacoustic effect remained largely a scientific curiosity until Kreuzer combined photoacoustic detection with lasers to demonstrate extremely sensitive gas detection in 1971 [3]. Kreuzer then went on to demonstrate photoacoustic detection for pollution monitoring [4, 5], and for gas chromatograph detection [6]. Dewey et al. [7] improved the sensitivity of photoacoustic detection systems further by modulating the optical excitation at a frequency corresponding to an acoustic mode in the cavity. In the late seventies, the field of photoacoustic exploded, finding commercial applicability today in trace gas analysis, pollution monitoring, biological studies and material characterization [8], which section 1.3 goes to detail.

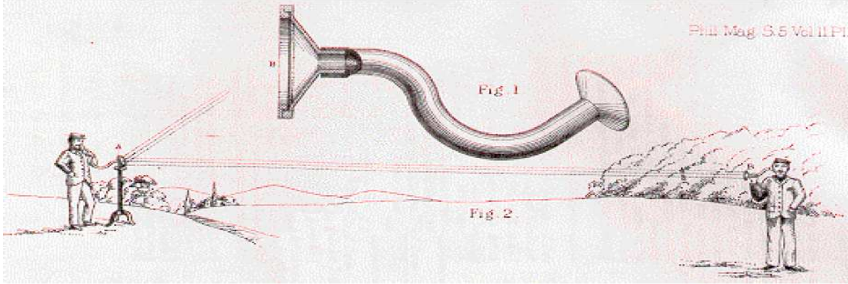


Figure 1.1: Bell's plan for photophone[2]

1.2 Physics Behind Photoacoustic

Much theory has been developed to explain the phenomenon of photoacoustics [8]. This section presents the fundamental elements in understanding the physics behind photoacoustics. Generally high energy light is used, usually a laser, firing into the test sample. When the laser hits the sample, some of the energy is absorbed by the molecules in the media resulting in a region of higher temperature. The rise in temperature will generate an expanding region and a pressure wave will propagate away from the source. This decaying pressure wave can be then picked up by a transducer or a laser imaging system. The phenomenon is due to the fact that the molecules in the sample being monitored have a quicker response to light than to the surrounding media. The process is shown in Fig. 1.2.

The thermal elastic expansion mechanism is an interesting choice for material characterization and medical diagnosis for a variety of reasons. Firstly, it does not break or change the properties of the object under study. Secondly, it has a linear or a definite relationship with many of the physical parameters of diverse materials. Thirdly, it is non-destructive or non-invasive in applications such as materials test and medical diagnosis.

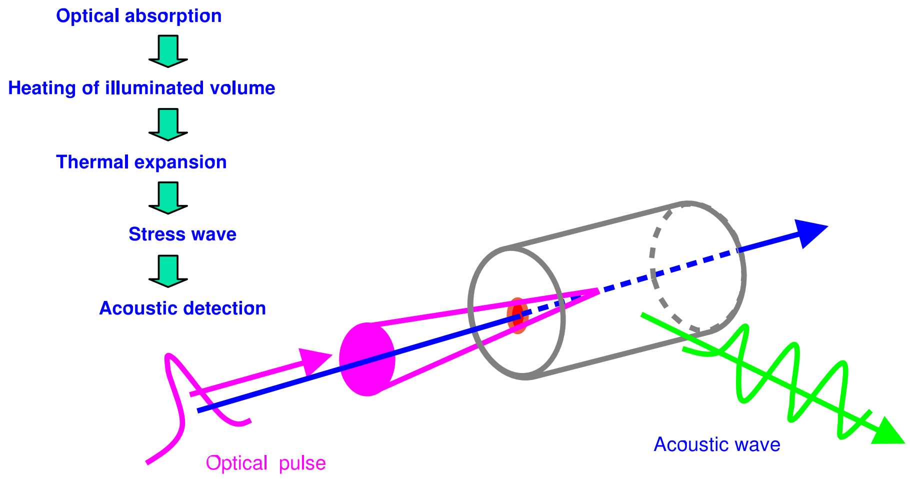


Figure 1.2: Photoacoustic generation

1.3 Non-Destructive Test (NDT)

NDT involves testing an object for its integrity and fitness without damaging the object. There are many different general methods of testing objects nondestructively. These methods include optical inspection, radiographic testing, magnetic particle inspection, holographic testing, acoustic emission and, of course, ultrasonics testing [9]. The role of ultrasonics in NDT can be split broadly into two areas; that of flaw detection, and material characterization. There is also a slight crossover in the area of flaw characterization. The interaction of ultrasonic waves with the material and features therein can be used to ascertain the presence of flaws (and perhaps their nature) or some physical or chemical characteristic of the material under test.

1.3.1 Flaw Detection

A common requirement for a NDT process is to determine whether any flaws are present that will adversely affect the operation or lifespan of a component. Typical flaws include surface-breaking or interior cracks, weak joints, coating delamination or loss of adhesion between coating and substrate.

There are several methods of detecting flaws with acoustic waves, and obviously different methods are appropriate for different types of flaws, the use of a particular technique depends on many factors, including sample geometry, attenuation, and the characteristics of any possible flaws. In a named pulse-echo techniques, the existence of a flaw is determined by the reflection of ultrasound, which is detected by the combined transmitter/receiver, the location of the flaw from the transmitter can be calculated by the ”time of flight” of the received ultrasound if the velocity of the acoustic waves in the material under investigation is known.

In general, the higher the acoustic waves frequency, the higher the resolution, both spatial and temporal, but with the disadvantage that losses caused by scattering and aberration are higher [9].

1.3.2 Material Characterization

The characterization of materials is another important area of NDT. Characterization can include ascertaining various dimensions (for example the thickness of the steel), coating thickness, porosity, residual stress, grain size in the case of multi-grained materials such as metals, or grain orientation in single-grained materials such as silicon. The characterization may be required mid-process, in which case the ultrasonic inspection may provide a means of process control, or post-process, in which case it provides a means of quality management [9].

In terms of geometry characterization, the pulse-echo method, is the most common form of ultrasonic inspection, whereby the time taken for an ultrasonic pulse to be reflected off the rear surface of the material is related to the thickness of the material. Knowledge of the wave velocity is necessary, but for many materials this has been well known for many years.

The velocity of ultrasonic waves is the primary means of many forms of characterization, since it is a function of the moduli of elasticity and density. Changes in velocity can therefore correlate with changes in the material properties such as porosity, residual stress or, in the case of surface waves, coating thickness. If the change in property is the result of a process that the material is being subjected to, then examination of the velocity may provide a means of monitoring the process concerned.

1.4 O-SAM in University of Nottingham

An all-Optical Scanning Acoustic Microscope System (O-SAM), which involves a totally non-contact method of acquiring images of the interaction between surface acoustic waves (SAWs) and a solid material has been set up by applied optics group in the University of Nottingham [12]. The system uses lasers to generate and detect the SAWs, and has been developed from an accurate non-contacting laser based system for surface wave velocity measurement [13].

Fig. 1.3 shows the key parts of the O-SAM system. Two lasers are involved, one for ultrasound generation, the other for detection of the ultrasound. Both are associated with some optics and mechanical stages (for adjustment and scanning) and some electronics and data acquisition equipment.

The energy from the pulsed generation laser is applied to the surface of the sample via optics that both focus the light onto the sample and control its spatial distribution (CGH means Computer Generated Hologram). This control of the generation profile is necessary to obtain the large surface acoustic wave amplitudes required. The sample is mounted on automated mechanical stages to enable the formation of scanned images. The excited surface acoustic waves propagate on the sample surface, and are then detected by focusing a continuous wave laser onto the sample surface and measuring the angular deflection. Analogue amplitude and phase detection electronics then amplify and convert the signals from the detectors into a form suitable for acquisition by a host computer.

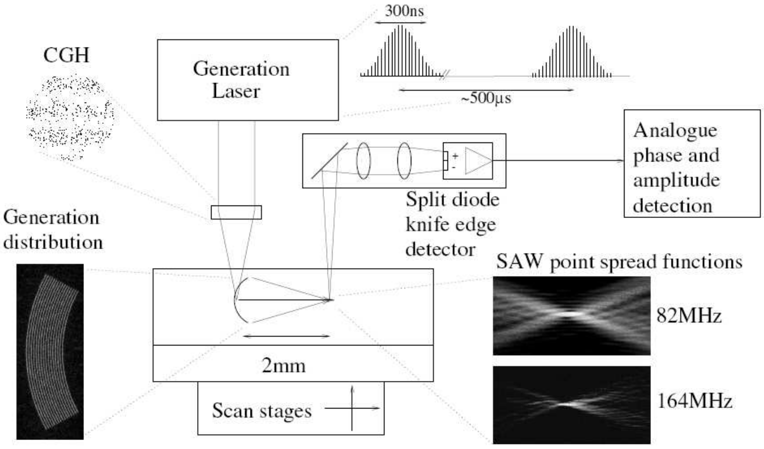


Figure 1.3: Diagram of all-Optical Scanning Acoustic Microscopy (O-SAM) [12]

The acoustic frequencies are determined by the generation laser. This is a Q-switched, mode locked laser, producing a tone-burst of approximately 30 very short pulses (200 *ps*) separated by 12.1 *ns* with a variable repetition rate (up to 5 KHz). The average power output of the laser is around 2 *W* (depending on the repetition rate); at low repetition rates (below 2 KHz) the peak power reaches a maximum, whereas above this it decreases with increasing repetition frequency. The signal has a fundamental frequency of 82 MHz (which is determined by the laser generation source) and contains harmonics of this. Higher frequencies have been observed up to 10 GHz [12]. This thesis is working on the implementation of high speed optical front end to allow greater spatial resolution of the acoustic wave imaging system.

Chapter 2

Photodetector

2.1 Introduction

Photodetectors are the components in an optical receiver system that convert optical energy into electrical energy. They are involved in applications of imaging, communications, and optical sensing. CCD (Charge Coupled Device) and CMOS sensors represent the widely used photodetectors for imaging system, and the competition between the two technologies continues today. Low noise, low power, low cost, and high sensitivity are always of great concern for these systems.

Different from imaging systems, the speed of the photodetectors in optical communication and sensing systems is generally the first priority instead. High speed, high sensitivity photodetectors are generally made from III-V compound to achieve the highest possible performance for long haul optical communications. However, the high cost of these materials limits their use in optical sensing systems, as in this thesis. Short wavelength light sources (shorter than 850nm) and a high speed, inexpensive photodetector which can be integrated in low cost photo receivers is the design target for this work.

2.2 Principles of Photodetection

A photodetector converts light energy into electrical energy, manifested usually in the form of photocurrent[23]. Semiconductor photodetectors rely on the absorption of incident photons with energy greater than the semiconductor bandgap energy to generate electron-hole pairs (EHPs). Photodetection broadly involves three processes [24]:

- Absorption of optical energy and generation of carriers;
- Transportation of photogenerated carriers away from the absorption region;
- Carrier collection and generation of photocurrent.

The performance of a photodetector can be characterized by various figures of merit. These include the responsivity of the detector, its bandwidth, and the noise added to the signal by the detector.

Responsivity Responsivity is a measure of light-to-current conversion efficiency of the detector. A high detector responsivity improves the signal-to-noise ratio of the receiver system. It is possible to have gain in photodetectors (as in avalanche photodetectors) due to impact ionization and avalanche multiplication that can lead to very high responsivities. But these mechanisms are usually accompanied by a penalty in bandwidth and noise performance. Mathematically, the responsivity of a detector, R, is defined as [36]

$$R = \frac{I_{ph}}{P_{inc}}$$

where I_{ph} is the photocurrent, P_{inc} is the incident optical power. η is sometimes defined as the external quantum efficiency representing the fraction of incident photons leading to I_{ph}. In reverse biased junction photodiodes where the depletion region (high field region) constitutes the bulk of the absorption region, it can be approximated by

$$\eta = \eta_i(1-R)(1-\frac{e^{-\alpha d}}{1+\alpha L_n}) \tag{2.1}$$

where R is the optical reflectivity between air and the semiconductor, α is the absorption coefficient of the intrinsic region, d is the depletion region thickness and η_i is the internal quantum efficiency defined as the ratio of number of EHPs created to the number of absorbed photons. In pure material, η_i is almost unity. L_n is the minority carrier diffusion length, *i.e.* for electrons in the p-type substrate. In the absence of diffusive transport, the term αL_n is missing in (2.1), and $(1-e^{-\alpha d})$ represents the fraction of light absorbed in the depletion width of the detecting junction. Note $L_n = \sqrt{D_n \tau_n}$, where D_n is the diffusion constant and τ_n is the minority carrier lifetime [46]. Fig. 2.1 shows the optical absorption coefficient for general semiconductor materials.

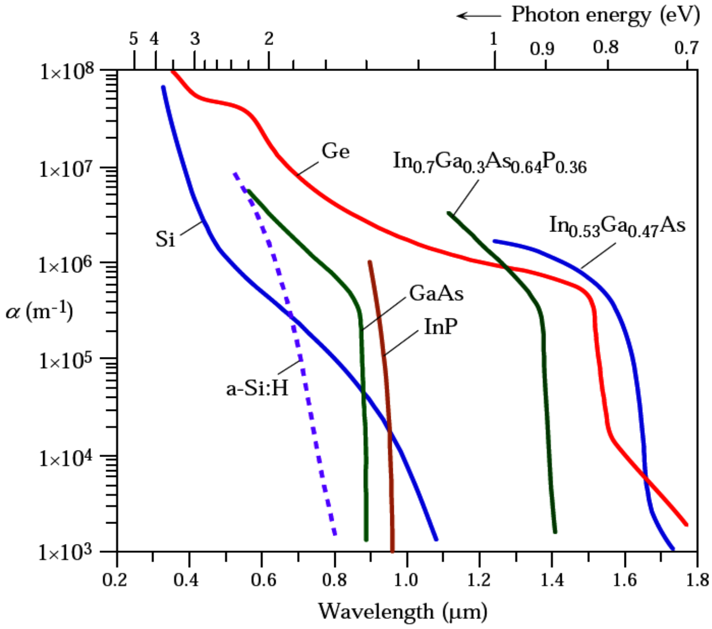


Figure 2.1: Optical absorption coefficients for various materials[45, 46]

Bandwidth The bandwidth of a photodetector is defined as the frequency at which the responsivity of the detector has fallen by 3-dB from its low frequency value. It is limited mainly by carrier transit time, RC time constant, diffusion time by photogenerated carriers outside the depletion region [38]. Carrier transit time is the time taken by photogenerated carriers to travel across the high-field region. It is usually dominated by hole transit time, as holes typically have a lower drift velocity than electrons in common photodetector materials. The RC time constant is determined by the equivalent circuit parameters of the photodiode and the load circuit. Diode series resistance (due to ohmic contacts and bulk resistances), load impedance, and the junction and parasitic capacitances contribute to the RC time constant. Diffusion time becomes important when the photocurrent due to carriers absorbed in the p and n contact regions within about one diffusion length of the edge of the depletion region becomes comparable to the current arising from photogenerated carriers within the depletion region. In Fig. 2.2 (a), the depletion and diffusion regions of a typical p-n junction photodiode are illustrated where d stands for depletion region width, L_n, L_p for diffusion length of electrons and holes in the semiconductor (practically, L_n, L_p are much longer than d, which is not as shown as in Fig. 2.2). The transit time then can be written as d/υ (υ is the saturation velocity of holes); diffusion time is determined by the time that the electrons take to get through the lower diffusion region in Fig. 2.2 (a) or the holes via the upper one. Because the diffusion process is driven by the random motion of carriers, it is a slow transport mechanism. This was interpreted and experimentally confirmed in the work of [39, 40]. In this case, the diffusion component contributes a slow but long tail to the detector's impulse response which significantly limits its high frequency response. Detailed analysis is given in following sections.

A simple equivalent circuit of a p-n photodiode is given in Fig. 2.2 (b). If the parasitic capacitance C_p and series resistance R_s are negligible, then the RC-limited bandwidth can be given by

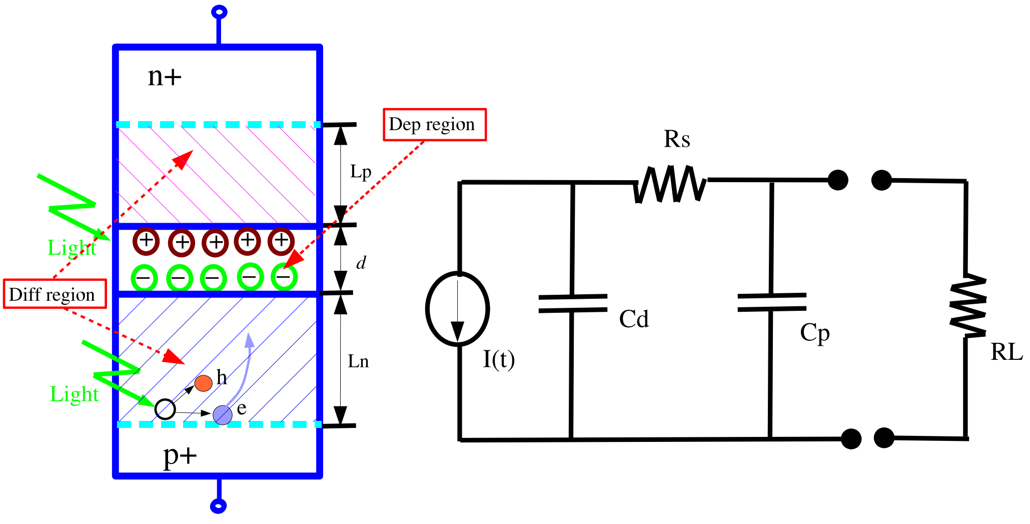
$$f_{RC} = \frac{1}{2\pi R_L C_d} \tag{2.2}$$

$$C_d \propto \frac{A}{X_d} \tag{2.3}$$

$$X_d = \sqrt{\frac{2\varepsilon_0\varepsilon_i}{q}(\frac{1}{N_n}+\frac{1}{N_p})\cdot(\triangle V+\phi_i)} \tag{2.4}$$

$$\phi_i = \frac{kT}{q} lg\left(\frac{N_n \cdot N_p}{N_i^2}\right) \tag{2.5}$$

where R_L is the load resistance, C_d is the photodiode junction capacitance, A is the area of the diode, X_d is the depletion width of the junction, N_p and N_n are impurity concentration of *p, n* region, $\triangle V$ is the reverse biased voltage applied to the photodiode, and ϕ_i is the junction potential across the *p-n* junction. ε_0 is the dielectric constant in vacuum, ε_i is the relative dielectric constant of the semiconductor, k is the Boltzmann constant, T is the absolute temperature in K, q is the electronic charge, $N_i \approx 1.3 \times 10^{10} cm^{-3}$ is the intrinsic carrier density of silicon at T. It can be seen from equations (2.3), (2.4), shrinking the size of the diode, increasing the reverse bias voltage and lowering doping level lead to smaller C_d. But the smaller diode area generally means smaller window to the light incident and degrades the responsivity; reverse bias voltage is limited by breaking down voltage; changing doping level is not always available for integrated photodetectors such as in CMOS. Hence there are lot of trade offs in bandwidth optimization of the photodiodes.



(a) (b)

Figure 2.2: (a) PN junction as a photodiode (b) equivalent circuit for a p-n junction photodiode

For discrete photodetectors, R_L is usually assumed to be 50 Ω because most of the RF test instruments used for device characterization are designed to have an input impedance of 50 Ω. In photo receivers, the input impedance of the amplifier following the photodetector is often designed to be lower than 50 Ω to improve the circuit bandwidth.

Noise The photodetector not only produces the signal current I_s, but also noise. Shot noise and thermal noise are two dominant sources of noise in high-speed photodetectors. Shot noise current is the result of the photocurrent being composed of a large number of short pulses that are distributed randomly in time. Each pulse corresponds to an electron-hole pair created by a photon. With white noise approximation, its mean-square value turns out to be [41]

$$\overline{i_n^2} = 2qI \cdot BW_n \tag{2.6}$$

where q is the electron charge, I is the signal current and BW_n is the bandwidth in which we measure the noise current[1]. The photodetector produces a small amount of current even when it is in total darkness. This so called *dark current*, depends on the junction area, temperature, and processing. It manifests itself as shot noise, contributes to the total system noise and gives random fluctuations about the average particle flow of the photocurrent [41]. Since most photodiodes are operated under reverse bias conditions, the dark current is usually quite small when compared to the that of the following transimpedance amplifier in operating condition[2].

Thermal noise in photodetector is caused by the various resistances in diode's equivalent circuit. Diode shunt resistance and series resistance both contribute to the thermal noise. If the diode is followed by an amplifier in a receiver circuit, then the input resistance of the amplifier and the amplifier's noise also contribute to the thermal noise of the overall photo receiver.

Apart from these noise sources, there is also a contribution from the background radiation in the ambient where the detector is placed. Usually, the contribution of the amplifier dominates noise performance in a photo receiver at telecommunication wavelengths. Detailed noise analysis of the optic front end is presented in chapter 3, the transimpedance amplifier design.

2.3 Overview of High Speed Photodetectors

There are many different types of high speed photodetectors possible with the most appropriate kind of detector being determined by the different application. With architecture concerned, *p-i-n*, Metal-Semiconductor-Metal (MSM), and Avalanche Photodiodes [44, 23] are generally referred to and are discussed below.

---

[1] It is noted that avalanche photodiode has a much bigger noise current than that calculated from (2.6) because of the excess noise factor introduced by randomness of avalanche multiplication process [41].

[2] This would be another case in the application of CMOS photodetector as APS (Active Pixel Sensor) at low frequency range.

2.3.1 *p-i-n* Photodiode

A *p-i-n* photodiode consists of a *p-n* junction with a layer of intrinsic or lightly doped semiconductor sandwiched between the *p* and *n* layers. The intrinsic region has a small number of carriers and is easily depleted of any charge. Therefore, the depletion region is almost entirely contained in the intrinsic region. Fig. 2.3 shows a *p-i-n* photodiode together with its electric field [25]. Light absorbed in the semiconductor produces electron-hole pairs, however, due to the substantial electric field in the depletion (intrinsic) region, pairs produced in the depletion region or within a diffusion length of it, will be separated by the electric field leading to current flow in the external circuit.

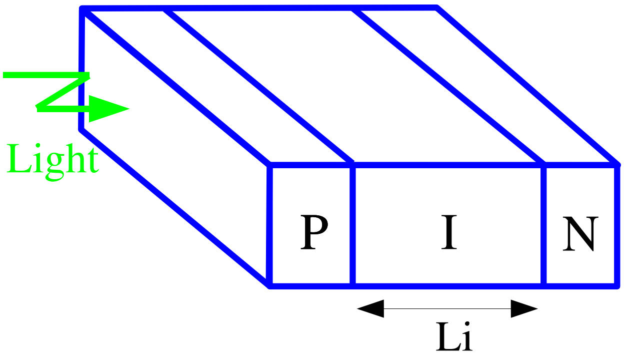


Figure 2.3: *p-i-n* photodiode

Very high speed and high sensitivity compound semiconductor p-i-n photodiodes have been reported in literature. For example, a waveguide integrated *p-i-n* photodiode (WG PIN PD) with a 3 dB bandwidth of > 40 GHz at 1550 *nm* was demonstrated by Wang *et al* [26]. In [27], 1550 *nm* InP/GaInAs/InP *p-i-n* photodiodes have been fabricated with speeds of up to 60 GHz. However, despite the high bandwidth and high sensitivity clearly achievable with compound semiconductors, there is an eagerness to produce silicon long wavelength photodetectors. This is, again, due to the low cost, large scale integration achievable with silicon. A possible means of extending the spectral response of silicon into the long wavelength region has been to use heterostructures composed of silicon (Si) and germanium (Ge). Many successful reports of *p-i-n* SiGe/Si superlattice photodetectors have been made for wavelengths of 850-1300 *nm*. For instance, Temkin *et al* developed a 1300 *nm* GeSi/Si waveguide *p-i-n* photodiode that operated at speed > 1 GHz and had an internal quantum efficiency of 40% [29], while Tashiro et al demonstrated a 10.5 GHz, 980 *nm* planar *p-i-n* SiGe/Si diode with an external quantum efficiency of 25-29% [30].

2.3.2 MSM Photodiode

a MSM PD is comprised of back-to-back Schottky diodes that use an interdigitated electrode configuration on an undopped semiconductor layer, as shown in Fig. 2.4. When light with energy $hv > E_g$ is incident, the light that hits the semiconductor surface is absorbed and creates electron-hole pairs (EHs) within the active region, and then one set of electrodes acts as a cathode and the other as an anode. The holes drift toward the negative electrodes, and electrons travel to the positive electrodes under the influence of an electric field by an applied reverse bias voltage [32, 33].

The metal electrode fingers have finger width w and are separated by a distance s. High photodiode quantum efficiency requires $s \gg w$, or low electrode shadowing as it is termed. Absorbed photons generate electron-hole pairs in the semiconductor, the holes drift with the applied electric field to the negative contacts while the electrons drift to the positive contacts forming a current (i_{photo}). The bandwidth of a MSM diode is similar to that of a *p-i-n* detector in that it is both RC time constant and transit time limited.

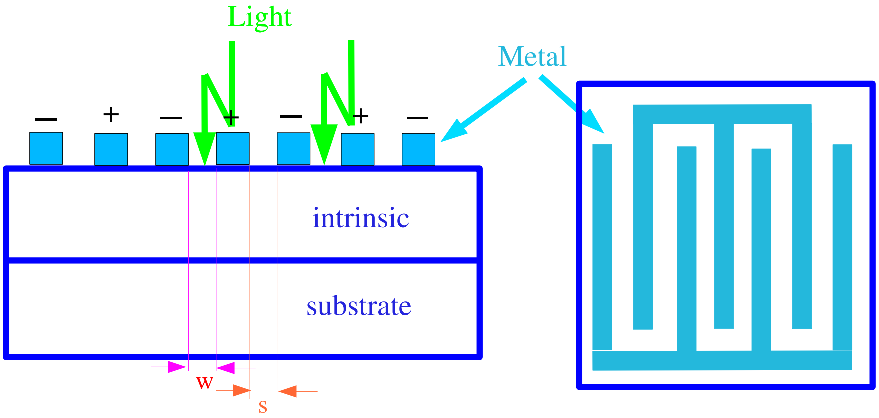


Figure 2.4: MSM photodiode cross section and top view

MSM photodiodes have simple, planar structures and can easily be fabricated with FET (field effect transistor) processes provided that the substrate is highly resistive or intrinsic epitaxy layer is available. They have also been shown to operate to very high speeds with bandwidths of 510 GHz on low temperature (LT) GaAs and 110 GHz on bulk silicon reported [34, 35]. The silicon photodiode, however, achieved this speed at very short wavelengths (400 *nm*) and at longer infrared wavelengths (800 *nm*) this speed became diffusion limited.

Although superior for its least complexity and high speed performance, MSM photodiodes are always associated with problems of quite low responsivity because of the reflection from the surface metals and semiconductor surface; the finite carrier lifetime as the carriers traverse the gap between the electrodes before being collected; absorption of incident light outside the region in which photogenerated carriers can be collected by the electrodes; and surface recombination currents and deep traps within the semiconductor material which may lower the detected optical signal. Furthermore, similar to *p-i-n* photodiode, it is not compatible with CMOS process.

2.3.3 Avalanche Photodiode

Avalanche photodiodes differ from *p-i-n* and MSM diodes in that they incorporate a high field region that multiplies the photocurrent through the avalanche generation of additional electron-hole pairs. The operation of PIN diodes is based on the generation of one electron-hole pair for each photon entering the lattice. avalanche photodiodes (APDs), on the other hand, the generated electrons and holes carry so much energy that they themselves can stimulate other electrons and holes, creating an avalanche effect.

APDs operate with controlled avalanche, that is, with a multiplication factor, M, of several hundred. Thus, each photon entering the device may create hundreds of of electron-hole pairs, providing a large output current. Shown in Fig. 2.5 is the structure of a typical APD, consisting of a sandwich of $n+$, p, and i layers atop a $p+$ substrate. As in PIN diodes, the intrinsic region enables generation of electron-hole pairs. Grown as a very uniform and thin layer, the p region supports a high electric field to create avalanche.

Besides the internal gain and high responsivity, APDs suffer from ”gain-bandwidth” trade off. That is, the higher the multiplication factor is, the longer the avalanche persists, limiting the changes of the current and the response to high frequency signal. Secondly, the reverse bias voltage applied to APDs must be controlled precisely so as to achieve avalanche while avoiding break down. Furthermore, it is incorporated with significant noise because of the noisy avalanche process [25].

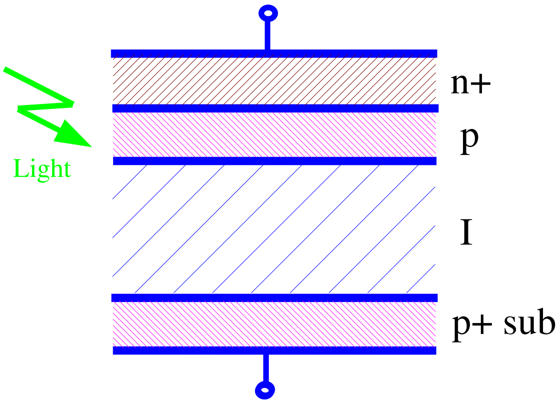


Figure 2.5: Avalanche Photodiode

2.4 Overview of CMOS Photodetector

Since the invention of the integrated circuit, the electronics industry has been relentless in its quest to integrate the functions of the electronic systems onto single silicon chip. This process has generally resulted in both increased utility and decreased system costs, an outcome that tends to accelerate the trend [42]. CMOS technology has achieved the greatest success in this regard [43]. Aggressive scaling resulting from competition to follow Moore's law has improved the integration density of CMOS transistors by more than 3 orders of magnitude in the last 30 years, moreover, advanced CMOS devices can now achieve cutoff frequencies comparable to that of III-V compound devices.

A CMOS monolithically integrated photo receiver is therefore desirable for use in short distance connection systems and optical sensing systems based on its low cost, low power consumption, high reliability and potential for large scale integration. Here monolithic integration is necessary because integrated photodiodes can be beneficial since the capacitance at the input node of the preamplifier circuit is greatly reduced due to the elimination of bonding wire parasitics.

It is not yet possible, however, to realize a high speed, high sensitivity photodetector in a production CMOS process without adding complexity to the process. Presently, no commercial silicon photo receiver uses an integral, monolithic detector, instead an external detector is hybridized to the receiver resulting in increased cost, and decreased bandwidth due to interconnection parasitics [44].

To address the challenge for implementing integrated photodetectors in silicon, first we rewrite the equation (2.1) here:

$$\eta = \eta_i(1-R)(1-\frac{e^{-\alpha d}}{1+\alpha L_n}) \tag{2.7}$$

Two characteristic lengths associated with the problem are L_n and α^{-1}. For d on the order of α^{-1}, most of the carriers are collected rapidly by the depletion field, and diffusive transport can be made small. At moderate voltages, transport at saturation velocity is typical, and the cutoff frequency associated with such transport is proportional to υ_{sat}/d, which easily scales to very high speeds. On the other hand, in reasonably undamaged silicon the minority carrier lifetime can be in the milliseconds range, and diffusion lengths of $100nm$ or more are common. The cutoff frequency will scale as $\tau_n^{-1} = D_n/L_n^2$, and may be even in the kilohertz range. In a detector with a thin depletion region, diffusive transport will be a major fraction of the total current, making the device unsuitable for high-speed application.

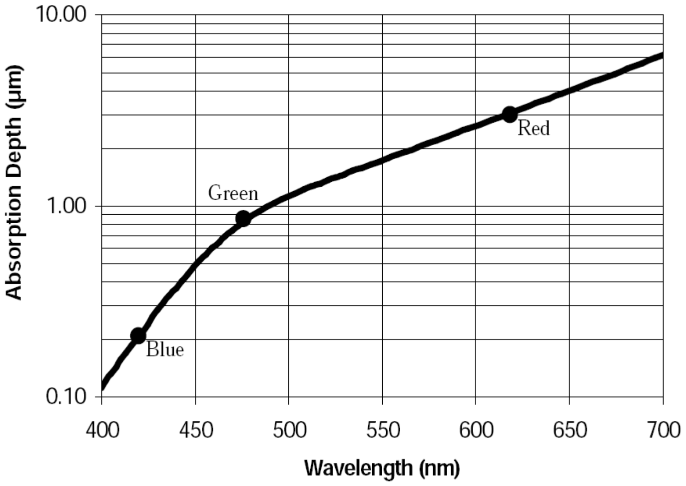


Figure 2.6: Silicon absorption length[44]

Fig. 2.6 shows the silicon absorption length against the light wavelength. With wavelength of $800nm$, the silicon absorption length is about $15\mu m$, while the depth of N-Well in a typical $0.35\mu m$ CMOS is only $1.8\mu m$. From Fig. 2.1, it is clear that as an indirect gap material, silicon has a smaller optical absorption coefficient (thus longer absorption length) than the III-V compounds at the light wavelength above $800nm$. This has a great impact on silicon photodetector's high frequency performance since lights with long wavelength penetrate deeper into silicon, thus causing unwanted diffusion current generated in the substrate. Consequently, it takes a long time for these carriers to reach the contact junctions. When light pulse impinges on the photodetector, the junction is intended to collect the early arriving *drift* carriers, as well as the delayed *diffusion* carriers. Sometimes, the delayed carriers are associated with previously received light pulses. This produces a slow frequency gain effect that limits the cutoff frequency, while in time domain the impulse response has a long tail effect as shown in Fig. 2.7. Slow response times due to deferred carriers in the substrate are the major obstacle to high-speed fully integrated silicon-based receivers.

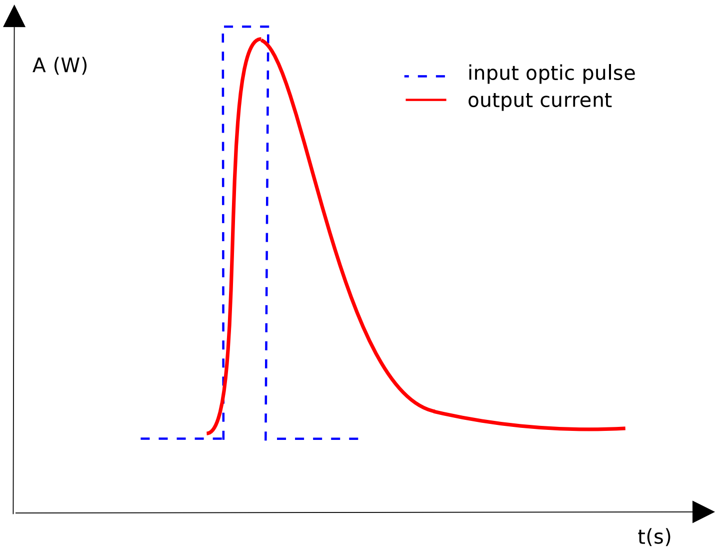


Figure 2.7: Diffusion long tail

However, despite the difficulties mentioned, continuously emerging techniques and solutions have been proposed [42, 52, 65, 61, 58, 67, 49, 50, 51], for integrating high speed photodetector in CMOS:

- 1. *Screen the slow bulk carriers by buried layer or oxide layer*

Workers at IBM first proposed this technique in detail for a fully integrated receiver in an unmodified BiCMOS process [47, 48]. This report explicitly describes the role of the buried collector contact as a screening terminal to remove the slow photogenerated bulk carriers, as shown in Fig. 2.8 (a), (b). The amplifier was entirely CMOS-based. The active detector was formed between the n-well of the BiCMOS process and p regions in the n-well. The buried collector layer, which makes contact to the n-well, forms a parasitic junction to the substrate, to which most of the slowly diffusing carriers are collected and screened from the amplifier, so that a high speed of response is obtained. Although the shallow nature of the 0.5 μm BiCMOS employed results in a low responsivity, their 75 $\mu m \times 75\ \mu m$ detector achieved a 3-dB bandwidth of 700 MHz at 850 nm, Performance could be improved by moving to shorter wavelengths as noted by authors. The stated goal of the work was not the highest possible speeds, but the most affordable overall data-link costs, which gave rise to their use of a large photodiode area and short wavelength operation.

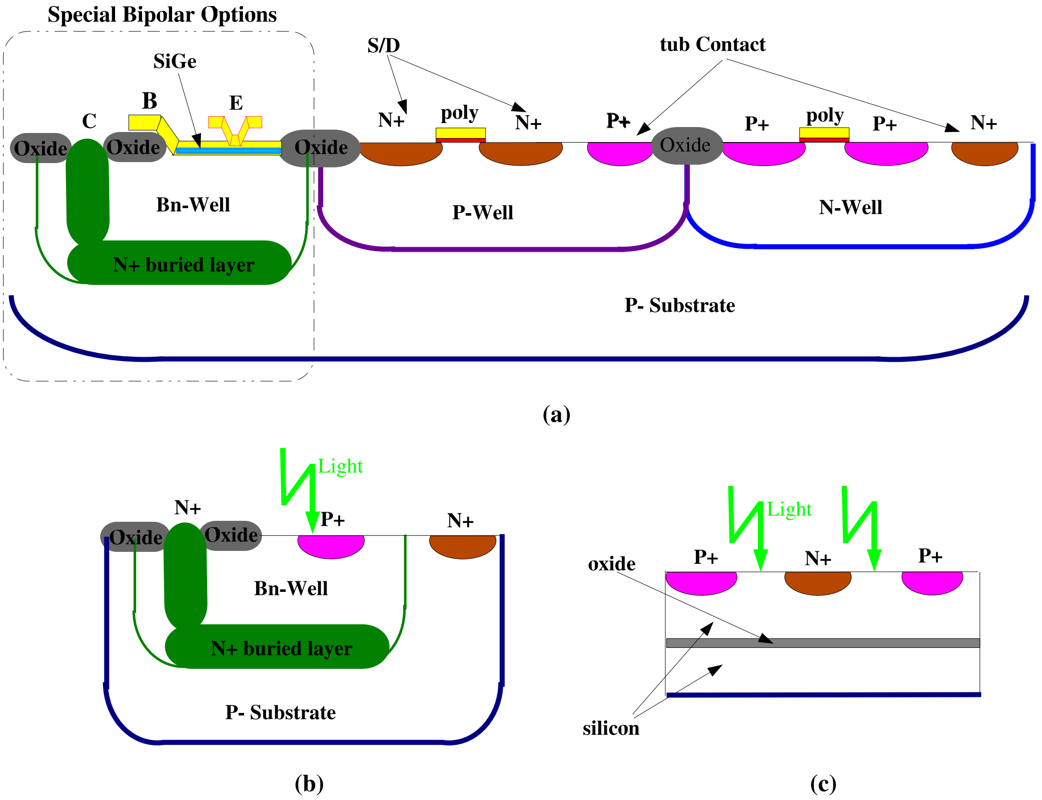


Figure 2.8: (a) simplified cross section of SiGe BiCMOS realization of NMOS, PMOS, and *n-p-n* device. (b) photodiode in BiCMOS, where the N+ buried layer is utilized as screen terminal which blocks slow bulk carriers to high field region. (c) photodiode in CMOS SOI, where the insulator layer (i.e. oxide) acts as the screen terminal.

S.M. Csutak, B. Yang, *et al.* presented several interesting papers [49, 50, 51] which reported the monolithically integrated photodiode in CMOS SOI technology.

In [51], interdigitated silicon photodiodes fabricated on standard Silicon-On-Insulator substrates with area of 144 μm^2 , 1 $\mu m \times 0.2\, \mu m$ (finger width * finger spacing), and a 2-μm-thick active layer achieved a bandwidth of 8 GHz at 9 V with a peak efficiency of 12% at 850 *nm*. The devices with the same geometry fabricated on a 200-*nm*-thick active layer exhibited a 15-GHz bandwidth and 3% quantum efficiency for a reverse bias of 9 V. the simplified cross section of the SOI CMOS technology is shown as Fig. 2.8 (c). Here the screen terminal for the photodetector is the oxide layer which has a similar in function as N+ buried layer in BiCMOS process. In [52] the photodiodes implemented in commercial 130 *nm* CMOS technology with 2 μm SOI substrate achieved -3 dB bandwidth of 8GHz with the optimized area of $50\, \mu m \times 50\, \mu m$, finger width and spacing of $2\, \mu m \times 2\, \mu m$, but under the reverse bias voltage as high as 28 V.

R.Swoboda, H. Zimmermann, *et al.* successfully integrated the photodiode together with Transimpedance Amplifier (TIA) and post processing circuits in BiCMOS for optical interconnections operating up to 5-Gb/s [53, 54, 55]. It used a p-type substrate, and the cathode was formed by a *buried n*+ layer on which a low-doped *n-* epitaxial layer with a thickness of about 15 μm was grown. The doping concentration of this epi-layer is in the order of $10^{13} cm^{-3}$. A thin *p*+ junction sitting on top of the epitaxial layer, formed the anode of the photodiode. This vertical PIN photodiode with a thick intrinsic-layer, with moderate modification on BiCMOS process, can achieve a quantum efficiency of 74% at a wavelength of 670 *nm* and 50% at 850 *nm*.

Although the above mentioned photodiodes achieved even 15GHz bandwidth, they were fabricated in BiCMOS, or in CMOS SOI technology, which are similar, but potentially incompatible with commercial CMOS. T. K. Woodward in [42] reviewed the monolithically integrated photodetectors, and reported his own work of a photodiode in 0.35 μm CMOS technology without any process modification. N-Well was used as the screen terminal in the same manner as the buried layer in [53, 54, 55], which was connected to a fixed potential while intrawell p-diffusion's defined the active contact of the detector. The photodiode had a size of $16.54\, \mu m \times 16.54\, \mu m$, and was employed in a photo receiver which worked at 1Gb/s, with a reverse bias of 9 V.

- 2. *SML (Spatial Modulated Light), eliminating the long current tail effect by differential approach*

The SML-detector is a fully-differential interdigitated structure that have junction regions that are alternatively exposed[57]. The exposed regions collect both the early and late carriers while the unexposed regions would collect only the diffusion carriers. In Fig. 2.9, the top view and the cross sectional view of the detector is shown. In the figure, the odd fingers are exposed to light while the even fingers are protected by a floating metal mask. In general, this approach is an attractive solution because it is compatible with standard CMOS process without any modification steps [58].

The problems associated with SML photodiodes are obvious. First, it has a low responsivity since only half of the optical light is detected; second, it requires current subtraction at the input to the Transimpedance amplifier, which complicates the circuits and increases the power consumption. However in case of a fully differential transimpedance amplifier following the SML photodiode, this would be beneficial because of its suppression of the noise.

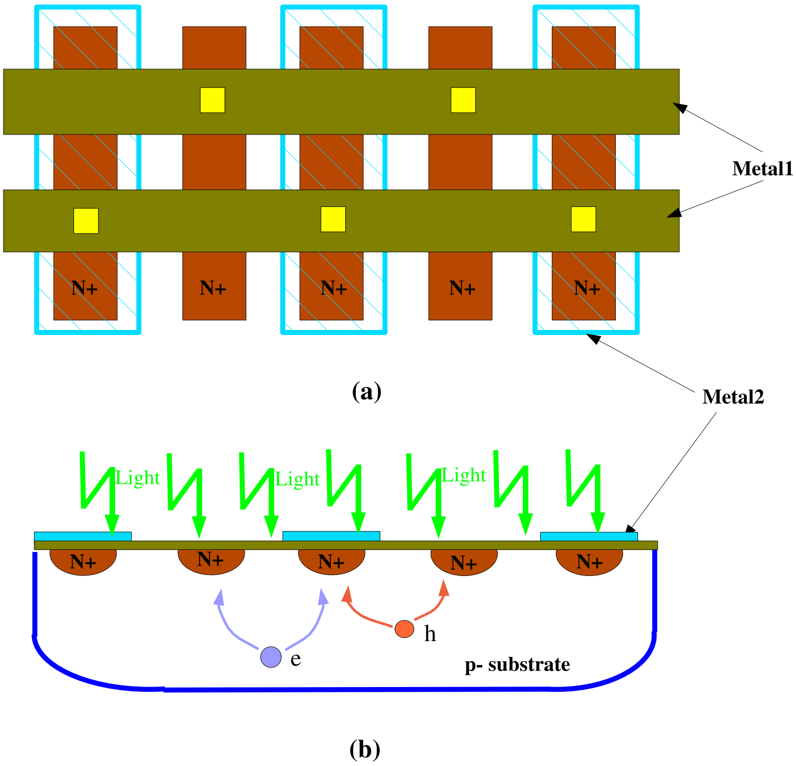


Figure 2.9: Modulated Spatial Light Photodiode in CMOS (a) top view of SML photodiode, the $n+$ active region are alternatively exposed by the floating shading of the Metal 2. Metal 1 collects the pnotogenerated currents from the shaded $n+$ regions and exposed ones. (b) cross section of SML photodiode. The p-substrate contact (anode of the photodiode) are not shown here, which is generally connected to a fixed potential. The difference of the two cathodes gives the output of the photodiode.

- 3. *Using polysilicon as photodetector*

In current MOS processes, polysilicon is used as a gate terminal for both NMOS and PMOS transistors. The lateral doping concentration of the polysilicon layer is high with the doping corresponding to the type of the MOS transistor. Using these two opposite types of poly layers, a p-n junction can be fabricated as a photodiode [59, 60, 61].

The carrier lifetime in polysilicon diode depends on recombination rates of holes and electrons and it is proportional to the concentration of recombination centers

[62]. It is also inversely proportional to the grain size of polysilicon. In deep sub micron CMOS technology, the grain size is on the order of several tens of *nm* [64], which causes the carrier lifetime to be very short [63]. Since in this case the diffusion speed of carriers is mainly determined by their lifetime, the diffusion bandwidth will be far in the GHz range.

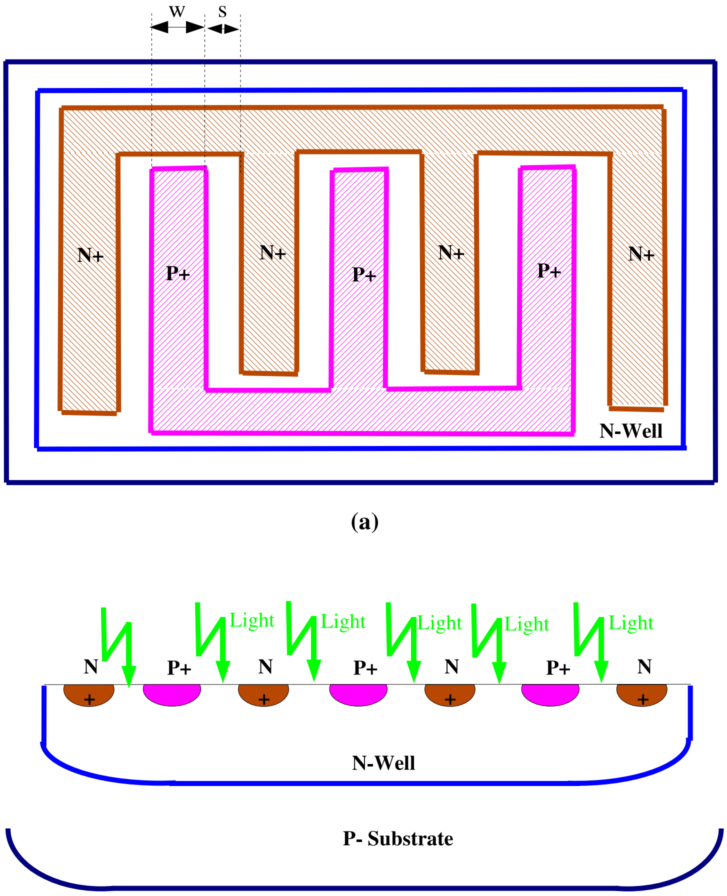
Although polysilicon photodiode is potentially faster than junction diode in CMOS process, it has a poor sensitivity to the light incident since a poly photodiode consist only of a small depletion region area plus the area outside this region proportional to a diffusion length of holes and electrons. The depth of the polysilicon in standard CMOS technology is about 0.2 μm and it also contributes to a poor responsivity of poly photodiode to vertical incident light.

2.5 Photodetector in AMS 0.35μm CMOS

2.5.1 Introduction and calculation

In order to achieve high-speed performance, such as that required for several GHz applications, whilst catering for the commercial AMS CMOS 0.35μm technology with no process modification, we have taken the approach mentioned previously of treating the n-well in standard CMOS process as a screening terminal to block the slow bulk carriers [42]. SML-detector [58] is an other potential candidate which also trades responsivity for speed, but it necessitates the differential system architecture, while in our optical front end system, stringent power consumption requirement prefers the single-ended implementation.

A sketch of the fabricated detector structure is shown in Fig. 2.10. The n-well in which the detector was made forms the active region that senses the incident light. The $p+$, $n+$ diffusion forms the anode and cathode of the photodiode, and are interdigitated with a separation width of s. The width w of $p+$ diffusion was chosen to have the smallest value based on the design rules in order to minimize the collection time of the photogenerated carriers, but trade offs exist since a smaller w leads to larger parasitic resistance of the $p+$ electrodes. The separation width, s, degrades speed because of the increased traveling time for the carriers through depletion region, while smaller s induces the risk of punch through with the high reverse bias voltage across the photodiode. Two versions of photodiodes with different values of w and s have been fabricated intending to find optimum values.



(a)

(b)

Figure 2.10: Photodiode implemented in AMS C35 (a) top view of the interdigitated photodiode (b) crossection of the photodiode.

The photodiode's parasitic capacitance is determined by its size, in other words, by the length and strips of the $p+$, $n+$ diffused regions since the junction capacitance dominates in reverse bias condition. Assuming an input impedance of 50 Ω for the transimpedance amplifier and a -3 dB bandwidth of 5 GHz, the parasitic capacitance of the photodiode can not exceed:

$$C_d = \frac{1}{2\pi R_L f} = \frac{1}{2\pi \times 50 \times 5 \cdot 10^9} = 0.64\,pf \tag{2.8}$$

To reserve some margin, the calculated capacitance should be smaller than 0.3 pf.

The junction capacitance C_d can be estimated as follows:

First the junction potential ϕ_i across the *p*-*n* junction given by:

$$\phi_i = \frac{kT}{q} ln \left(\frac{N_n \cdot N_p}{N_i^2} \right) \tag{2.9}$$

where k is the Boltzmann constant, T is the absolute temperature, q is the electronic charge, N_p and N_n are impurity concentration of *p, n* region, $N_i \approx 1.3 \times 10^{10} cm^{-3}$ is the intrinsic carrier density of silicon at room temperature.

For AMS CMOS 0.35 μm technology, the *p*- substrate effective doping level N_{sub} is $212 \times 10^{15} cm^{-3}$, *n*-well effective doping level N_{well} is $101 \times 10^{15} cm^{-3}$, the shallow junction $p+$ inside n-well has an impurity concentration N_p of $3.3 \times 10^{19} cm^{-3}$ [66, 67]. Since the desired photodiode [3] is formed by the junction between $p+$ and *n*-well, N_p and N_n have the values of $3.3 \times 10^{19} cm^{-3}$, $101 \times 10^{15} cm^{-3}$.

$$\phi_i = \frac{1.38 \cdot 10^{-23} \times 300}{1.6 \cdot 10^{-19}} ln \left(\frac{101 \times 10^{15} \cdot 3.3 \times 10^{19}}{(1.3 \times 10^{10})^2} \right) = 0.96\,V \tag{2.10}$$

With the junction potential known, the total width of the depletion region of an applied reverse bias voltage $\triangle V$ of 9 V can be derived from equation 2.4 :

---

[3] It is noted that the junction between *n*-well and *p*- substrate forms another photodiode which gives a larger responsivity than the desired one. But this photodiode is not utilized in our application because of its slow frequency response.

$$X_d = x_n + x_p = \sqrt{\frac{2\varepsilon_0\varepsilon_i}{q}(\frac{1}{N_n}+\frac{1}{N_p})\cdot(\triangle V+\phi_i)} = w_{d0}\sqrt{1+\frac{\triangle V}{\phi_i}}$$

$$= \sqrt{\frac{2\cdot 8.85\times 10^{-12}\cdot 11.7}{1.6\times 10^{-19}}\cdot(\frac{1}{101\times 10^{15+6}}+\frac{1}{3.3\times 10^{19+6}})\cdot(9+0.96)}$$

$$\approx 0.36\,\mu m \qquad (2.11)$$

where x_n and x_p are the junction depths from the metallurgical junction on the n-type material and p-type material respectively. ε_0 is the dielectric constant in vacuum ($8.85\times 10^{-12} F/m$), ε_i is the relative dielectric constant of the semiconductor, here 11.7 for silicon, $w_{d0}$is the width of the depletion region with zero bias.

The capacitance of the reverse biased *p-n* junction is given by:

$$C_d = \varepsilon_0\varepsilon_i\cdot\frac{A}{X_d} = \varepsilon_0\varepsilon_i\cdot\frac{A}{w_{d0}\sqrt{1+\frac{\triangle V}{\phi_i}}} \qquad (2.12)$$

where A is the junction area. Fig. 2.11 indicates that the junction capacitance is dependent on the reverse bias voltage. As shown, it is desirable to work at higher reverse bias voltage. However, the maximum reverse bias voltage is limited by the break down potential of the photodiode. Since higher doping level lowers the break down voltage [68], this makes the shallow junction based photodiode vulnerable.

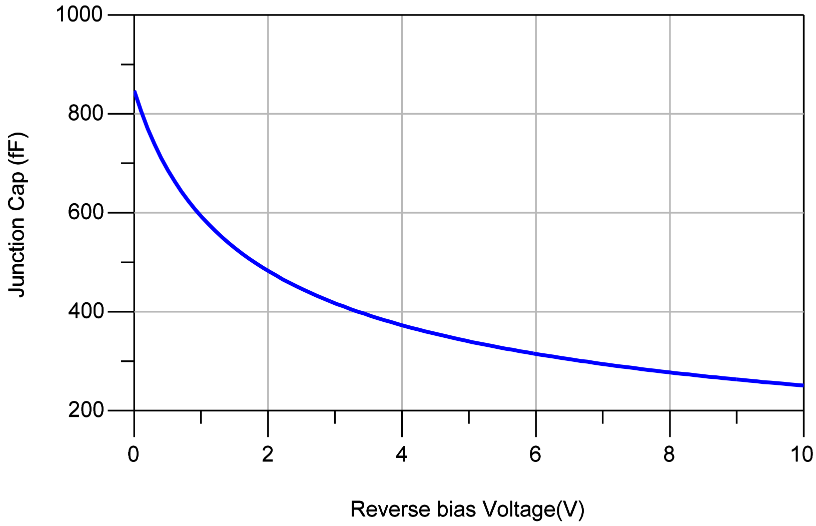


Figure 2.11: Junction capacitance vs. reverse-bias voltage for the chosen CMOS technology (the effective area of the photodiode is 900 um^2)

Referring to [66], the break down voltage across $p+$ diff - n-well is only 9 V. Assuming a photodiode area of $30\,\mu m \times 30\,\mu m$ with reverse bias voltage of 9 V, the junction capacitance of the diode is:

$$C_d = \varepsilon_0 \varepsilon_i \cdot \frac{A}{X_d} = 8.85 \times 10^{-12} \cdot 11.7 \cdot \frac{30 \times 10^{-6} \cdot 30 \times 10^{-6}}{0.36 \times 10^{-6}} \approx 0.26\,pf$$

2.5.2 Simulation

The calculation result ($0.26\,pf$) of the photodiode parasitic capacitance with a diode area of $30\,\mu m \times 30\,\mu m$ is quite close to our target of $0.3\,pf$, and seems to be an optimal value. But it is still not accurate enough since only the area junction capacitance of the diode was included, while the side wall junction capacitance was ignored. In fact, the value of side wall junction capacitance depends heavily on the layout of the photodiode, and not convenient for hand calculation.

For photodiode shaped with interdigitated $p+$ and $n+$ diffusion regions, it would be more convenient to build a diode model and extract the parasitic capacitances from the simulation results. The Cadence Analog Design Environment (ADE) provides the operating point analysis for CMOS transistors. We tried to model the photodiode parasitic capacitance with the bulk junction capacitance of a PMOS transistor (which takes both the area junction capacitance and the side wall junction capacitance into account).

The schematic for simulation is shown in Fig. 2.12. The source, gate, and bulk (n-well) of the PMOS are all connected together to a DC voltage of 9 V except only drain of it is grounded. It assures the operating condition of the drain-bulk junction identical to that of the reverse biased photodiode, and the gate does not effect the junction capacitance in this case for its same potential with the n-well.

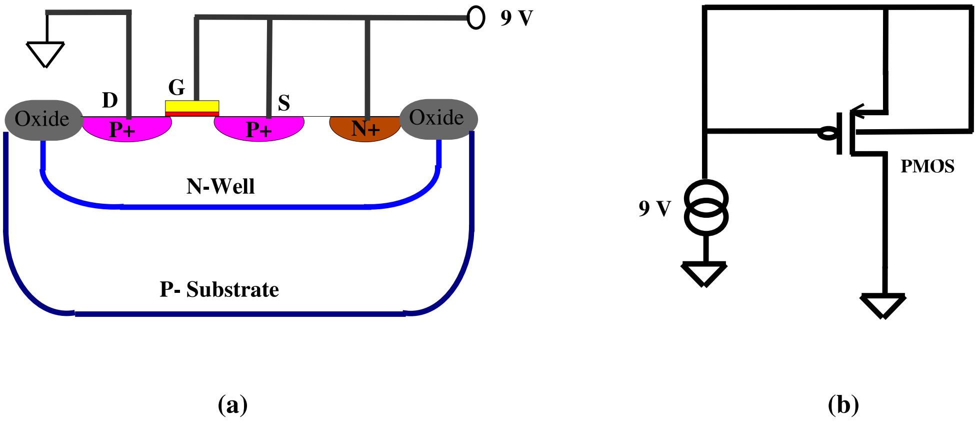


Figure 2.12: Schematic for simulation (a) the cross section of PMOS transistor; (b) schematic for simulation of photodiode's parasitic junction capacitance modeled by a PMOS transistor.

PMOS transistor's diffusion extension width has a default value of 0.85 μm in AMS C35 and can not be changed, while the width w of $p+$ diffusion region in Fig. 2.10 is set to be 2.3 μm and 1.7 μm for the two versions of photodiode. So the simulation results of the extracted parasitic capacitance will be multiplied by the ratio of photodiode electrode width to PMOS transistor's default diffusion extension width. In simulation file, the width of PMOS transistor is fixed as

$30\,\mu m$ based on the hand calculation result, and the number of finger is fixed to 1. The bias condition of 9 V is used to emulate the maximum voltage that can be applied across the photodiode. With the tool of result browser in Analog Design Environment, the drain to bulk junction capacitance of the PMOS transistor , C_{jd}, can be extracted.

Table 2.1 shows the extracted C_{jd} of the PMOS transit or with different width and different transistor model. The parasitic capacitances have smallest values in worse power model (cmoswp) and biggest ones in worse speed model (cmosws), which is in line with the concept that larger parasitic capacitance leads to slower frequency response. here cmostm means CMOS typical means model.

capacitance	$10\mu m$	$20\mu m$	$30\mu m$	$40\mu m$	$50\mu m$	$60\mu m$
cmostm	4.6	9.0	13.4	17.9	22.3	26.7
cmosws	5.2	10.1	15.1	20.0	24.9	30.0
cmoswp	4.1	7.9	11.8	15.7	19.6	23.5

Table 2.1: Extracted PMOS drain-bulk parasitic capacitance (fF)

The strip number N_f in Fig. 2.10 is given by :

$$N_f = \frac{C_{target}}{C_{jd} \cdot r} = \frac{0.3\,pf}{C_{jd} \cdot r} \tag{2.13}$$

where r is the ratio of the width of the electrodes to the PMOS diffusion extension width in simulation. Table 2.2 shows the derived strip number N_f of photodiodes with different strip width based on table 2.1. In practical design of the photodiode, the strips length and number have been finely tuned to make the layout of the photodiode have a square shape. The total parasitic capacitance of the photodiode in both versions has been kept close to 0.3 pf. The final strip numbers are 6 and 8 for version 1 and version 2. But both strip length are around 40 μm.

strip nums	$10\mu m$	$20\mu m$	$30\mu m$	$40\mu m$	$50\mu m$	$60\mu m$
ver 1	24	12	8	6	5	4
ver 2	33	17	11	8	7	6

Table 2.2: Calculated strip numbers of photodiode based on the extracted C_{jd}

2.5.3 Implementation

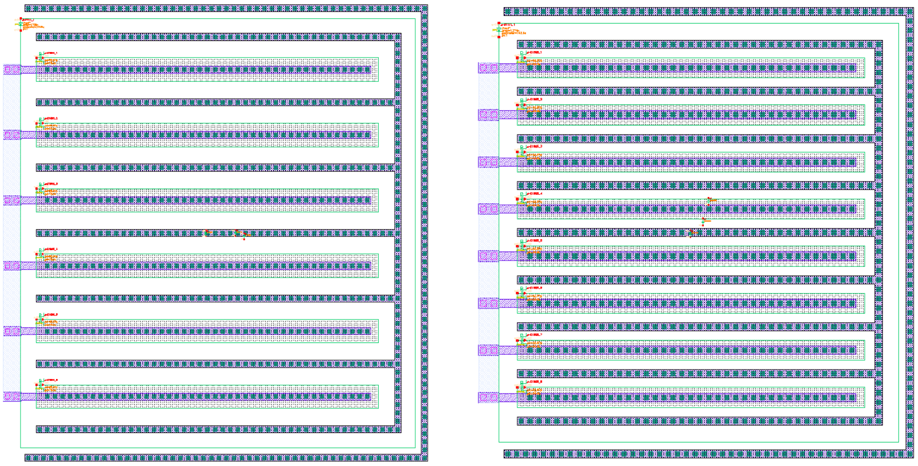
Based on previous hand calculations and simulation results, two versions of photodiode have been fabricated in AMS 0.35 μm CMOS process[4], which are shown in Fig. 2.13. In version 1, the width of the electrode ($p+$ diffusion) w are chosen to be 2.3 μm, with separation distance between $p+$ diffusion and $n+$ diffusion s of 1.7 μm. These values are quite conservative compared with the calculated depletion width of 0.36 μm , but it guarantees that the photodiode will not break down with the large reverse bias voltage. In version 2, w and s are shrunk to 1.7 μm and 0.8 μm, which is believed to improve the speed, and responsivity of the photodiode in some degree, since the carriers travel less distance and more strips fit in within the same area. But it is associated with the risk of Punch through under high reverse bias voltage. Punch through occurs when the depletion regions of two adjacent p-n junction merge into a single one, the diode fails to function in this case.

PD	w	s	strips	area	C_{jd}	note
ver. 1	2.3 μm	1.7 μm	6	$41\mu m \times 42\mu m$		integrated
ver. 2	1.7 μm	0.8 μm	8	$36\mu m \times 36\mu m$		test chip only

Table 2.3: Comparison of the two fabricated photodiodes

---

[4] More test chips of the photodiode would be beneficial for optimization of the photodiode's performance, but it is limited by the fabrication cost and the available silicon space of the final tape out.



(a) (b)

Figure 2.13: Photodiodes layout in AMS C35 (a) version 1; (b) version 2.

The layout of the photodiode test circuits are shown in Fig. 2.14. The test structures are shielded with top metal (Metal 4 in AMS C35) except the photodiode window to minimize the influences of the incident light on the circuit functional devices. Since photodiodes are to be measured at high frequency domain, G-S-G mode pads are adopted for the output of the photodiodes ($p+$ diffusion in Fig. 2.10), which give 50Ω impedance match to the testing instrument. There is no need of impedance match for the anodes ($n+$ diffusion's in Fig. 2.10, which define the potential of n-well), because this port is connected to a fixed potential, *i.e.* DC 9 V for the reverse bias voltage across the photodiode. However, on-chip decoupling capacitors in parallel with the anode output are necessary to cancel the effects of parasitic inductance of the needle probe.

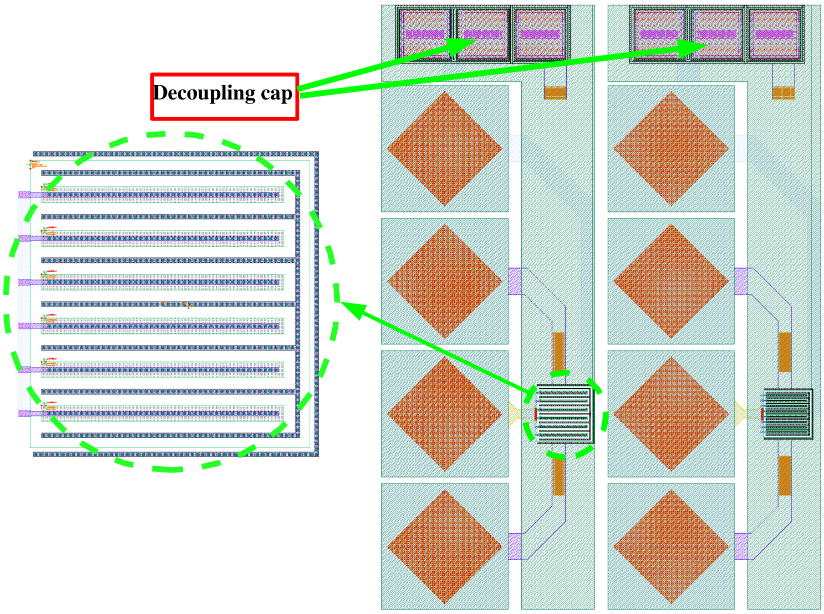


Figure 2.14: Photodiode test circuits for the two versions. Photodiode are zoomed in left-sided. G-S-G pads are used here for cathode outputs of the photodiodes which give 50 Ω impedance match to measurement setup; the anodes are taken out by normal analog pads, with 3 pf on-chip decoupling capacitors.

With the slow photogenerated carriers from the substrate screened by n-well, the photodiode's bandwidth is dominated by the transit time, which refers to the time taken by the carriers to drift across the high field depletion region. the transit time τ_{tr} is then described as :

$$\tau_{tr} = \frac{X_d}{\upsilon_s} \tag{2.14}$$

where υ_s are the carriers (electrons or holes) saturation velocity in silicon. As one applies an electric field to a semiconductor, the electrostatic force causes the carriers to first accelerate and then reach a constant average velocity, υ, as the carriers scatter due to impurities and lattice vibrations. The velocity saturates at high electric fields reaching the saturation velocity. Generally, when the electrical field exceeds $10^5\,Vcm^{-1}$, both electrons and holes have a saturation velocity of about $10^7\,cm/s$ at room temperature ($300K$) [69, 70].

Further we relate the transit time τ_{tr} with the time constant τ_0

$$\tau_{tr} = \tau_0 ln \frac{0.9}{0.1} \approx 2.2\tau_0 \tag{2.15}$$

In AMS $0.35\mu m$ process, the n-well depth is 2.0 *um*, the $p+$ junction inside n-well is about 0.2 *um*. The depletion width of the photodiode, with a reverse bias voltage $\triangle V$ of 9 V , can be derived from equation (2.4) as 0.36 *um*.

The transit time dominated -3 *dB* bandwidth of the photodiode then is :

$$f_{-3dB} = \frac{1}{2\pi\tau_0} = 0.34 \cdot 1/\tau_0 > 10\,GHz$$

2.6 Measurement

2.6.1 Laser source

Most of the high speed commercial laser diodes are found in applications for high volume, long haul optical communications. The emission light wavelength are generally around 1300 *nm* or 1500 *nm* for low loss and dispersion, and the output power is on the order of 100 mW. Typical product such as ML9XX37 single mode distributed feedback (DFB) laser diode from Mitsubishi Electric & Electronics, supports up to 40Gbps data rate within temperature range from 0 to 50 degree Celsius. Expensive butterfly package, incorporating thermal-electric cooler system and back facet monitor diode are employed to meet the stringent requirements for long haul communications.

On the other hand, the development of low cost sources with low power consumption hastens a new concept of laser, the VCSEL (Vertical Cavity Surface Emitting Laser). VCSELs are very well adapted to a great variety of fiber communication systems due to some inherent advantages of the structure over that of edge-emitting lasers, i. e., intrinsically low and circular divergence for the output beam; efficient coupling, the increase of positioning tolerances and high density of device integration resulting a sharp cost reduction since test, mounting, and assembling time represent more than half of the cost of an optical source; lower power consumption as well as the possibility of operation over a wide range of temperature. Today VCSEL has found popular applications in short distance optical communication systems, partially because of its emission light wavelength of 850 *nm* which limits its application for long haul communications.

Two kinds of VCSEL diodes have been used for characterizations of the high speed photodiodes fabricated in AMS C35, HFE4080-32X-XBA and HFE6391-561, both from Honeywell (later taken over by AOC (Advanced Optical Components Ltd)). HFE4080-32X-XBA is a high radiance VCSEL packaged on a TO-46 header with a metal can. It produces a circularly symmetric, narrow divergence light beam which is coupled into the industry standard ST fiber connector (Fig. 2.15). HFE4080-32X-XBA has an analog bandwidth of over 6GHz at a small signal sinusoidal modulation input, which suits our 5GHz photodiode test purpose. Another important feature for HFE4080-32X-XBA is its low drive current, the operating forward current is only 10 mA which makes direct drive possible and simplifies the test module. The important technical parameters of HFE4080-32X-XBA are listed in table 2.4.

Parameter	value	unit	test conditions
output power	800	μW	
threshold current	3.5	mA	
slope efficiency	0.3	mW/mA	$I_F = 10$mA
forward voltage	1.75	V	$I_F = 10$mA
reverse breakdown voltage	10	V	$I_R = 10\mu$A
wavelength	850	nm	$I_F = 10$mA
rise and fall time	100	ps	T=25°C, 10-90%
analog bandwidth	6	GHz	$I_F = 10$mA
series resistance	30	Ω	DC

Table 2.4: Main performance parameters of HFE4080-32X-XBA

To drive the HFE4080-32X-XBA, a test board was made in the EEE department, University of Nottingham, which connects the cathode and anode of the laser diode to 50Ω output impedance of high frequency signal generator. The picture of the test board is shown in Fig. 2.15. There is only one important device on the test board, the 50Ω transmission line which connects the anode of the laser to a standard SMA connector. The important technical parameters for the PCB manufacturing include the material (FR4), laminate thickness (1.6 mm) and the dielectric constant of FR4 (4.5). Modelled as a microstrip, the width of the transmission line can be calculated by the LineCal in ADS to be 2.8 mm when 50Ω impedance is expected.

Besides laser diode HFE4080-32X-XBA, another faster laser diode, HFE6391-561 from Advanced Optical Component Ltd., was also used as the high frequency modulated optical source in the photodiode characterizations and the optical receiver measurements. Compared with HFE4080-32X-XBA, HFE6391-561 has a larger bandwidth which supports data rate up to 10Gbps, and it consumes less DC current. Finally another important feature of HFE6391-561 is its flexible interface to the evaluation board (Fig. 2.16). Important technical parameters in the data sheet of HFE6391-561 are summarized in table 2.5.

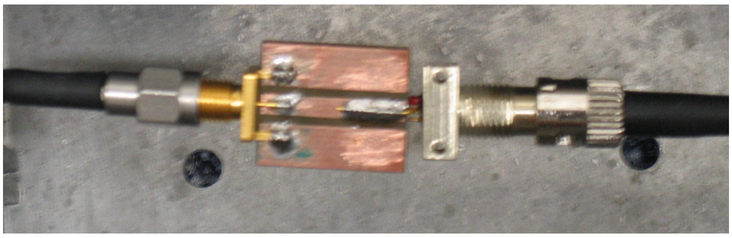


Figure 2.15: Picture of the test board for laser diode HFE4080-32X-XBA

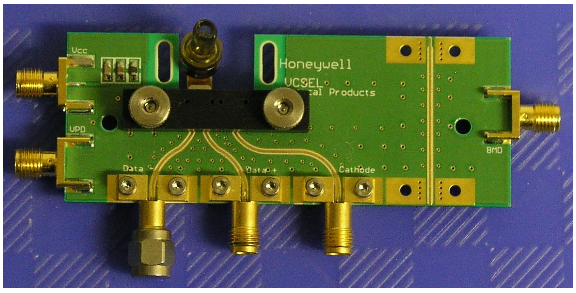


Figure 2.16: Picture of the evaluation board for laser diode HFE6391-561

Parameter	value	unit	test conditions
output power	600	μW	
threshold current	1	mA	
slope efficiency	0.075	mW/mA	$I_F = 6.5$mA
forward voltage	1.8	V	$I_F = 6.5$mA
reverse breakdown voltage	10	V	$I_R = 10\mu$A
wavelength	840	nm	$I_F = 6.5$mA
rise and fall time	40	ps	T=25°C, 20-80%
Total	0.5	pF	forward biased
series resistance	60	Ω	$I_F = 6.5$mA

Table 2.5: Main performance parameters of HFE6391-561

2.6.2 DC characterizations of PDs

As mentioned in section 2.4, two kinds of high speed photodiode were fabricated in AMS C35 process. Their DC performances were first characterized, i.e., dark current, I-V curve, responsivity etc. The setup of the DC measurement of the photodiodes is shown in Fig. 2.17. Two voltage sources (Keithley 230) were used to provide the DC bias of the DUT (photodiode) and the laser source (HFE4080-32X-XBA). For monitoring purpose, two current sources (Keithley 617) are connected in series with the voltage sources to make sure that the biasing conditions are in line with the specifications. The die micrograph of the photodiodes can be seen in Fig. 2.18.

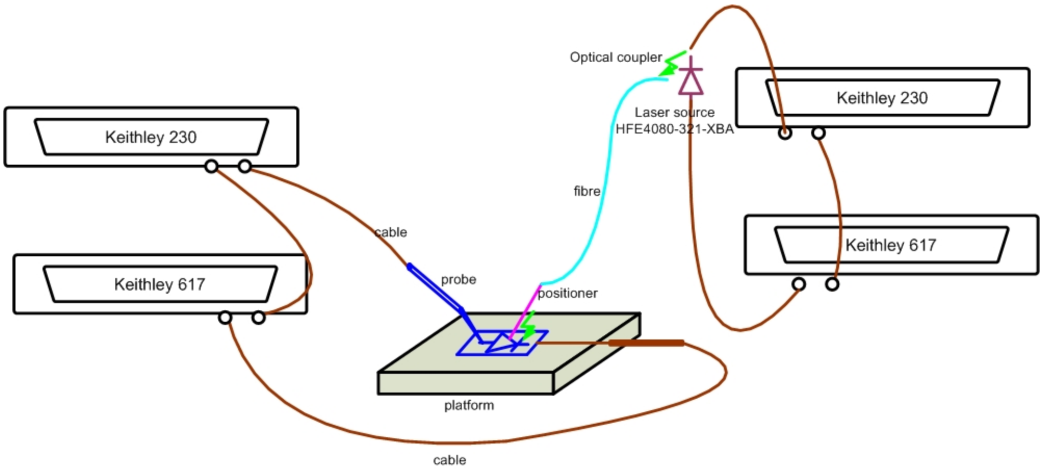


Figure 2.17: Photodiode DC test set up

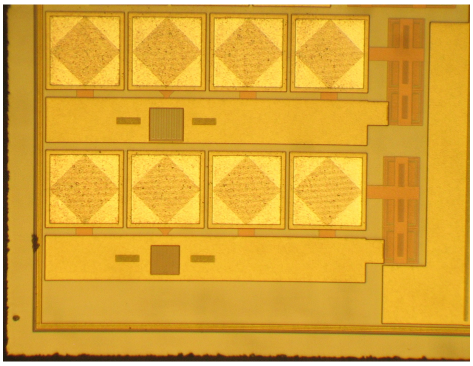


Figure 2.18: Die micrograph of the photodiodes

Prior to the optical characterization for the photodiodes, I-V curves of the diodes were first derived from the DC test of the photodiodes as a functionality verification. For comparison, Fig. 2.19 shows the measured I-V curves of p+ nwell photodiode and nwell-sub photodiode. It is noted that the two versions of photodiode (namely, PD1 and PD2) give almost the same I-V curves as they have almost the same saturation I_s.

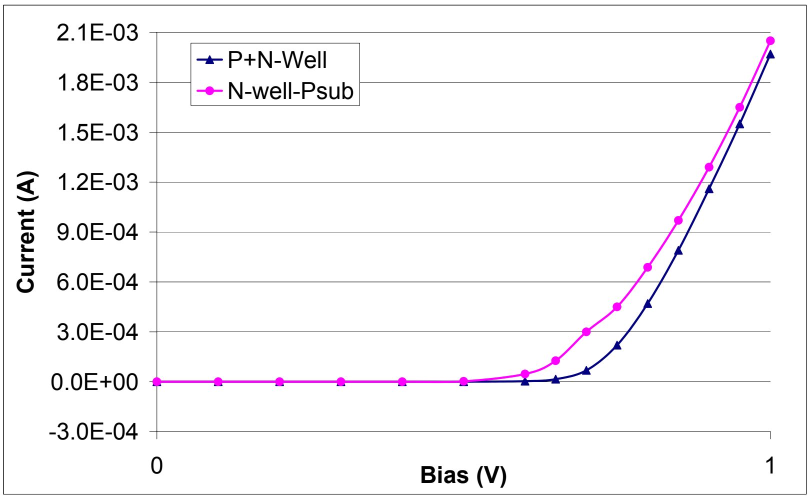


Figure 2.19: I-V curves of the p+ nwell and nwell psub photodiode

The dark current of the two photodiodes were then measured with all the lights off in the measurement area (since these devices were measured on wafer, it is not applicable to totally shield the test chip by a black box because of the the probes, positioners, cables and fibres) . With the reverse bias voltage varying from 1V to 5V, the dark current for the photodiodes change from 0.5 pA to 77 pA, as shown in table 2.6.

PD version	1V	2V	3V	4V	5V
PD1 ($42\mu m \times 42\mu m$)	0.5 pA	7 pA	12 pA	32 pA	77 pA
PD2 ($36\mu m \times 36\mu m$)	0.3 pA	3 pA	9 pA	25 pA	65 pA

Table 2.6: Dark current of the photodiodes with different reverse biased voltage applied

Responsivity is an important specification for photodiodes. The 850 nm light generated from the laser source HFE4080-32X-XBA was coupled to 50 μm fibre with standard ST optical adapter while the other end of the fibre is flat cleaved for the convenience of alignment. Since the photodiodes fabricated on chip have square shapes with the size of $42\mu m \times 42\mu m$ and $36\mu m \times 36\mu m$ respectively, they can not be fully covered by the 50 μm fibre. Further more, some safety distance between the flat cleaved fibre and the die surface should be reserved, 50 μm was estimated in this measurement. The fibre used presents a NA (numerical angle) of 0.2. With these parameters, the coupling efficiency from the end of the flat cleaved fibre to the photodiode 1 (PD1, $42\mu m \times 42\mu m$) under test can be calculated as below

$$\begin{aligned} \eta_1 &= 42\mu m * 42\mu m/\pi * (25\mu m + 50\mu m * tg\theta)^2 \\ &= 45\% \end{aligned}$$

where the $tg\theta$ can be calculated to be 0.204 from NA of the fibre being equal to 0.2 ($sin\theta = 0.2$). Similarly, the coupling efficiency of photodiode 2 (PD2, $36\mu m \times 36\mu m$) is

$$\begin{aligned} \eta_2 &= 36\mu m * 36\mu m/\pi * (25\mu m + 50\mu m * tg\theta)^2 \\ &= 33\% \end{aligned}$$

For calibration purpose , first the optical output power from the flat cleaved end of the fibre was first measured by Anritsu ML9001A optical power meter. The output current from the photodiodes under test then were measured with Keithley current meter. The responsivity was calculated with the following equations

$$\Re = \frac{I_{PD}}{P_{in}} \cdot \frac{1}{\eta} \tag{2.16}$$

where I_{PD} denotes the output DC current from the photodiode; P_{in} denotes the measured optical power at the end of the flat cleaved fibre; η is the coupling efficiency from the end of the fibre to the photodiode, here equal to 45% and 33% for PD1 and PD2, respectively.

It can be seen from all three photodiodes' response that with higher reverse biased voltage applied, better responsivity can be achieved. This is because when the reverse biased voltage across the photodiode increases, the depletion region expands correspondingly, resulting the increase of responsivity. For PD1, it gives a responsivity of 0.011 A/W at 1 V reverse biased voltage, and 0.016 A/W at 5 V reverse biased voltage; PD2 gives a responsivity of 0.013 A/W at 1 V reverse biased voltage, and 0.019 A/W at 5 V reverse biased voltage. As expected, the n-well psub photodiode demonstrates the best responsivity among the three, the responsivity is 0.30 A/W at 1 V reverse biased voltage, and 0.31 A/W at 5 V.

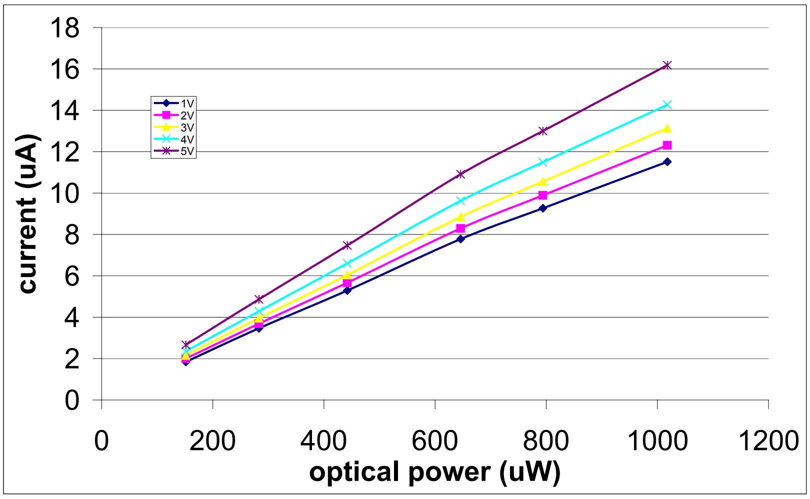


Figure 2.20: Responsivity of PD1 with different reverse biased voltage applied

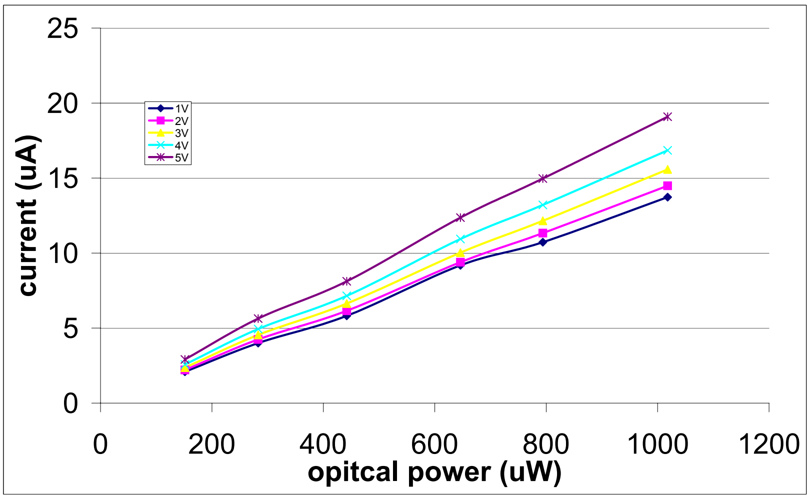


Figure 2.21: Responsivity of PD2 with different reverse biased voltage applied

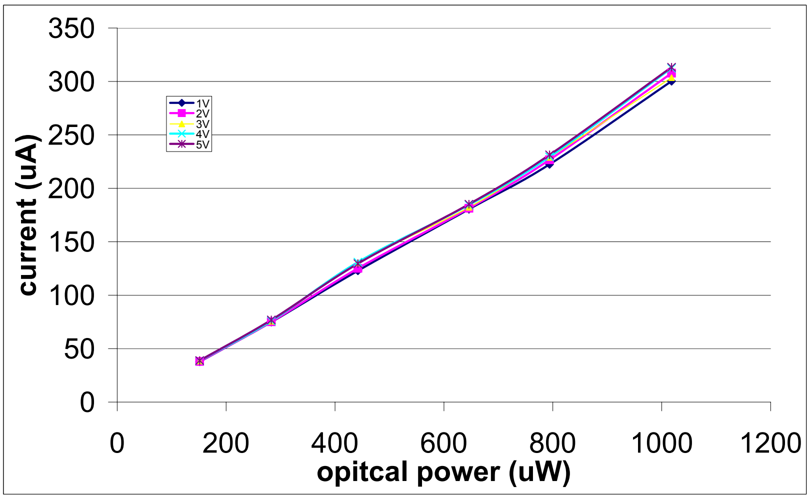


Figure 2.22: Responsivity of nwell-psub photodiode with different reverse biased voltage applied

2.6.3 High frequency characterizations of PDs

To characterize the high frequency performances of the photodiodes, the test set up was changed and it is shown in Fig. 2.23.

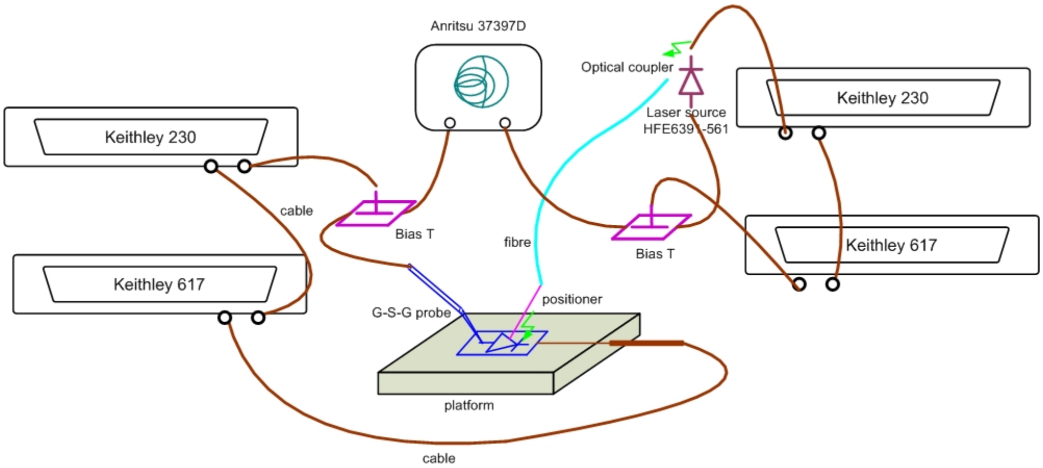


Figure 2.23: Photodiode high frequency response measurement set up

The frequency response of the two versions of photodiodes (PD1 and PD2) were measured with 10 Gbps laser source HFE6391-561. Anritsu VNA 3739D was used as the high frequency signal generator which drives the laser HFE6391-561 directly with RF power fixed at 3 dBm. A bias Tee was placed between the high frequency generator and laser source to give a DC bias, 6 mA in this case. The output from the laser diode was then coupled into a 50 μm fibre through an optical adapter. The alignment of the optical fibre to the photodiode on chip was controlled by a x-y-z three dimensional positioner. For high frequency measurement, the output of the photodiode (anode) needs to be terminated by 50 Ω, hence a 50 Ω G-S-G probe made by cascade microtech was employed here, which features a pin separation of 100 μm. Similar to the biasing purpose of the laser source driving, another bias Tee was necessary at the anode of the photodiode. After the bias Tee, the signal was fed into another port of the VNA, as shown in Fig. 2.23. With this set up, the S11 parameters gives the input impedance variation of the lase diode from 50 Ω while S21 parameters then can be used to describe the frequency response of the photodiode.

The frequency responses of the photodiodes were measured under different bias voltage. Fig. 2.24 gives the typical frequency response of PD1 at a reverse biased voltage of 9 V. The output power measured from the photodiode has been normalized since we have no reference photodiode to characterize the power loss of the laser source. Anyway, we still can extract the speed performance (-3 dB bandwidth) of the photodiodes under test from the diagram of output power changes versus frequency sweep. It should be noted here, the curves in Fig. 2.24 was not de-embedded from the effects of laser source's frequency response, although its bandwidth is up to 10 GHz as mentioned in its data sheet.

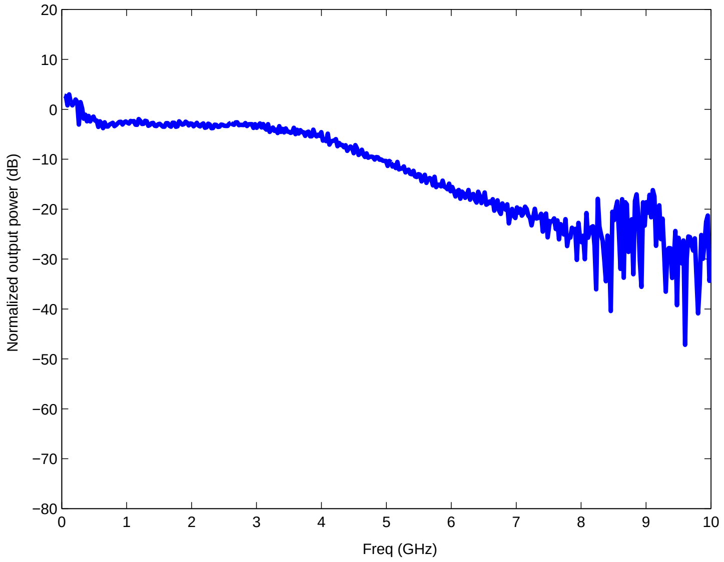


Figure 2.24: Frequency response of PD1 with reverse biased voltage of 9 V.

At very low reverse biased voltage (i.e., 1 V and 3 V), the output spectrum presents slowly dropping slope, which could not be solved by increasing the RF sweep power. This could be caused by the very thin depletion width of the photodiode, in this case, the slow diffusion photogenerated carriers dominates the current through the photodiode. Fig. 2.25 and Fig. 2.26 show the measured results of the frequency response of the two photodiodes under different reverse biased voltages. No clear bandwidth increase can be observed from the increase of the reverse biased voltage. This further proves that the bandwidth of the CMOS photodiode is not dominated by the traveling time through the depletion region, but by how much of the slow diffusion carriers being blocked. If all the diffusion carriers are screened from the depletion region, the frequency response of the photodiode can be modeled by one pole (in transfer function) system, where the dominant pole is caused by the time constant of the carriers traveling across the depletion region. Finally it should be noted here, that because of the normalization on the output power, the curves do not reflect the responsivity changes to the reverse biased voltage. In fact, with reverse biased voltages increases from 7 V to 9 V, the measurement showed an output power increment by around 10 dB.

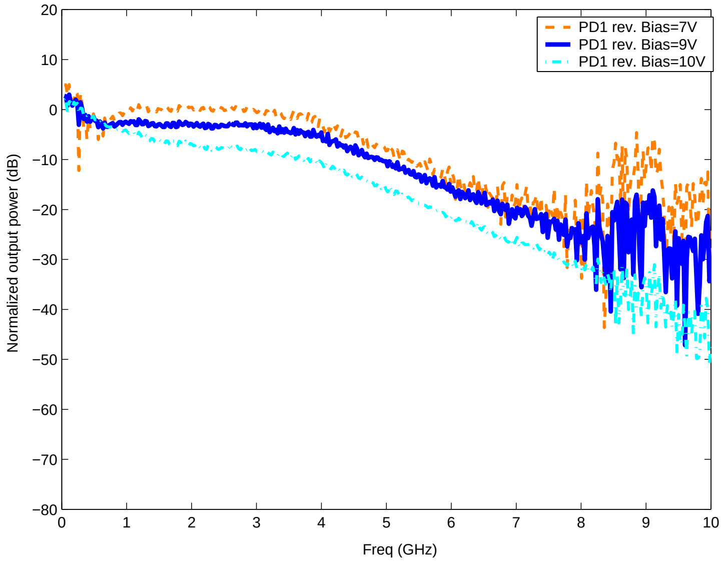


Figure 2.25: Frequency response of PD1 with reverse biased voltage of 7 V, 9 V and 10 V.

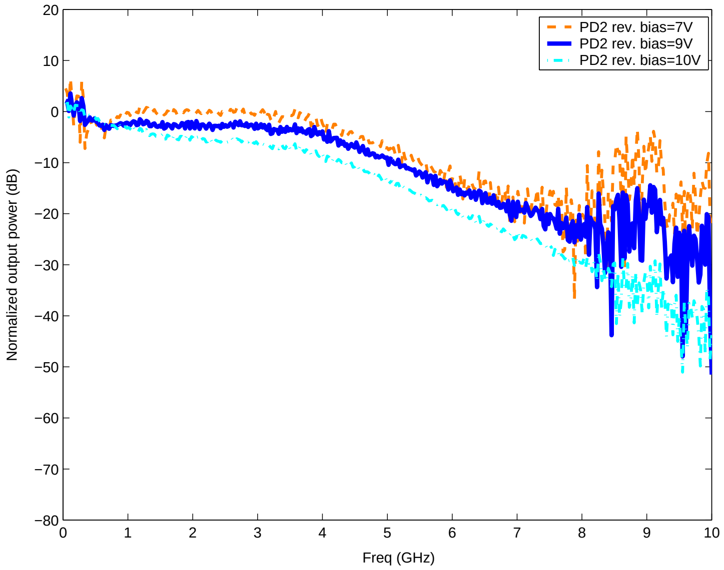


Figure 2.26: Frequency response of PD2 with reverse biased voltage of 7 V, 9 V and 10 V.

In Fig. 2.27, comparison of the frequency response between PD1 and PD2 was given. Here no normalizations were applied. With 3 dBm RF power from VNA 37397D to the laser source, and then the light being conducted through $50\mu m$ optical fibre and finally aligned to the photodiode on chip, around -70 dBm output power was measured at a reverse biased voltage of 9 V for both PD1 and PD2. Basically PD1 presents a -3 dB bandwidth of 4.5 GHz at a reverse biased voltage of 9 V, while PD2 presents a -3 dB bandwidth of 4.6 GHz.

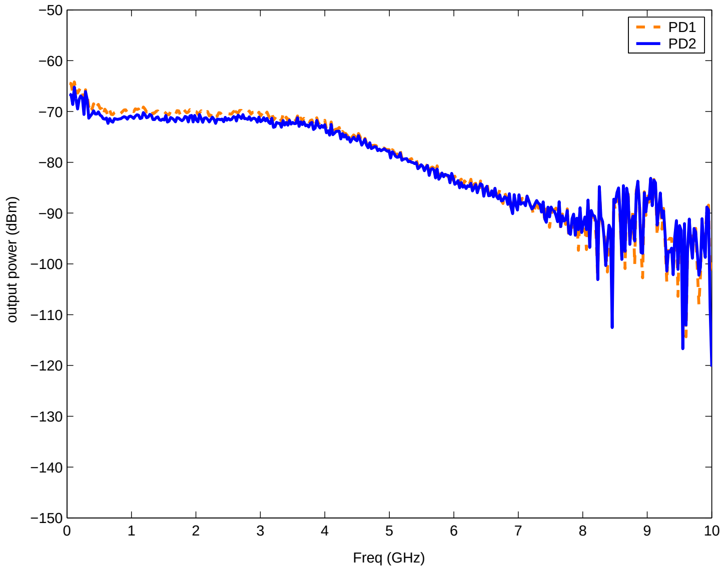


Figure 2.27: Measured output power of both PD1 and PD2 at reverse biased voltage of 9 V, the RF input power is 3 dBm to laser source HFE6391-561.

In all the measured results, the outputs at high frequency (higher than 8 GHz) are relatively noisy, which is believed to be caused by the signal being immersed in the noise floor. Further more, in the frequency range from 40 MHz (which is the VNA's lowest frequency limit) to around 500 MHz, the measured output was not stable and flat, compared to the response at frequency range from 500 MHz to around 8 GHz. The reason for this instability is still not clear, but one of the potential possibilities is the effect from the laser source, as mentioned earlier that its frequency response effect has not been deducted in all the measured results. The measured performances of the photodiodes are summarized in table 2.7.

parameter	PD1	PD1
size	$36\mu m \times 36\mu m$	$42\mu m \times 42\mu m$
dark current	$77pA$ @ -5 V	$65pA$ @ -5 V
responsivity	$0.016A/W$ @ -5 V	$0.019A/W$ @ -5 V
- 3 dB bandwidth	4.5 GHz @ -9 V	4.6 GHz @ - 9 V

Table 2.7: The measured performances of the photodiodes

Chapter 3

Transimpedance Amplifier

3.1 Introduction

In optical sensing systems, the generated current from photodiodes are generally small and most of the subsequent processing occurs in the voltage domain, thus it needs to be converted into voltage. A current to voltage converter is also called a transimpedance amplifier (TIA). The simplest circuit is a resistor which is added to the output node of the photodiode, as depicted in Fig. 3.1 (a). The small signal equivalent circuit is shown as Fig. 3.1 (b). This circuit provides a transimpedance gain of R_L, and a time constant of R_LC_d, where R_L is the resistance of the resistor connected, C_d is the parasitic capacitance of the photodiode. Modelling the thermal noise of R_L as $\overline{I_n^2} = 4kT/R_L$, we have the noise voltage $\overline{V_{n,out}^2}$ at the output :

$$\overline{V_{n,out}^2} = \int_0^\infty \frac{4kT}{R_L} \left| R_L \parallel \frac{1}{j \cdot 2\pi f \cdot C_d} \right|^2 df \tag{3.1}$$

$$= \int_0^\infty \frac{4kT}{R_L} \frac{R_L^2}{R_L^2 C_d^2 \cdot 4\pi^2 f^2 + 1} df \tag{3.2}$$

Noting that

$$\int \frac{1}{1+u^2} = tan^{-1}u \tag{3.3}$$

we can write

$$\begin{aligned} \overline{V_{n,out}^2} &= \frac{2kT}{\pi C_d} tan^{-1}u \mid_0^\infty \\ &= \frac{kT}{C_d} \end{aligned} \tag{3.4}$$

This equation reveals that the total integrated noise is independent of R_L. It is because, as R_L increases, the increment of the transimpedance gain cancels the effect of the degradation of $\overline{I_n^2}$, resulting in a constant $\overline{V_{n,out}^2}$. For a fair comparison of different designs [71], the noise is always referred to the input so that it does not depend on the gain[1]. Since the circuit of Fig. 3.1 (a) has a transimpedance gain of R_L, its total input-referred noise current is equal to

$$\overline{I_{n,in}^2} = \frac{\overline{V_{n,out}^2}}{R_L^2} \tag{3.5}$$

$$= \frac{kT}{R_L^2 C_d} \tag{3.6}$$

indicating that $R_L$must be maximized. However remember that the bandwidth of the circuit is related to the time constant $R_L C_d$ as

$$f_{-3dB} = \frac{1}{2\pi R_L C_d} \tag{3.7}$$

---

[1] Input-referred noise is defined as the value that, if applied to the input of the equivalent noiseless circuit, produces an output noise equal to that of the original, noisy circuit.

The fundamental trade-offs expressed by equations (3.6) and (3.7) suggest that the diode/resistor network of Fig. 3.1 is ill-suited to high-performance applications, and other circuit topologies, named as TIAs, must be used in optical front end to relax these constraints.

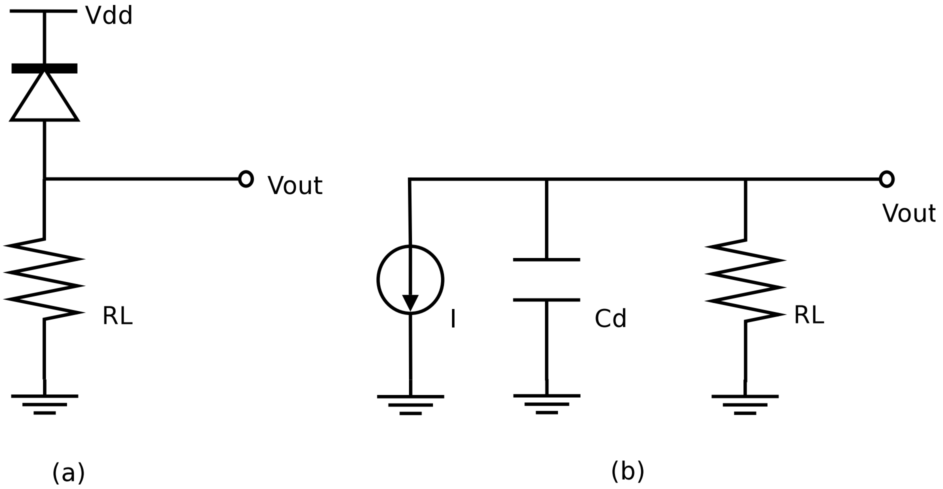


Figure 3.1: Simple resistive optical front end

3.2 Overview of the Transimpedance Amplifier

The strong trade-off between the bandwidth and sensitivity of a front-end with a simple resistor makes it impractical for many applications. The effective input resistance of the front-end can be reduced significantly by adding an active component to the design, resulting a transimpedance architecture. A Transimpedance Amplifier (TIA) is an analog front-end with reduced input impedance and a relatively high current-to-voltage gain. The addition of active components, like transistors, will add to the noise. However, with a careful design, low noise, high gain and high bandwidth TIAs can be achieved. Detailed analysis of TIAs are covered in numerous publications [72, 73, 74, 75, 76, 77, 78]. For TIAs, like any other receiver, the most important specifications are noise, bandwidth, sensitivity, gain, output impedance and power consumption.

Noise The noise performance of TIA is generally quoted as the input-referred noise current, $I_{n,in}$, which determines the minimum input current that yields a recognizable output voltage. The stringent TIA noise requirements limit the choice of circuit topologies and in particular the number of devices in the signal path. As will seen throughout this chapter, the problem of noise becomes more severe when applications come to high frequency range.

Bandwidth The bandwidth of TIA refers to the frequency at which the transimpedance gain of the amplifier drops by 3 dB. The challenges of design of broadband TIAs differ from that of narrow band amplifier design used in lots of wireless communication systems. The later requires a good impedance match and a high gain for only a narrow band, whereas the former demands a high, flat gain over a wide frequency range, and there are always trade-offs between the gain and the bandwidth of the amplifier if it is power constrained.

Gain A transimpedance amplifier converts an input current, I_{in}, to an output voltage, V_{out}. The circuit is characterized by a "transimpedance gain", defined as $R_T = \partial V_{out}/\partial I_{in}$. For example, a gain of 1 kΩ means the TIA produces a 1-mV change in the output in response to a 1-μA changing in the input. High gain TIA design is always desirable for high speed applications due to the signal loss in optical communication or sensing systems and the low responsivity of the photodetector.

Input and Output Impedance In order to isolate the photodiode capacitance from the TIA, the input impedance of the TIA is normally designed to be as small as possible, whereas an output impedance of 50-Ω is generally required for a stand-alone TIA over the entire interested frequency range in order to match the subsequent stage or the test equipment. A S_{22} of lower than -10dB is acceptable in most cases.

Power Consumption Most of the high speed optical receivers have a larger power consumption compared to that of the Intermediate-Frequency (IF) analog circuits because of the large bandwidth required. But the power consumption of the TIA is limited by the system specification, particularly when it is integrated with other functional circuits.

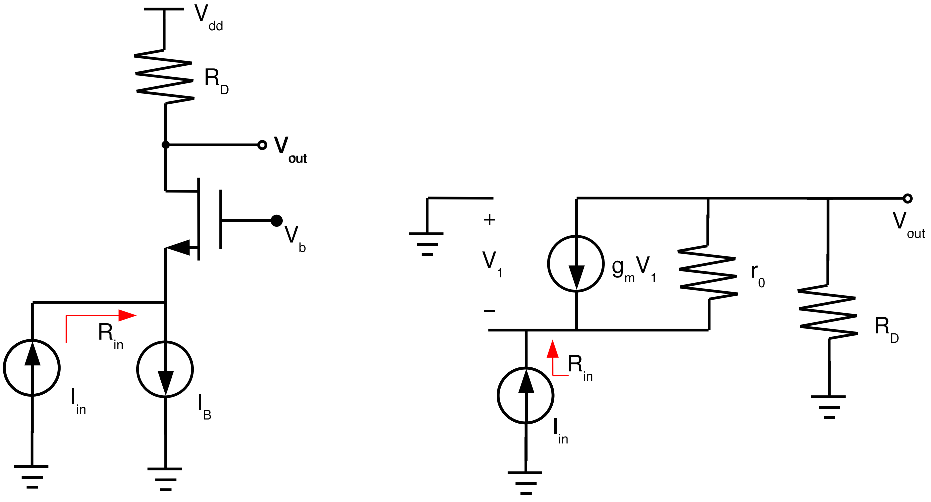
The stringent gain-bandwidth and noise-bandwidth trade-offs prohibit complex configurations that introduce many devices in the signal path, constraining TIA circuits to primarily two topologies: open-loop (common-gate/base) stages and current-feedback (shunt-shunt) feedback amplifiers.

3.2.1 Common-Gate TIA

An amplifier stage that exhibits a low input impedance is the common-gate (CG) (for field-effect devices) or common-base (CB) (for bipolar devices) topology. A common-gate stage transimpedance amplifier is shown in Fig. 3.2 (a). Neglecting second-order effects in the transistors, it is well known that the input resistance is approximately equal to $1/g_m$, where g_m denotes the transconductance of the input transistor. In fact, it ignores the load effect, below a detailed analysis of the common-gate stage performances as a transimpedance amplifier is given. Excluding body effect (g_{mb}) and assume I_B is an idea current source, the small-signal equivalent circuit of the CG stage is shown in Fig. 3.2(b).

Transimpedance gain Since all of I_{in}flows through R_D the transimpedance gain is

$$R_T = R_D \tag{3.8}$$



(a) (b)

Figure 3.2: Common-gate stage (a) schematic (b) small-signal equivalent circuit.

Input impedance To obtain the input resistance, which is equal to $-V_1/I_{in}$, we recognize that the current through r_o is equal to $(I_{in}+g_mV_1)$. Adding the voltage drops across r_o and R_D , equating the result to $-V_1$, we have

$$(I_{in}+g_mV_1)r_o+I_{in}R_D=-V_1 \tag{3.9}$$

Thus,

$$R_{in} = \frac{-V_1}{I_{in}} \tag{3.10}$$

$$= \frac{r_o+R_D}{1+g_mr_o} \tag{3.11}$$

Since typically, $g_mr_o \gg 1$,

$$R_{in}=\frac{1}{g_m}+\frac{R_D}{g_mr_o} \tag{3.12}$$

The second term in equation (3.12) indicates that the input resistance of the common-gate stage depends on load resistance R_D. Only when $R_D \ll r_o$, R_{in} can be simply equal to $1/g_m$. With short-channel devices, on the other hand, this may not be the case.

Bandwidth With input resistance known, it is easy to derive the -3 dB bandwith

$$\begin{aligned} f_{-3dB} &= \frac{1}{2\pi R_{in} C_D} \\ &\approx \frac{g_m}{2\pi C_D} \end{aligned} \tag{3.13}$$

where C_D denotes the total capacitance at the input node, including the photodiode capacitance and the gate-source capacitance of M1. In a typical design, the magnitude of input pole, g_m/C_D, is smaller than that of the output pole, $(R_D C_{out})^{-1}$, since the photodiode capacitance is quite large (ranging from a few hundred fF to several pF depending on applications) . Therefore the frequency response is usually determined by the input.

Noise In order to find out the noise performance of the CG stage, Fig. 3.3(a) depicts a more realistic implementation, with M2 operating as the bias current source. Adding the noise source as in Fig. 3.3(b) and neglecting channel-length modulation and body effects ($r_o = \infty$ and $g_{mb} = 0$) for simplicity, we compute the output noise current by superposition that from each noise sources. All $\overline{I^2_{n,M2}}$ flows through R_D, generating an output noise of $R_D^2\overline{I^2_{n,M2}}$; $\overline{I^2_{n,M1}}$ doest contribute to the output because it consists a circuit loop with M1, based on Kirchhoff's law, the current through R_D equals to that through M2, which means $\overline{I^2_{n,M1}}$ does not appear at the output; Finally all of $\overline{I^2_{n,RD}}$ flows through R_D since we have assumed the output resistance of M1, r_o, is infinite. It follows that the output noise voltage is

$$\overline{V^2_{n,out}} = (\overline{I^2_{n,M2}} + \overline{I^2_{n,RD}})R_D^2 \tag{3.14}$$

$$= 4kT(\gamma g_{m2} + \frac{1}{R_D})R_D^2 \tag{3.15}$$

Dividing this quantity by the transimpedance gain yields the input-referred noise current :

$$\overline{I^2_{n,in}} = 4kT(\gamma g_{m2} + \frac{1}{R_D}) \tag{3.16}$$

$$= \overline{I^2_{n,M2}} + \overline{I^2_{n,RD}} \tag{3.17}$$

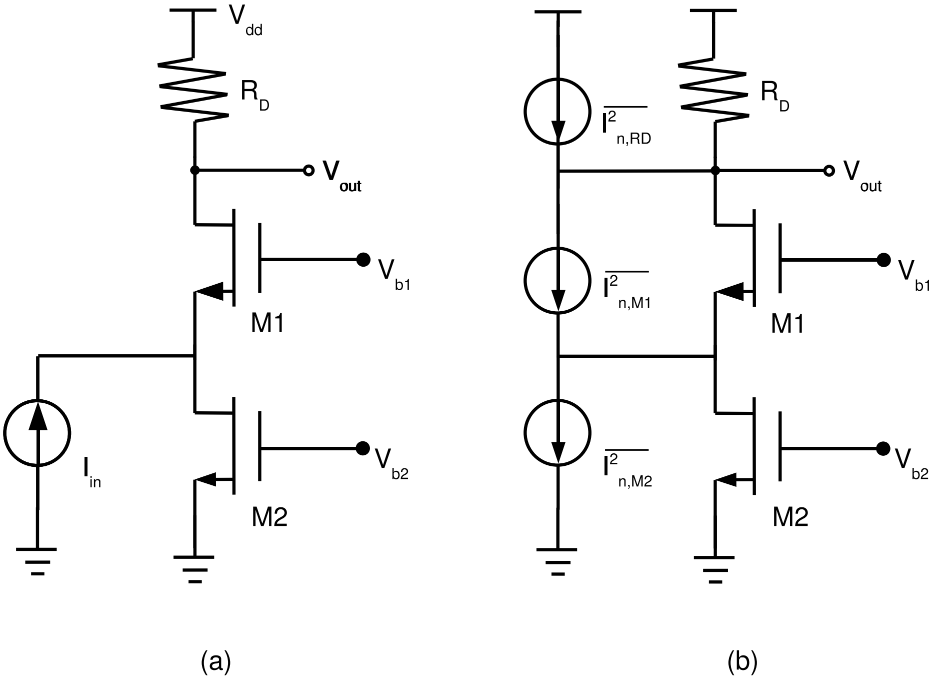


Figure 3.3: (a) Common-gate stage with bias current source (b) equivalent circuit with noise sources.

Equation (3.17) suggests that the noise currents of M2 and R_D are referred to the input with a unity factor. This is the principal drawback of common-gate TIAs. Also there are trade-offs between $\overline{I^2_{n,M2}}$ and $\overline{I^2_{n,RD}}$. This is because for a fixed biased current for the common gate stage, we need to maximize the V_{dsat} of the M2 to make g_{m2}, thus $\overline{I^2_{n,M2}}$ smaller, while this means R_D must be decreased since the headroom for R_D is shrunk by increase of V_{dsat}. Finally, the decrease of R_D introduces more noise current to the output.

3.2.2 Regulated Cascode TIA

The CG input configuration relaxes the effect of large input parasitic capacitance on the bandwidth of the front-end. However, generally g_m of MOS transistors are limited by the constraints of power consumption and the Common gate input stage cannot totally isolate the parasitic capacitance. Moreover, this small g_m deteriorates the noise performance of the amplifier. A regulated cascode (RGC) configuration mitigates these issues and is found to be used in some TIA designs [76, 77, 78]. The RGC input mechanism enhances the effective transconductance significantly. As a result, the input node of the amplifier can sit at virtual ground and higher bandwidths are feasible.

Fig. 3.4 shows the schematic diagram of the RGC circuit. The optically generated current is converted to be a voltage at the drain of M2. The M1/Rb stage operates as a local feedback and thus reduces the input impedance by the amount of its own voltage gain.

Transimpedance gain the transimpedance gain of a RGC stage in Fig. 3.4 is R_1.

Input impedance the input impedance of the RGC circuit can be approximated by

$$R_{in} \simeq \frac{1}{g_{m2}(1+g_{m1}R_b)} \tag{3.18}$$

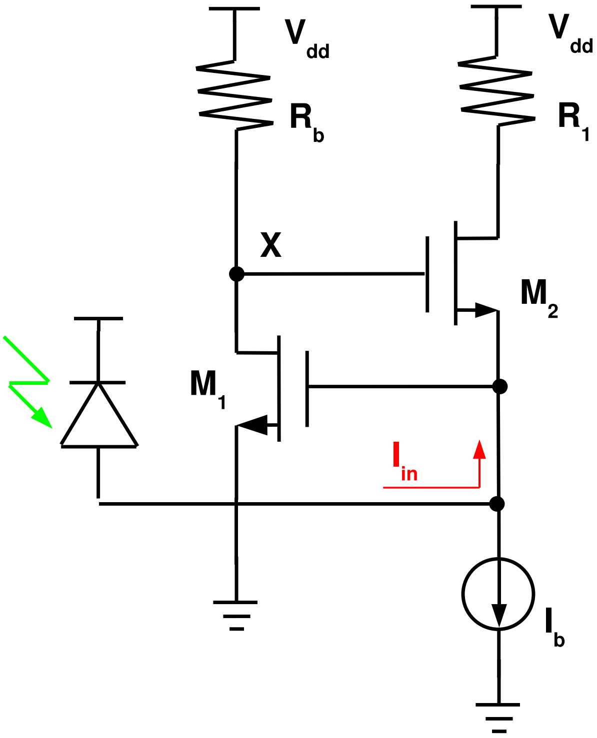


Figure 3.4: Regulated cascode TIA input stage

Clearly the input impedance is $(1+g_{m1}R_b)$ times smaller than the CG configuration. However, the pole in the local feedback stage inherently produces a zero in the transfer function of the system, causing a peaking in the frequency response at the frequency of $1/[2\pi R_b(C_{gs2}+C_{db1})]$. In order to avoid this peaking and stability concerns, either the resistance R_b or the gate width of M2 should be reduced. Reducing R_b decreases the input transconductance G_m almost linearly. In this case, in order to obtain the same $(1+g_{m1}R_b)$, the drain bias current of M1 needs to be increased, thus resulting in larger power consumption. Reducing the width of transistor M2 decreases G_m more slowly. However, it may lead to the increase of the channel thermal noise contribution from M2 due to smaller g_{m2}[2]. Generally in RGC TIAs a second stage voltage amplifier usually follows this first stage RGC, to boost gain.

---

[2] in first order approximation, the thermal noise of M1 does not contribute to the noise of the system, but it does when C_{gs2} and C_{gd1} are taken into account. Further analysis can be found in section 3.3.3.

Bandwidth and noise The bandwidth and noise performance of RGC stage will be analyzed in detail in section 3.3.3. Similar to CG topology the noise of bias resistor and gain resistor directly contribute to the total noise of the TIA. However, the enhanced input transconductance reduces the high-frequency noise contribution from transistor M2, and the trade-off between the noise and bandwidth is inevitable in some cases.

3.2.3 Shunt-Shunt Feedback TIA

Besides common-gate circuits, the most common TIA configuration is the "voltage-current" or "shunt-shunt" feedback topology, where a negative feedback network senses the voltage at the output and returns a proportional current to the input. This type of feedback is chosen because it lowers both the input impedance–thus increasing the input pole magnitude and allowing the amplifier to absorb the big photodiode capacitance– thereby yielding better drive capability.

A typical simplified topology of a shunt-shunt feedback transimpedance amplifier is shown in Fig. 3.5 (a), where R_F provides feedback around a voltage amplifier. It is assumed the core amplifier exhibits an open loop DC gain of A_0, and a single pole ω_0, therefore its transfer function is given by

$$A(s) = \frac{A_0}{1 + s/\omega_0} \tag{3.19}$$

Assume the open loop DC gain of the core amplifier, $A_0 \gg 1$, we have the transimpedance gain Z_T of the feedback TIA

$$\begin{aligned} \frac{V_{out}}{I_{in}} &= -\frac{A(s)R_F}{A(s) + 1 + R_F C_D s} \\ &= -\frac{A_0 R_F}{\frac{R_F C_D}{\omega_0} s^2 + (R_F C_D + \frac{1}{\omega_0})s + A_0 + 1} \end{aligned} \tag{3.20}$$

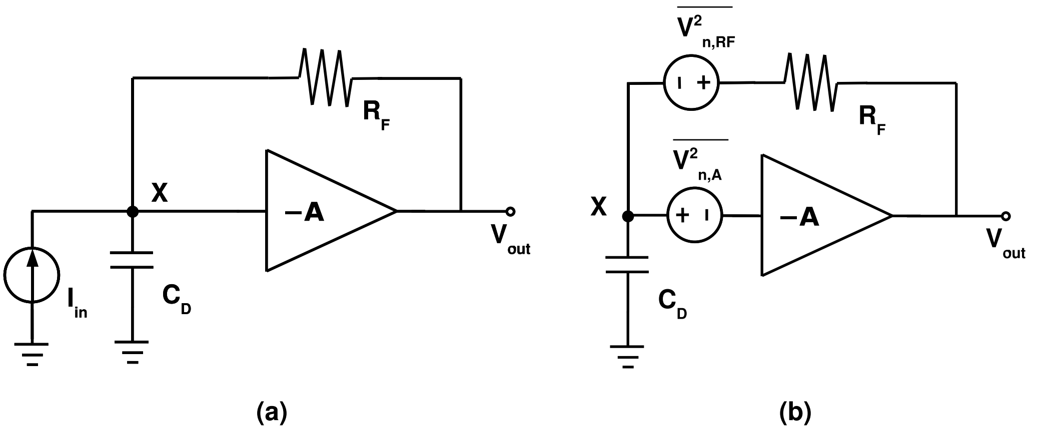


Figure 3.5: Shunt-shunt Feedback TIA (a) the simplified schematic with a feedback resistor (b) the equivalent circuit of (a) with noise sources added

Input impedance from the principle of feedback system [80], the input of the shunt-shunt feedback amplifier and the photodiode can be approximated as[3] :

$$Z_{in} = \frac{R_F}{A_0} \parallel \frac{1}{sC_D} \tag{3.21}$$

Transimpedance and bandwidth In order to quantitize the bandwidth of feedback TIA, we examine equation (3.20) for two cases. First, suppose the second pole of the closed-loop system is much higher in magnitude than the first one (named as ω_{p1} and ω_{p2} respectively). Noting that

$$(\frac{s}{\omega_{p1}}+1)(\frac{s}{\omega_{p2}}+1) = \frac{s^2}{\omega_{p1}\omega_{p2}} + (\frac{1}{\omega_{p1}} + \frac{1}{\omega_{p2}})s + 1 \tag{3.22}$$

and $\frac{1}{\omega_{p1}} + \frac{1}{\omega_{p2}} \approx \frac{1}{\omega_{p1}}$ if $\omega_{p2} \gg \omega_{p1}$, we have

$$\omega_{p1} \approx \frac{A_0+1}{R_F C_D + \frac{1}{\omega_0}} \approx \frac{A_0}{R_F C_D + \frac{1}{\omega_0}} \tag{3.23}$$

the second pole is obtained by recognizing that $(\omega_{p1}\omega_{p2})^{-1} = R_F C_D/[(A_0+1)\omega_0]$,

---

[3] The input impedance of the core amplifier and the load effect of the feedback resistor on it are ignored here.

$$\omega_{p2} \approx \omega_0 + \frac{1}{R_F C_D} \tag{3.24}$$

thus the second pole is equal to the sum of the open-loop poles of the circuit (so long as the assumption $\omega_{p2} \gg \omega_{p1}$remain valid). Here if we assume the core amplifier is an ideal one (which means $\omega_0 = \infty, A_0 \gg 1$), we have $\omega_{p1} \approx A_0/R_F C_D$, $\omega_{p2} \approx \infty$, which is as expected since the shunt-shunt feedback TIA has an input impedance of R_F/A_0.

With the assumption of $\omega_{p2} \gg \omega_{p1}$, it means

$$\omega_0 + \frac{1}{R_F C_D} \gg \frac{A_0 + 1}{R_F C_D + \frac{1}{\omega_0}} \tag{3.25}$$

and hence

$$(R_F C_D \omega_0 + 1)^2 \gg R_F C_D \omega_0 (A_0 + 1) \tag{3.26}$$

If $\omega_0 \gg (R_F C_D)^{-1}$, then $R_F C_D \omega_0 + 1 \approx R_F C_D \omega_0$, and

$$\omega_0 \gg \frac{A_0 + 1}{R_F C_D} \tag{3.27}$$

In other words, the open-loop pole of the amplifier must be much higher than the pole resulting from R_F and C_D. In practice, the above condition may not hold. In this case the transfer function can be treated as a typical second-order system. To ensure a well-behaved time response, the system is preferred to have a critically-damped behavior. In a basic control theory, if the denominator of a second-order transfer function is expressed as $s^2 + 2\zeta\omega_n + \omega_n^2$, then ζ, the "damping factor", must be equal to $\sqrt{2}/2$ for critical damping. If $\zeta < \sqrt{2}/2$, then the step response exhibits ringing, or even causes stability problem. If $\zeta > \sqrt{2}/2$, the response is "over damping", slowing the system speed. Rewriting equation (3.20), we have

$$\frac{V_{out}}{I_{in}} = -\frac{A_0\omega_0/C_D}{s^2 + \frac{R_F C_D + 1/\omega_0}{R_F C_D/\omega_0} s + \frac{(A_0+1)\omega_0}{R_F C_D}} \tag{3.28}$$

concluding that $\omega_n^2 = (A_0 + 1)\omega_0/(R_F C_D)$ and hence

$$\zeta = \frac{1}{2} \frac{R_F C_D \omega_0 + 1}{\sqrt{(1 + A_0) R_F C_D \omega_0}} \tag{3.29}$$

If $\zeta = \sqrt{2}/2$, then

$$(R_F C_D \omega_0 + 1)^2 - 2A_0 R_F C_D \omega_0 + 1 = 0 \tag{3.30}$$

and

$$\omega_0 = \frac{A_0 \pm \sqrt{A_0^2 - 1}}{R_F C_D} \tag{3.31}$$

Recognizing that only the sum gives a valid solution and noting that $A_0 \gg 1$, we obtain

$$\omega_0 \approx \frac{2A_0}{R_F C_D} \tag{3.32}$$

Thus, the -3-dB bandwidth of the core amplifier must be chosen equal to twice the closed-loop bandwidth of the ideal TIA ($A_0/(2\pi R_F C_D)$) to ensure a critically-damped response. Intuitively, we expect that as ω_0 exceeds this value, the feedback becomes overdamped, a trend indeed observed in equation (3.29). Conversely, if $\omega_0 < 2A_0/(R_F C_D)$, then the TIA suffers from the potential instability.

With above condition for critical damping, the other parameters of the circuit can be determined. In particular, we have

$$\omega_n = \sqrt{\frac{(A_0 + 1)\omega_0}{R_F C_D}} \tag{3.33}$$

$$\approx \frac{\sqrt{2} A_0}{R_F C_D} \tag{3.34}$$

More importantly, the -3 dB bandwidth of the TIA is obtained by setting the magnitude of (3.28) to $\sqrt{2}/2$ times its low-frequency value. Writing (3.28) in a more general form,

$$\frac{V_{out}}{I_{in}} = -\frac{R_0\omega_n^2}{s^2 + 2\zeta\omega_n + \omega_n^2} \tag{3.35}$$

where R_0 is the low frequency transimpedance gain of the shunt-shunt feedback TIA :

$$R_0 = \frac{A_0}{(A_0+1)}R_F \tag{3.36}$$

further we have

$$\left|\frac{R_0\omega_n^2}{-\omega_{-3dB}^2 + 2\zeta\omega_n\omega_{-3dB} + \omega_n^2}\right| = \frac{\sqrt{2}}{2}R_0 \tag{3.37}$$

That is

$$(\omega_n^2 - \omega_{-3dB}^2)^2 + 4\zeta^2\omega_n^2\omega_{-3dB}^2 = 2\omega_n^4 \tag{3.38}$$

which, along with $\zeta = \sqrt{2}/2$, yields

$$\omega_{-3dB} = \omega_n \tag{3.39}$$

Thus, the -3-dB bandwidth of the second-order TIA is equal to

$$f_{-3dB} = \frac{1}{2\pi}\frac{\sqrt{2}A_0}{R_F C_D} \tag{3.40}$$

Plotted in Fig. 3.6 is the -3-dB bandwidth of the shunt-shunt feedback TIA as a function of ω_0, say, the -3-dB bandwith of the core amplifier. The maximum TIA bandwidth is achieved at a critical-damped response, which is alternatively expressed as $\omega_0 \approx \sqrt{2}A_0/R_F C_D$.

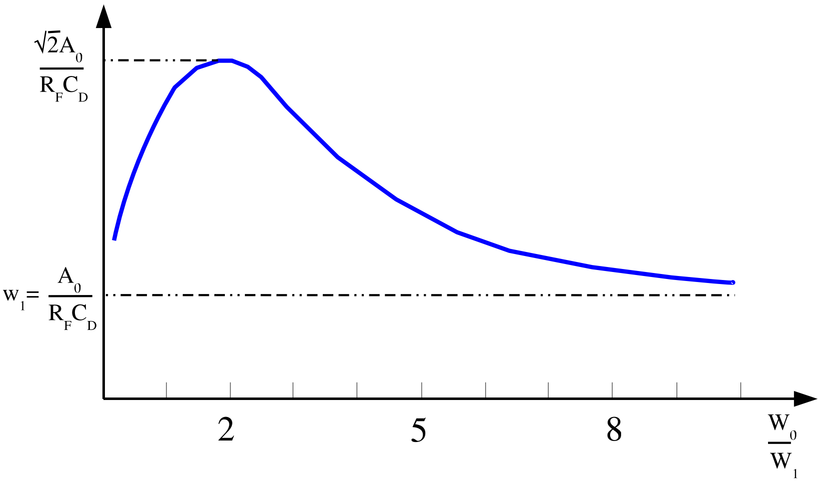


Figure 3.6: Shunt-shunt Feedback TIA bandwidth

Noise It is also instructive to examine the noise behavior of the circuit. Modelling the input referred noise of the amplifier by a voltage source $\overline{V_{n,A}^2}$ and neglecting its input noise current[4], we arrive at the equivalent circuit shown in Fig. 3.5(b). Since $V_x = V_{n,out}/(-A) + V_{n,A}$, we have

$$(V_{n,oout} + \frac{V_{n,out}}{A} - V_{n,A} - V_{n,RF})\frac{1}{R_F} \quad = \quad (-\frac{V_{n,out}}{A} + V_{n,A})C_D s \tag{3.41}$$

Note that the two noise sources ($\overline{V_{n,RF}^2}$ and $\overline{V_{n,A}^2}$) are uncorrelated, we must compute the total noise by superposition in *rms,* which can be solved as

$$\overline{V_{n,out}^2} = (\frac{1}{1+R_F C_D s/A})^2 \overline{V_{n,RF}^2} + (\frac{1+R_F C_D s}{1+R_F C_D s/A})^2 \overline{V_{n,A}^2} \tag{3.42}$$

where it is assumed $A \gg 1$. From equation (3.42), if $C_D = 0$, then $\overline{V_{n,out}^2} = \overline{V_{n,RF}^2} + \overline{V_{n,A}^2}$, yielding an input noise current (per unit bandwidth) of

---

[4] Note that only when the input impedance of the core amplifier is much higher than that of the source we can ignore the input equivalent noise current, here the input impedance of the core amplifier is assumed to be much larger than that of R_F and C_D in not very high frequencies.

$$\overline{I_{n,in}^2} = \frac{\overline{V_{n,in}^2} + \overline{V_{n,A}^2}}{R_F^2} \tag{3.43}$$

$$= \frac{4kT}{R_F} + \frac{\overline{V_{n,A}^2}}{R_F^2} \tag{3.44}$$

Thus, the noise of R_F is directly referred to the input, and the noise voltage of the amplifier is divided by R_F. While the first term of equation (3.44) may appear to play the same role as the term $4kT/R_D$ in equation (3.16) for the common-gate structure, the critical difference is that in the topology of shunt-shunt feedback, R_F needs not carry a bias current and its value does not limit the voltage headroom. Also, if R_F is large, the second term in equation (3.44) may be much smaller than the contribution from the bias current source in CG circuits. In the other extreme case, if C_D or s approach infinity, consequently, (3.42) approaches $A^2\overline{V_{n,A}^2}$. This is because C_D provides a low impedance from node X in Fig. 3.5 (b) to ground, allowing $V_{n,A}$ to be amplified by A.

3.2.4 Calculations on a typical feedback TIA

To quantize the noise $\overline{V_{n,A}^2}$ contributed by the core amplifier, a typical implementation of the feedback TIA with the core amplifier consisting of a common source NMOS transistor is analyzed. The circuit is shown in Fig. 3.7, where M1 is the amplification core with feedback resistor R_F to lower the input impedance, R_D is the load resistor. This configuration is also a broadband amplifier and it finds popular applications in UWB, optical communications and other broad band systems. Possible variations are cascode topology and that with a source follower added between the output and the feedback resistor.

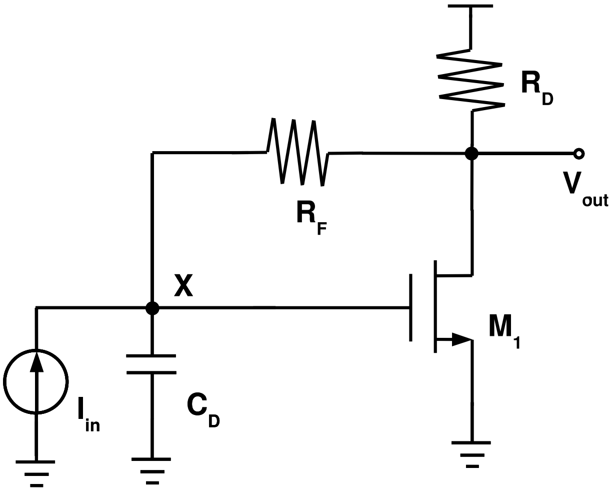
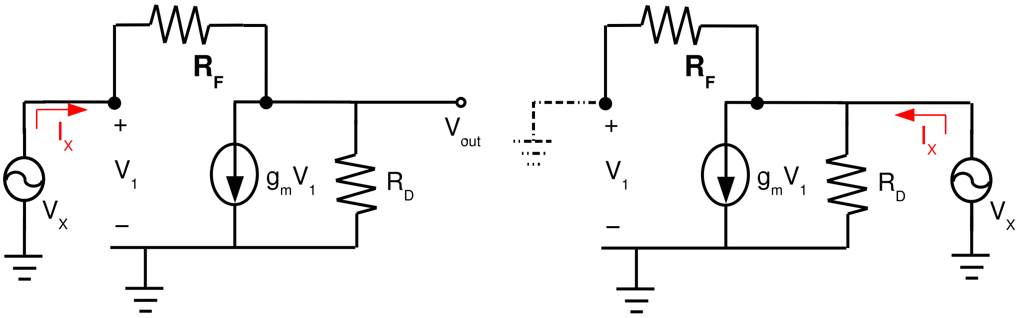


Figure 3.7: Implementation of feedback TIA



(a)

(b)

Figure 3.8: Equivalent small signal circuit of feedback TIA in Fig. 3.7 for calculating (a) input impedance (b) output impedance.

Input and output impedance To characterize the input impedance, output impedance of the TIA, the equivalent small signal circuit are shown in Fig. 3.8. In Fig. 3.8 (a), a voltage source V_X is applied to the input of the TIA for calculating input impedance. Based on Kirchhoff law (KCL), we have :

$$\begin{cases} V_{out} = (I_X - g_m V_X) R_D \\ V_X = V_{out} + R_F I_X \end{cases} \tag{3.45}$$

Thus, the input impedance R_{in} is

$$\begin{aligned} R_{in} &= \frac{V_X}{I_X} && (3.46) \\ &= \frac{R_F + R_D}{1 + g_m R_D} && (3.47) \end{aligned}$$

To calculate the output impedance, the small equivalent circuit is shown in Fig. 3.8 (b), where the input of the TIA is left as open circuit because the input to a TIA is typically modeled as a current source[5]. In this case the R_F does not effect the output impedance, we have

$$V_X = (I_X - \frac{V_X}{R_F})R_D \quad (3.48)$$

thus

$$\begin{aligned} R_{out} &= \frac{V_X}{I_X} \\ &= R_D \| (1/g_m) && (3.49) \end{aligned}$$

---

[5] It should be noted here that practically the photodiode is modeled as a parallel circuit of a current source and a parasitic capacitance. Here for simplicity we only calculate the input and output impedance of the TIA at low frequency, where the parasitic capacitance has been ignored. The high frequency analysis can be found in the section of TIA implementation.

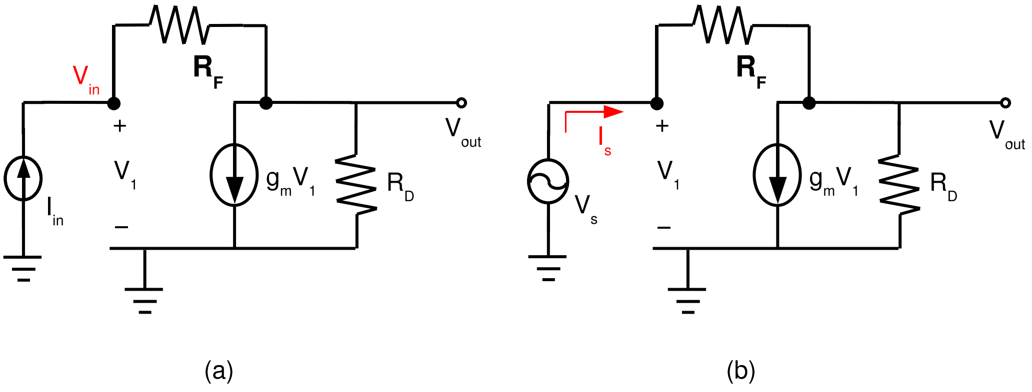


Figure 3.9: Equivalent small signal circuit of feedback amplifier in Fig. 3.7 for calculating (a) transimpedance gain as a TIA (b) voltage gain as a broadband amplifier

Transimpedance gain The transimpedance gain, R_T can be derived from the equivalent small signal circuit in Fig. 3.9(a). This is similar to the calculation of the input impedance, from Kirchhoff law, we have

$$\begin{cases} V_{in} = V_{out} + I_{in}R_F \\ V_{out} = R_D(I_{in} - g_m V_{in}) \end{cases} \tag{3.50}$$

then

$$R_T = \frac{V_{out}}{I_{in}} = \frac{R_D(g_m R_F - 1)}{1 + g_m R_D} \tag{3.51}$$

Voltage gain It is also instructive to compute the input, output impedance and the voltage gain of the circuit in Fig. 3.7 as a broadband amplifier instead of a TIA. First, it is easy to find that the input impedance R_{in} is the same as in equation (3.47). For output impedance calculation, the only difference between broadband amplifier and TIA is the termination of the input node in Fig. 3.8(b), which is typically terminated by 50Ω source for the case of broadband amplifier and open for the case of TIA. If R_F is much bigger than 50Ω, the output impedance for a broadband amplifier is given by

$$R_{out} = R_D || R_F \tag{3.52}$$

The transimpedance gain of the broadband amplifier is computed from the schematic in Fig. 3.9 (b). Note that $I_s = (V_s - V_{out})/R_F$, we have

$$\frac{V_s - V_{out}}{R_F} - g_m V_s = \frac{V_{out}}{R_D} \tag{3.53}$$

then the voltage gain of the broadband amplifier is

$$\begin{aligned} G &= \frac{V_{out}}{V_s} \\ &= \frac{R_D(g_m R_F - 1)}{R_F + R_D} \end{aligned} \tag{3.54}$$

From another aspect, we can also compute the voltage gain of the broadband amplifier as the product of the transimpedance gain of the TIA and its input admittance :

$$\begin{aligned} G &= R_T \cdot \frac{1}{R_{in}} \\ &= \frac{R_D(g_m R_F - 1)}{1 + g_m R_D} \cdot \frac{1 + g_m R_D}{R_F + R_D} \\ &= \frac{R_D(g_m R_F - 1)}{R_F + R_D} \end{aligned} \tag{3.55}$$

which is the same as (3.54).

Bandwidth The bandwidth of the circuit in Fig. 3.7 is easily obtained from equation (3.40) :

$$\begin{aligned} f_{-3dB} &= \frac{1}{2\pi}\frac{\sqrt{2}A_0}{R_F C_D} \\ &= \frac{1}{2\pi}\frac{\sqrt{2}\,g_m R_D}{R_F C_D} \end{aligned} \tag{3.56}$$

From equation (3.56), it seems the bandwidth is maximized by increasing R_D and reducing feedback resistor R_F. In fact, it is not true because equation (3.56) is achieved by the assumption in equation (3.32), $\omega_0 = 1/R_D C_L \approx 2A_0/R_F C_D$, which is

$$\frac{2g_m R_D}{R_F C_D} \approx \frac{1}{R_D C_L} \tag{3.57}$$

where C_L is the total capacitance at the output node, which includes the parasitic capacitance from the drain of M1 and R_D, and the input capacitance of the subsequent stage. Obviously big C_L limits the bandwidth of the circuit.

Another way to calculate the bandwidth is to get the transimpedance gain transfer function of the TIA with all the parasitic capacitance taken into account. For this case, the parasitics are the input photodiode parasitic capacitance C_D and output parasitic capacitance C_L. Adding C_D and C_L to the small equivalent circuit in Fig. 3.9(a), we have

$$\begin{cases} (V_{in} - V_{out}) \cdot 1/R_F = I_{in} - V_{in} \cdot sC_D \\ V_{out}/(R_D + sC_L) = I_{in} - V_{in} \cdot sC_D - g_m V_{in} \end{cases} \tag{3.58}$$

by solving the equations, we can have the transimpedance gain

$$Z_T = \frac{V_{out}}{I_{in}} = Z_0 \cdot \frac{1}{s^2/\omega_n^2 + s/Q_n\omega_n + 1} \tag{3.59}$$

where

$$Z_0 = \frac{R_D(g_m R_F - 1)}{1 + g_m R_D} \tag{3.60}$$

$$\omega_n^2 = \frac{1+g_mR_D}{R_FR_DC_LC_D} \tag{3.61}$$

$$(Q_n\omega_n)^{-1} = \frac{1+g_mR_D}{(R_DC_L+(R_F+R_D)C_D)}$$

It can be seen that from 3.60, the DC transimpedance gain is the same as that in equation (3.51). while when $\omega_0 = 1/R_DC_L \approx 2A_0/R_FC_D$, which is

$$\frac{2g_mR_D}{R_FC_D} \approx \frac{1}{R_DC_L} \tag{3.62}$$

we have ω_n^2 in equation (3.61) equal to $2g_m^2R_D^2/R_F^2C_D^2$, which gives the same bandwidth as in equation (3.56). This furthers proves the previous analysis results on the shunt shunt feedback TIA.

Noise Finally, we check the input-referred noise current of the simple resistor feedback TIA. In section 3.2.3, the noise of a resistive shunt shunt feedback TIA was presented where a generic amplifier was employed. To obtain the input referred noise of the circuit in Fig. 3.7, first we need to have the noise property of the amplifier core, which consists of the NMOS transistor M1 and the load resistor R_D. Neglecting channel-length modulation and body effect, the output noise voltage of the core amplifier is expressed as :

$$\overline{V_{n,out,core}^2} = 4kT(\gamma g_{m1} + \frac{1}{R_D})R_D^2 \tag{3.63}$$

Dividing (3.63) by the square of the voltage gain ($g_{m1}^2R_D^2$) yields the input referred noise of the core amplifier:

$$\overline{V_{n,in,core}^2} = 4kT\gamma\frac{1}{g_{m1}} + \frac{4kT}{g_{m1}^2R_D} \tag{3.64}$$

From equation (3.44), we have

$$\begin{aligned}\overline{I_{n,in}^2} &= \frac{4kT}{R_F} + \frac{\overline{V_{n,in,core}^2}}{R_F^2} \\ &= \frac{4kT}{R_F} + \frac{4kT}{R_F^2}\left(\frac{\gamma}{g_{m1}} + \frac{1}{g_{m1}^2 R_D}\right) \qquad (3.65)\end{aligned}$$

3.3 TIA in AMS 0.35μm CMOS

As discussed in last section, a variety of topologies are potential candidates for the implementation of high speed, low noise CMOS TIAs. Common-Gate (CG) configurations are not suitable for CMOS implementation because of the poor device characteristic of MOSFET, such as small g_m, which can not totally isolate the parasitic capacitance of the photodiode. Furthermore this small g_m deteriorates the noise and stability performance of the amplifiers. For this design, ReGulated cascode (RGC) is exploited as the input stage of the transimpedance amplifier since it mitigates the effect of large input parasitic capacitance from the bandwidth determination much better than the Common-Source (CS). As a result, the input impedance of the amplifier can sit at virtual ground and the high frequency noise contribution relating with the large input parasitic capacitance can be reduced.

The schematic of the TIA implemented in AMS 0.35μm is shown in Fig. 3.10. NMOS transistor M1, M2, together with resistor R_{b1}, R_1 form the RGC input stage, which is followed by a feedback amplifier consisting of NMOS transistor M3, M4, and resistor R_2 , R_f. The second voltage gain stage is required to boost gain boosting since the RGC input stage operates as a current buffer. A source follower (M5) is additionally employed as an output stage in order to improve the driving capability.

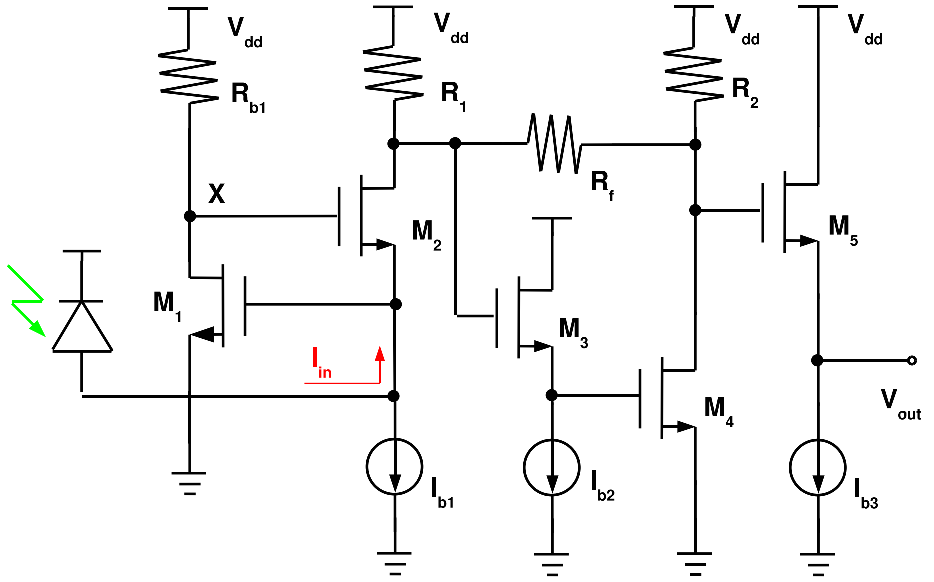


Figure 3.10: TIA schematic implemented in AMS C35

The photodiode converts the incoming optical power to a signal photocurrent, which is then amplified to be a voltage at the drain of M2. M1 and R_{b1} operate as a local feedback and thus reduces the input impedance by the amount of its voltage gain. Since the input impedance is very small due to the RGC input mechanism, it is not advantageous to apply the negative feedback to the input node. Rather, the feedback can be applied to a high impedance node that is the drain of M2. With this feedback application, the dominant pole of the amplifier moves to a higher frequency and thus wider bandwidth can be achieved [81, 82]. However, the circuit stability should be carefully examined because the dominant pole and the non-dominant pole of the amplifier move close to each other and may cause a peaking in the frequency response.

3.3.1 Small signal analysis

In order to find out the frequency response of the TIA, the small signal equivalent circuits for the input RGC stage and shunt-shunt feedback stage are first derived and analyzed individually since the outcome would be too complicated and not instructive if both stages are analyzed together.

Fig. 3.11 shows the simplified small signal equivalent circuit of the RGC input stage, where C_x denotes the total capacitance to ground at node X in Fig. 3.10, including C_{gb2}, C_{db1} and the parasitic capacitance of R_{b1}; C_y denotes the total parasitic capacitance to ground at input node, including C_{gs1}, C_{sb2}, and photodiode capacitance C_{PD}; C_{xy} denotes the sum of gate-source capacitance of M_2 and gate-drain capacitance of M_1 ($C_{xy} = C_{gs1} + C_{gd2}$). Here the Capacitance at the output node has been ignored, which would be taken care in analysis of subsequent stage – shunt-shunt feedback amplifier.

Based on Kirchhoff law, the summation of currents at node X, the input node and the output node should be equal to 0, thus we have

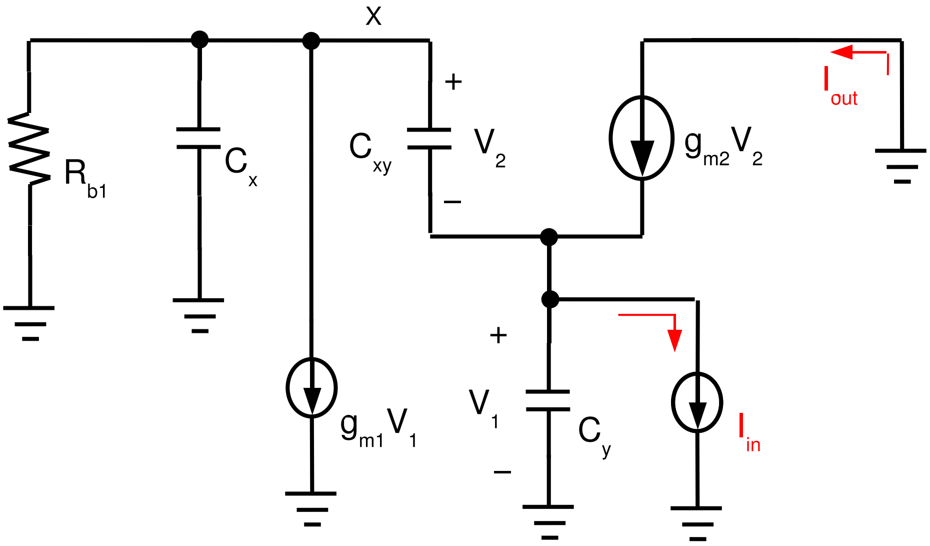


Figure 3.11: Small signal equivalent circuit of RGC input stage

$$\begin{cases} (V_1 + V_2) \cdot \frac{1}{R_{b1}} + (V_1 + V_2)sC_x + g_{m1}V_1 + V_2 sC_{xy} = 0 \\ V_2 sC_{xy} + g_{m2}V_2 = V_1 sC_y + I_{in} \\ I_{out} = g_{m2}V_2 \end{cases} \tag{3.66}$$

solving these equations to get

$$I_{out}(s) = I_{in} \cdot \frac{g_{m2}}{sC_{xy} + g_{m2} + sC_y \cdot [s(C_x + C_y) + 1/R_{b1}]/(sC_x + g_{m1} + 1/R_{b1})} \tag{3.67}$$

the transfer function of RGC stage is therefore

$$\begin{aligned} \frac{I_{out}}{I_{in}} &= \frac{g_{m2}(g_{m1} + \frac{1}{R_{b1}}) + g_{m2}C_x \cdot s}{(C_{xy}C_x + C_{xy}C_y + C_xC_y)s^2 + [(g_{m1} + \frac{1}{R_{b1}})C_{xy} + g_{m2}C_x + C_y\frac{1}{R_{b1}}]s + g_{m2}(g_{m1} + \frac{1}{R_{b1}})} \\ &= \frac{1 + s/z_1}{1 + s/(Q_1\omega_{n1}) + s^2/\omega_{n1}^2} \end{aligned} \tag{3.68}$$

where

$$z_1 = \frac{(g_{m1} + 1/R_{b1})}{C_x} \tag{3.69}$$

$$Q_1\omega_{n1} = [\frac{C_x}{(g_{m1} + 1/R_{b1})} + \frac{C_y \cdot R_{b1}}{g_{m2}(g_{m1} + 1/R_{b1})} + \frac{C_{xy}}{g_{m1}}]^{-1} \tag{3.70}$$

$$\omega_{n1}^2 = \frac{g_{m2}(g_{m1} + 1/R_{b1})}{C_{xy}C_x + C_{xy}C_y + C_xC_y} \tag{3.71}$$

It is clear from equation (3.68) that the zero of the RGC, z_1 comes from the pole of local feedback path, which is exactly the *RC* time constant of node X. As a current buffer, RGC has a unit current gain at low frequency. But as frequency increases, the current gain drops because of the parasitic capacitances at input node and X shunt signal to ground. Furthermore, the Miller capacitance C_{xy} lowers the local feedback gain $(1 + g_{m1}R_{b1})$ at high frequency, canceling the virtual ground effect of input node. The bandwidth of the RGC stage can be estimated as ω_{n1} based on the conclusion in section 3.2.3.

Since the input resistance of RGC stage is $1/g_{m2}(1 + g_{m1}R_{b1})$, we conclude from equation (3.71) that decrease the input resistance, minimizing the parasitic capacitance at input node and node X, result in wider bandwith. Since C_y is dominated by the large capacitance of the photodiode, it is usually much bigger than C_x, thus we have

$$\begin{aligned}\omega_{n1}^2 &= \frac{g_{m2}(g_{m1}+1/R_{b1})}{C_y(C_{xy}+C_x)} \\ &= \frac{1}{R_{b1}R_{in}} \cdot \frac{1}{C_y(C_{xy}+C_x)} \qquad (3.72)\end{aligned}$$

Fig. 3.12 shows the simplified small signal equivalent circuit of the shunt-shunt feedback amplifier stage of the TIA in Fig. 3.10, where the source follower transistor M3, M5 have been treated as unit gain buffer for simplicity, and they are not taken into account in Fig. 3.12. Similar to the analysis of RGC stage, the parasitic capacitances in the circuits are summed and include the device parasitic capacitances, i.e., C_2 denotes the total capacitance at the drain of M_2 in Fig. 3.10, including C_{db2}, C_{gs3} and the parasitic capacitance of R_1, R_f; C_4 denotes the total capacitance at the drain of M_4, including C_{db4}, C_{gs5} and the parasitic capacitance of R_2, R_f. With a current I_{in} from RGC stage, the feedback amplifier provides a voltage output V_{out} by a transimpedance gain of R_T. The summation of currents at input node and output node give

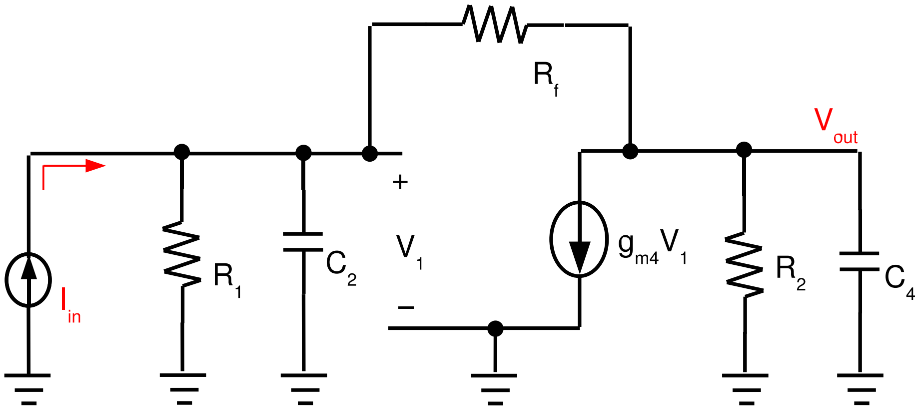


Figure 3.12: Small signal equivalent circuit of shunt-shunt feedback amplifier stage

$$\begin{cases} I_{in} - V_1(\frac{1}{R_1} + sC_2) - (V_1 - V_{out})\frac{1}{R_f} = 0 \\ (V_1 - V_{out})\frac{1}{R_f} = g_{m4}V_1 + V_{out}(\frac{1}{R_2} + sC_4) \end{cases} \tag{3.73}$$

solve these equations, we have

$$\begin{aligned} \frac{V_{out}}{I_{in}} &= \frac{1 - g_{m4}R_f}{(\frac{1}{R_2} + sC_4)(1 + \frac{R_f}{R_1} + sR_fC_2) + \frac{1}{R_1} + g_{m4} + sC_2} \\ &= \frac{1 - g_{m4}R_f}{R_fC_2C_4s^2 + (\frac{R_2+R_f}{R_2}C_2 + \frac{R_1+R_f}{R_1}C_4)s + g_{m4} + \frac{R_1+R_2+R_f}{R_1R_2}} \end{aligned} \tag{3.74}$$

To make it more intuitive, below can be obtained

$$\frac{V_{out}}{I_{in}} = \frac{1 - g_{m4}R_f}{g_{m4} + \frac{R_1+R_2+R_f}{R_1R_2}} \cdot \frac{1}{1 + s/(Q_2\omega_{n2}) + s^2/\omega_{n2}^2} \tag{3.75}$$

where

$$Q_2\omega_{n2} = (g_{m4} + \frac{R_1 + R_2 + R_f}{R_1R_2}) \cdot (\frac{R_2 + R_f}{R_2}C_2 + \frac{R_1 + R_f}{R_1}C_4)^{-1} \tag{3.76}$$

$$\omega_{n2}^2 = (\frac{g_{m4}}{R_f} + \frac{R_1 + R_2 + R_f}{R_1R_2R_f}) \cdot (C_2C_4)^{-1} \tag{3.77}$$

Combine equation (3.68) and (3.75), we have the transfer function of TIA in Fig. 3.10 :

$$\frac{V_{out}}{I_{in}} = R_T \cdot \frac{1 + s/z_1}{1 + s/(Q_1\omega_{n1}) + s^2/\omega_{n1}^2} \cdot \frac{1}{1 + s/(Q_2\omega_{n2}) + s^2/\omega_{n2}^2} \tag{3.78}$$

$$R_T = \frac{1 - g_{m4}R_f}{g_{m4} + (R_1 + R_2 + R_f)/(R_1R_2)} \tag{3.79}$$

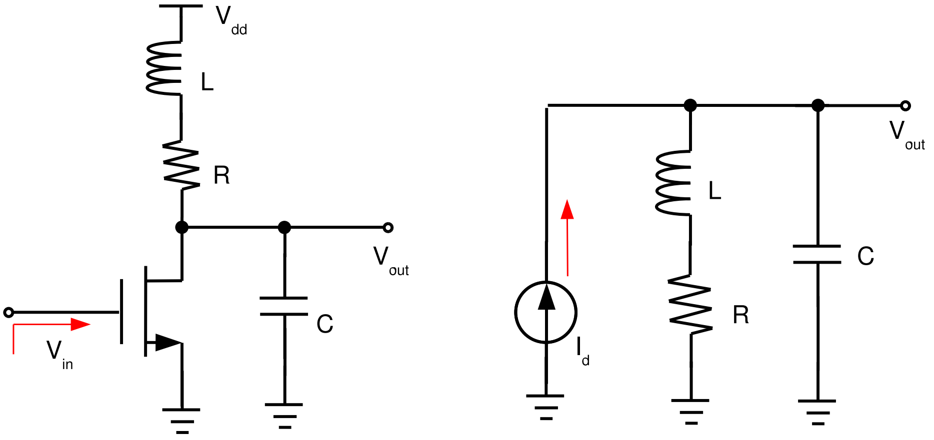
Transimpedance gain The TIA is basically a fourth-order system, a cascade of two second order low pass filters. The second term in (3.78) is contributed by the regulated cascode input stage, while the third term is determined by the shunt feedback amplifier in the second stage in Fig. 3.10. It has a transimpedance gain of R_T which is mainly decided by the transconductance of M_4 and the feedback resistor R_f in the second stage.

Bandwidth For a second order low pass filter which has a transfer function of $1/(1++s/(Q\omega_n)+s^2/\omega_n^2)$, it achieves a maximum bandwidth of ω_n when the filter presents two complex poles and $Q=\sqrt{2}/2$, corresponding to a butterworth response. By choosing appropriate g_{m1}, g_{m2} and R_{b1}, the impact of photodiode parasitic capacitance C_y on ω_{n1} can be minimized and the bandwidth of the TIA can be extended to ω_{n2} if $\omega_{n1} > \omega_{n2}$ [78]. In addition, the RGC input stage introduces another zero z_1, which can be placed at the roll-off region of the gain curve for bandwith extension. The expression of ω_{n2} in (3.77) reveals the trade offs between the transimpedance gain and bandwidth. For instance, under the constraint of power consumption, increasing the size of M_4, thus the transconductance, g_{m4}, improves the transimpedance gain of the amplifier, but it tends to deteriorate the bandwidth because of the increased C_4. Similarly, there are optimizations of the values of R_1, R_2. Considering the equation (3.79), high transimpedance gain R_T is achieved with large R_1, R_2 whilst bandwidth in (3.77) requires reasonally small R_1, R_2.

3.3.2 Inductive-series peaking

It has been shown in last section that the inherent parasitic capacitors of devices are the main cause of bandwidth limitation in transimpedance amplifiers. Several bandwidth enhancement methods have been proposed in the past[83, 84]. First-order shunt peaking has historically been used to introduce a resonant peaking at the output as the amplitude starts to roll off at high frequencies. It improves the bandwidth by adding an inductor in series with the output load to increase the effective load impedance as the capacitive reactance drops at high frequencies. Stripped to its essentials, a shunt-peaking amplifier is sketched in Fig. 3.13(a).

where capacitor C can be taken to represent all the capacitive loading on the output node, the resistor R is the effective load resistance at the output node and inductor L provides the bandwidth enhancement. We can model the amplifier in Fig. 3.13(a) for small signals as shown in Fig. 3.13(b), where the current source $i_d = g_m V_{in}$ (g_m is the transconductance of the transistor). It is then clear that the gain of the amplifier can be expressed as product of $g_m$and the impedance of the RLC network.



(a) (b)

Figure 3.13: Shunt-peaking (a) simplified shunt-peaking amplifier (b) model of shunt-peaking amplifier

The impedance of the RLC network can be written as

$$\begin{aligned} Z(s) &= (sL+R) \parallel \frac{1}{sC} \\ &= \frac{R(s(L/R)+1}{s^2LC+sRC+1} \end{aligned} \tag{3.80}$$

The magnitude of $Z(s)$ is

$$\begin{aligned}|Z(j\omega)| &= R\sqrt{\frac{(\omega L/R)^2+1}{(1-\omega^2 LC)^2+(\omega RC)^2}}\\ &= R\sqrt{\frac{(\omega\tau)^2+1}{(1-\omega^2\tau^2 m)^2+(\omega\tau m)^2}} \qquad (3.81)\end{aligned}$$

where τ is the defined as time constant of L/R, m is ratio of the RC time constant and τ (L/R). Different values of m leads to different desired behavior of the amplifier which is summarized in table 3.1. The maximum bandwidth is 1.85 times as large as the uncompensated bandwith with $m = 1.41$. But it is achieved at the cost of about 20% overshooting in frequency response.

Condition	$m = R^2C/L$	Norm. BW	Norm. peak response
Max BW	1.41	1.85	1.19
Max flat freq response	2.41	1.72	1
Best group delay	3.1	1.6	1
No shunt peaking	∞	1	1

Table 3.1: Shunt-peaking summary

Inductive shunt-peaking is widely used in wide-band amplifier design for its simplicity. But for multistage wide-band amplifier, inductive-series peaking finds more popular application for its efficiency. Considering a single-stage amplifier in Fig. 3.14, where the intrinsic output resistance and capacitance of the transistor, i.e. R_1 and C_1, are separated from those of the load, namely, $R_2$and C_2 by an inductor L. If there is no series inductor between the two stages (node A is shorted to node B), then the combination of capacitors C_1 and C_2 limits the bandwidth of the amplifier, i.e.,

$$\frac{V_{out}}{V_{in}} = \frac{-g_m(R_1 \parallel R_2)}{1+s(R_1 \parallel R_2)(C_1+C_2)}$$

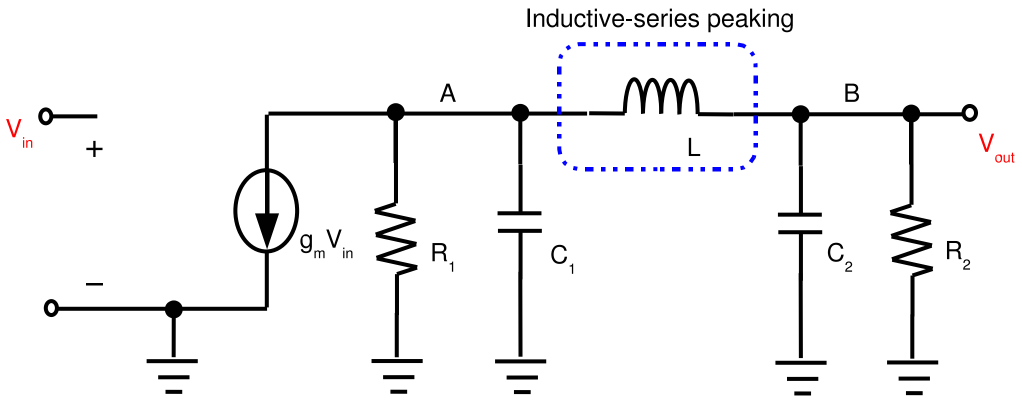


Figure 3.14: Small signal circuit of an amplifier with inductive-series peaking

A passive two-port network can be inserted between the transistor's intrinsic components (R_1 and C_1) and load (R_2 and C_2) to increase the bandwidth. This two-port passive network can be designed to maintain an constant impedance over a wide frequency range, as it separates and isolates C_1 and C_2. Therefore, C_1 is the only capacitor that affects GBW at the input port of the network. The passive two-port network is generally implemented by a simple inductor as shown in Fig. 3.14. The gain is then

$$\frac{V_{out}}{V_{in}} = \frac{-g_m R_p}{1 + s(C_t R_p + \frac{L}{R_1+R_2}) + s^2(\frac{R_p C_2 L}{R_1} + \frac{R_p C_1 L}{R_2}) + s^3 C_1 C_2 L R_p} \tag{3.82}$$

where $C_t = C_1 + C_2$, $R_p = R_1 \parallel R_2$. Equation (3.82) is a third-order system and so is computationally difficult to calculate the optimized component value L for the frequency response. Instead, graphical or numerical methods based on simulation can be used. Fig. 3.15 shows the simulated frequency responses in ADS of the single stage amplifier in Fig. 3.14. Here the input resistance of the second stage, R_2 is assumed to be infinite according to the implementation of the TIA in AMS C35. The resistance R_1, capacitances $C_1$and C_2 are normalized to be 1Ω, and $0.5\,pf$ to achieve 0-dB gain at low frequency and a 1-rad/s 3-dB bandwidth when $L = 0$. Values of normalized 3-dB bandwidth due to different L are summarized in table 3.2, where the bandwidth optimization assumes no gain peaking constraints.

L value (H)	0	0.2	0.4	0.6	0.8	1.0
3-dB bandwidth (rad/s)	1	1.2	3.4	2.8	2.4	2.17

Table 3.2: Normalized 3-dB bandwidth with different inductance L for amplifier in Fig. 3.14.

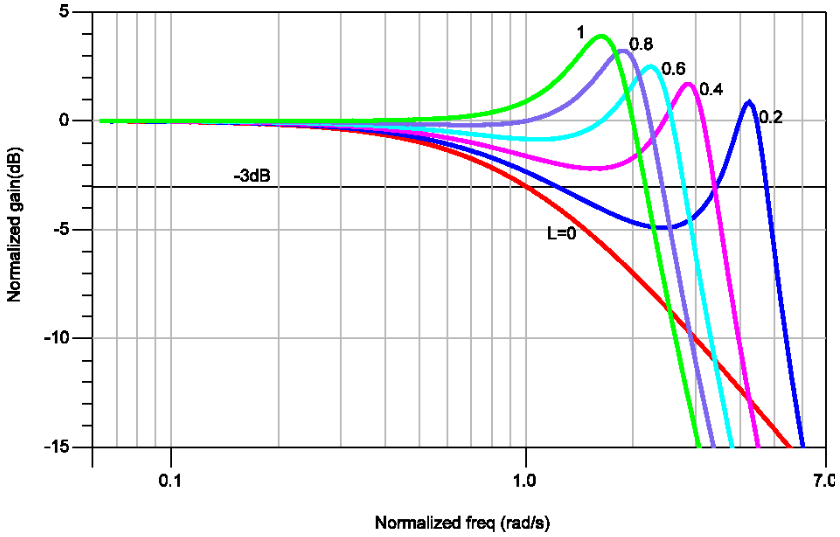


Figure 3.15: Simulated gain of amplifier in Fig. 3.14 with different inductor values ($R_1 = 1\Omega$, $R_2 = \infty, C_1 = C_2 = 0.5F$)

3.3.3 Noise analysis

It is preferred to do the noise analysis of the TIA step by step. First, assume the input equivalent current of TIA's second stage, the shunt feedback amplifier, is $\overline{i_{eq,fb}}$. Hence the input referred noise current of TIA can be computed based on the small signal equivalent circuit of RGC input stage, which is shown in Fig. 3.16. It takes the channel thermal noise sources of active devices and the thermal noise sources the resistors into account, whereas ignores the gate current noise of the CMOS transistors.

Superposition method is used to calculate the input equivalent noise current of TIA ($\overline{i_{eq}}$), i.e., the generated output noise currents $\overline{i_o}$ from different noise sources are mean squared at the output node and then referred back to input port by dividing by the transimpedance gain.

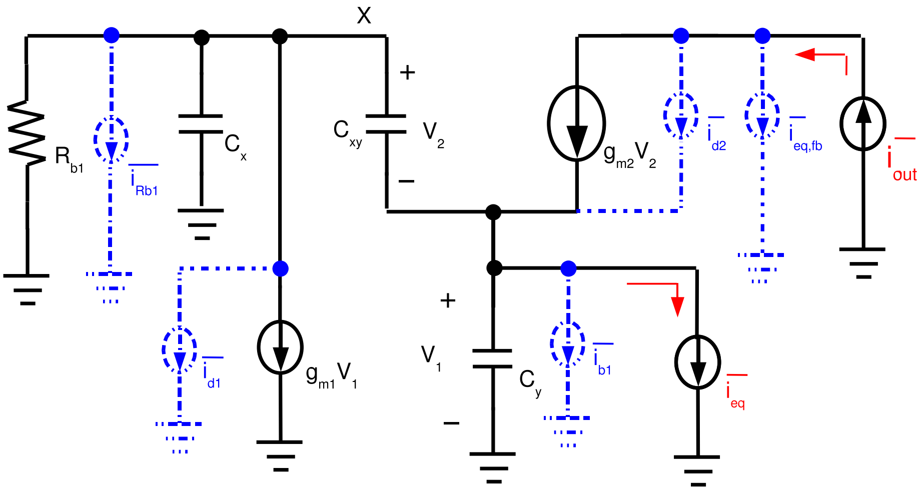


Figure 3.16: Noise equivalent circuit of the TIA (the input equivalent noise current of the shunt feedback amplifier is referred as $\overline{i_{eq,fb}}$)

First, the output noise current $\overline{i_{o1}}$ due to the thermal noise of current source $\overline{i_{b1}}$ can be directly obtained from equation (3.68) [6]:

$$\overline{i_{o1}} = \frac{[g_{m2}(g_{m1}+\frac{1}{R_{b1}})+g_{m2}C_x\cdot s]\cdot\overline{i_{b1}}}{(C_{xy}C_x+C_{xy}C_y+C_xC_y)s^2+[(g_{m1}+\frac{1}{R_{b1}})C_{xy}+g_{m2}C_x+C_y\frac{1}{R_{b1}}]s+g_{m2}(g_{m1}+\frac{1}{R_{b1}})}$$

Recall that $C_x = C_{gds}+C_{db1}$, which is much smaller compared with C_{xy}, C_y. Furthermore, if the circuit operation frequency ω is far below the transistor's cut off frequency f_T ($g_m/(C_{gs}+C_{gd})$), we have

---

[6] I_{b1} is a practically implemented by a NMOS transistor M_{b1}.

$$\begin{aligned}\overline{i_{o1}} &= \frac{g_{m2}(g_{m1}+\frac{1}{R_{b1}})\cdot\overline{i_{b1}}}{C_{xy}C_{y}s^{2}+[(g_{m1}+\frac{1}{R_{b1}})sC_{xy}+\frac{1}{R_{b1}}sC_{y}+g_{m2}(g_{m1}+\frac{1}{R_{b1}})} \\ &= \frac{g_{m2}(g_{m1}+\frac{1}{R_{b1}})\cdot\overline{i_{b1}}}{(g_{m1}+\frac{1}{R_{b1}})(g_{m2}+sC_{xy})+(\frac{1}{R_{b1}}+sC_{xy})sC_{y}} \qquad (3.83)\end{aligned}$$

To calculate output noise current $\overline{i_{o2}}$ due to channel thermal noise of M2 , $\overline{i_{d2}}$, we sum the currents at node X, input port and output port in Fig. 3.16 and obtain

$$\begin{cases} (V_1+V_2)\cdot\frac{1}{R_{b1}}(V_1+V_2)\cdot sC_x+g_{m1}V_1+V_2sC_{xy}=0 \\ V_2sC_{xy}+g_{m2}V_2+\overline{i_{d2}}=V_1sC_y \\ \overline{i_{o2}}=g_{m2}V_2+\overline{i_{d2}} \end{cases} \qquad (3.84)$$

solving the above equation (3.84) and ignoring the effect of C_x as mentioned before, we have

$$\overline{i_{o2}}=\frac{[(g_{m1}+\frac{1}{R_{b1}})sC_{xy}+(\frac{1}{R_{b1}}+sC_{xy})sC_{y}]\cdot\overline{i_{d2}}}{(g_{m1}+\frac{1}{R_{b1}})(g_{m2}+sC_{xy})+(\frac{1}{R_{b1}}+sC_{xy})sC_{y}} \qquad (3.85)$$

Similarly, the output noise current $\overline{i_{o3}}$ due to channel thermal noise of M1 and resistor R_1($\overline{i_B}=\overline{i_{d1}}+\overline{i_{Rb1}}$) is given by

$$\overline{i_{o3}}=\frac{g_{m2}sC_{y}\cdot(\overline{i_{d1}}+\overline{i_{Rb1}})}{(g_{m1}+\frac{1}{R_{b1}})(g_{m2}+sC_{xy})+(\frac{1}{R_{b1}}+sC_{xy})sC_{y}} \qquad (3.86)$$

The output noise current $\overline{i_{o4}}$ due to the equivalent noise current from next stage, $\overline{i_{eq,fb}}$, is

$$\overline{i_{o4}}=\overline{i_{eq,fb}} \qquad (3.87)$$

Finally, refer the total noise current at the output node $\overline{i_{o,tot}}$ to the input port of

TIA based on equation (3.68)[7]:

$$\begin{aligned}\overline{i_{o,tot}^2} &= \overline{i_{o1}^2}+\overline{i_{o2}^2}+\overline{i_{o3}^2}+\overline{i_{o4}^2}\\ &= \left|\frac{g_{m2}(g_{m1}+\frac{1}{R_{b1}})}{(g_{m1}+\frac{1}{R_{b1}})(g_{m2}+sC_{xy})+(\frac{1}{R_{b1}}+sC_{xy})sC_y}\right|^2\cdot\overline{i_{eq}^2} \qquad (3.88)\end{aligned}$$

where $\overline{i_{o1}}$, $\overline{i_{o2}}$, $\overline{i_{o3}}$, $\overline{i_{o4}}$ refer to (3.83), (3.85), (3.86), (3.87). Further we have

$$\begin{aligned}\overline{i_{eq}^2} &= \overline{i_{b1}^2}+\left|\frac{(g_{m1}+\frac{1}{R_{b1}})sC_{xy}}{g_{m2}(g_{m1}+\frac{1}{R_{b1}})}\right|^2\cdot\overline{i_{d2}^2}+\left|\frac{sC_y}{g_{m1}+\frac{1}{R_{b1}}}\right|^2\cdot(\overline{i_{d1}^2}+\overline{i_{Rb1}^2})\\ &\quad+\left|\frac{(g_{m1}+\frac{1}{R_{b1}})(g_{m2}+sC_{xy})}{(g_{m1}+\frac{1}{R_{b1}})g_{m2}}\right|^2\cdot\overline{i_{eq,fb}^2} \qquad (3.89)\end{aligned}$$

Here it is assumed in (3.89) that, g_{m1}, g_{m2} are much bigger than $1/R_{b1}$, sC_{xy}, sC_y. This simplicity is reasonable when TIA operation frequency is far below f_T, and more importantly, it makes the final noise expression more intuitive. It is also easy to find out

$$\overline{i_{eq,fb}^2}=\left|\frac{sC_2}{g_{m4}}\right|^2\cdot(\overline{i_{d4}^2}+\overline{i_{R2}^2})+\overline{i_{R1}^2}+\overline{i_{Rf}^2} \qquad (3.90)$$

where $\overline{i_{d4}^2}$, $\overline{i_{R2}^2}$, $\overline{i_{R1}^2}$, $\overline{i_{Rf}^2}$ are noise currents of transistor M4, resistor R_2, R_1, R_f; C_2 is the total parasitic capacitance at the drain of M2 in Fig. 3.10. The noise contribution of two source followers (M3, M5) are omitted in (3.90). Replace $\overline{i_{eq,fb}^2}$ in equation (3.89) by (3.90); transistor's channel thermal noise current by $4kT\gamma g_m$; resistor's noise current by $4kT/R$, we have

---

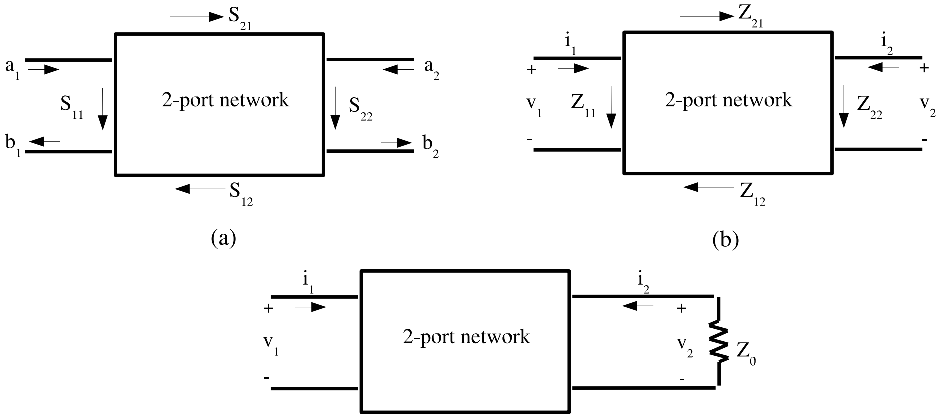
[7] Note that all the noise source are summed in *rms* since they are uncorrelated.

$$\begin{aligned}\overline{i_{eq}^2} &= 4kT\gamma g_{mb1} + 4kT\gamma g_{m2}\frac{\omega^2 C_{xy}^2}{g_{m2}^2} + 4kT(\gamma g_{m1} + \frac{1}{R_{b1}})\frac{\omega^2 C_y^2}{(g_{m1} + \frac{1}{R_{b1}})^2} \\ &\quad + (1 + \frac{\omega^2 C_{xy}^2}{g_{m2}^2})[4kT(\gamma g_{m4} + \frac{1}{R_2})\frac{\omega^2 C_2^2}{g_{m4}^2} + 4kT(\frac{1}{R_1} + \frac{1}{R_f})] \qquad (3.91)\end{aligned}$$

It can be seen from (3.91) that the noise currents from the input current source I_{b1}, resistor R_1, R_f directly refer to the input without any gain suppression. The noise from the the local feedback transistor M1 and resistor R_{b1} in RGC stage are attenuated by the transconductance of g_{m1}, which reminds us to choose g_{m1}as large as possible. It is also noted that channel thermal noise from M2 is referred to the input by the Miller capacitor C_{xy}, but suppressed by g_{m2}. Sizing M2 then should be very careful since it effects C_{xy}, g_{m2}, and the zero location in RGC transfer function (3.68), which further influences the TIA's noise, gain and bandwidth.

3.3.4 Design consideration

Transimpedance gain calculation based on S parameters TIA sub-circuits were fabricated separately for performance evaluation (tape out code: Atto1b). It is difficult to measure the transimpedance gain directly, S parameters of the circuits are generally first derived for the convenience, from which transimpedance gain is then calculated. But how to relate the transimpedance gain Z_T to S parameters? To do that, we first convert the S parameters into Z parameters for a two-port network. The diagrams for the definitions of S parameters and Z parameters are drawn in Fig. 3.17 (a) and Fig. 3.17 (b) respectively.



(a) (b)

(c)

Figure 3.17: 2-port network parameters (a) S parameters (b) Z parameters (c) Z parameters for calculating transimpedance gain Z_T

For S parameters:

$$\begin{aligned} b_1 &= S_{11} \cdot a_1 + S_{12} \cdot a_2 && (3.92) \\ b_2 &= S_{21} \cdot a_1 + S_{22} \cdot a_2 && (3.93) \end{aligned}$$

where a_j and b_j ($j = 1, 2$) are defined as[85] :

$$\begin{aligned} a_j &= \left[\frac{Z_{0j} + Z_{0j}^*}{2}\right]^{1/2} \cdot i_{ji} && (3.94) \\ b_j &= \left[\frac{Z_{0j} + Z_{0j}^*}{2}\right]^{1/2} \cdot i_{jr} && (3.95) \end{aligned}$$

where $*$ indicates complex conjugate and Z_{0j} is the normalized impedance for jth port. For two-port networks, Z_{01} and Z_{02} are the source and load impedance of the system in which the S parameters of the two-port are measured or calculated. i_{ji} and i_{jr} are the incident and reflected currents for the jth port. Knowing that,

$$i_j = i_{ji} - i_{jr} \tag{3.96}$$

we can solve (3.94), (3.95) for $i_{ji}$and i_{jr} and substitute them into (3.96) to get

$$i_j = \left[\frac{2}{Z_{0j} + Z_{0j}^*}\right]^{1/2} \cdot (a_j - b_j) \tag{3.97}$$

On the other hand,

$$v_j = v_{ji} + v_{jr} \tag{3.98}$$

where v_{ji} and v_{jr} are the incident and reflected voltage at the jth port, we can substitute the expressions for i_{ji} and i_{jr} along with $v_{ji} = i_{ji} \cdot Z_{0j}^*$ and $v_{jr} = i_{jr} \cdot Z_{0j}$, we have

$$v_j = \left[\frac{2}{Z_{0j} + Z_{0j}^*}\right]^{1/2} \cdot (a_j \cdot Z_{0j}^* + b_j \cdot Z_{0j}) \tag{3.99}$$

Solving (3.97) and (3.99) for a_j and b_j gives

$$\begin{aligned} a_j &= \frac{v_j + Z_{0j} i_j}{\left[2(Z_{0j} + Z_{0j}^*)\right]^{1/2}} && (3.100) \\ b_j &= \frac{v_j - Z_{0j} i_j}{\left[2(Z_{0j} + Z_{0j}^*)\right]^{1/2}} && (3.101) \end{aligned}$$

Substitute a_j and b_j in equation (3.92) and (3.93) we have the v_j and i_j expressed in S parameters. While from Z parameters in Fig. 3.17 (b), we have

$$\begin{aligned} v_1 &= Z_{11} \cdot i_1 + Z_{12} \cdot i_2 && (3.102) \\ v_2 &= Z_{21} \cdot i_1 + Z_{22} \cdot i_2 && (3.103) \end{aligned}$$

Finally we finish the conversion from S parameters to Z parameters, we have

$$Z_{21} = Z_0 \cdot \frac{2S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \tag{3.104}$$

$$Z_{22} = Z_0 \cdot \frac{(1-S_{11})(1-S_{22})+S_{12}S_{21}}{(1-S_{11})(1-S_{22})-S_{12}S_{21}} \tag{3.105}$$

To calculate the transimpedance gain Z_T, a load with impedance of Z_0 is added to the output of the two-port network in Fig. 3.17 (c), where we have

$$i_2 = -\frac{v_2}{Z_0} \tag{3.106}$$

replace i_2 in (3.103) and rearrange

$$\frac{v_2}{i_1} = \frac{Z_{21}}{1+Z_{22}/Z_0} \tag{3.107}$$

Substitute Z_{21} and Z_{22} with S parameters in equations (3.104), (3.105) to get

$$Z_T = Z_0 \cdot \frac{S_{21}}{1-S_{11}} \tag{3.108}$$

Note that the S parameters and Z parameters are not measured by open source and load. For S parameters, they are measured with source impedance and load impedance of Z_0. While for Z parameters, it is most convenient to open circuit the ports in succession to determine the various Z parameters experimentally, because various terms then become zero. For instance, determination of Z_{11} is easiest when the output port is open circuited because i_2 in equation (3.102) is zero under that condition [83]. Similarly to the determination of Z_{11}, open-circuiting the output port, driving the input port with a current source, and measuring the voltage of v_2 allows determination of Z_{21}. Note Z_{21} is not equal to transimpedance gain Z_T, which is defined as V_{out}/I_{in} with an appropriate load. Since we are using the S parameters to express the transimpedance gain Z_T, a load impedance of Z_0 is connected to the network as shown in Fig. 3.17 (c).

TIA Circuit implementation The final schematic of the TIA is shown in Fig. 3.18. Current sources I_{b1}, I_{b2} and I_{b3} are all implemented by single NMOS transistor. An external voltage bias controls the input biasing current, which is then replicated to the three current sources. M1-5 are implemented by NMOS transistor in AMS C35 with fixed length of 0.35μm and finger width of 5μm, while their sizes are determined by different finger numbers. No PMOS transistors are used for TIA because of their relative large parasitic capacitance than the poly resistors.

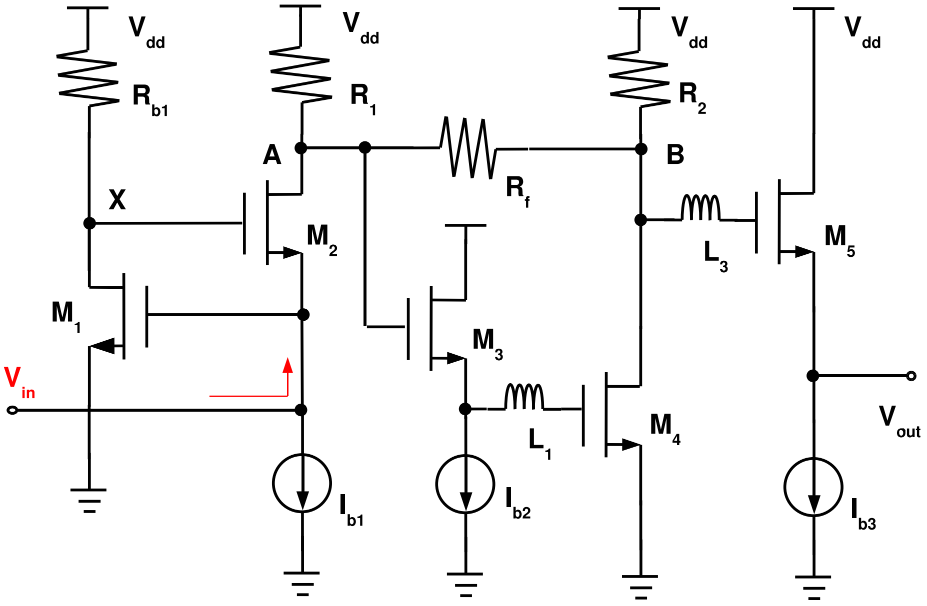


Figure 3.18: TIA implemented in AMS C35 with inductor-series peaking

Two kind of poly resistors are available in AMS C35 process, poly resistor (rpoly2) and high resistive poly module (rpolyH), both are silicide-blocked and implemented by poly 2. Rpoly2 has a typical sheet resistance of $50\Omega/\Box$, while rpolyH has a typical value of $1.2K\Omega/\Box$. For small resistance values of a few hundred Ω, like R_1, R_2 and R_f, rpoly2 is preferred for higher precision. But another design constraint that has to be taken into account is the current density of poly2, is only 0.3 $mA/\mu m$. To accommodate a current of $2\sim3$ mA, a 500Ω resistor must have at least a width of 10μm, and a length about 100μm. This introduces an intolerable parasitic capacitance into signal path of the circuit. Alternatively, rpolyH can be used to implement a 500Ω resistor. With the same width of $10\mu m$, the resistor has only a length of about $4.2\mu m$, reducing the parasitic capacitance magnificently. In AMS C35 design rules, it is not allowed to make the length of rpolyH resistor shorter than the width for precision consideration. In this design, this warning was ignored since trade off has to be made between the large parasitic capacitance of poly resistors and the resistor value variations (absolute resolution). From another point of view, it is possible to replace R_{b1} and R_1 with PMOS transistors as the load of the amplifier stage. However simulation shows that the PMOS transistor introduces higher parasitic capacitance compared with rpolyH resistor. For speed consideration, rpolyH resistors are used instead.

Two inductors are inserted between the stages of the TIA as shown in the Fig. 3.18. As mentioned in section 3.3.2, this inductive-series peaking technique effectively improves the bandwidth of the multistage amplifier. L_1 serves as the passive network between the parasitic capacitance of M3, Mb2 and the input parasitic capacitance of M4; L_3 isolates the parasitic capacitance at the drain of M4 from the input parasitic capacitance of source follower M5. Fig. 3.19 shows the simulation results with inductive-series peaking and without inductive-series peaking. A bandwidth enhancement ratio (the ratio of the bandwidth of the TIA with inductive-series peaking to the bandwidth of the TIA without) of 1.9 is observed.

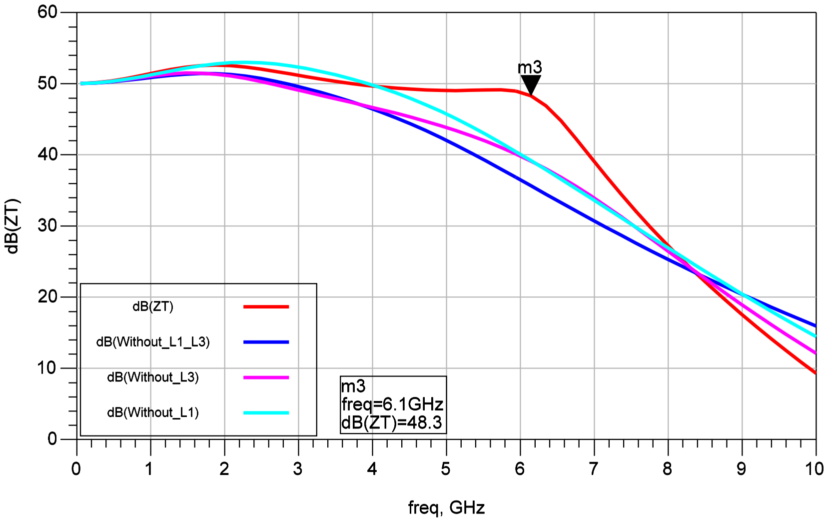


Figure 3.19: Simulation results of TIA with inductor-series peaking and without inductor-series peaking

L_1 and L_3 are implemented by the top metal of AMS C35 process, which has the largest thickness of 900 nm among the 4 metal layers. All the inductors provided by AMS C35 RF library are square shaped, with a metal width of 15 μm or 8 μm. The inductors have inductance ranging from 1.4 nH to 9.0 nH, Q factor from 2.5 to 6.2, respectively. In practical design, L_1 and L_3 are chosen to be 3.7 nH and 5.0 nH for optimizing the bandwidth and gain flatness.

There are many factors that need to be taken into consideration when choosing the size of the NMOS transistors (M1-M5) (schematic shown in Fig. 3.18). M1 is expected to be as large as possible in order to minimized the noise from M1 and R_{b1}, but if the size of M1 is too large, its parasitic capacitance adversely affects the bandwidth performance due to the parasitic capacitance at node X. Equation (3.91) indicates by increasing the transconductance of M2, g_{m2}, to reduce the noise contribution from M2 and the components of shunt feedback stage. On the other hand, simulation reveals that a large value of g_{m2} makes it difficult to match the input impedance to 50Ω at low frequency. For transistor M3, it was chosen to provide an appropriate DC operating point for M4 while not consuming too much current. M4 is critical to the transimpedance gain of the TIA, based on equation (3.79). g_{m4} was chosen to be large to increase transimpedance gain, whilst keeping its input gate capacitance C_{gs4} as small as possible to mitigate its impact on bandwidth. M5 is chosen to give a 50Ω output impedance over a wide frequency range from DC to -3dB bandwidth with a limited current consumption.

Table 3.3 gives the optimized component values of the TIA. All the transistors have a minimum length of $0.35\mu m$. The width of resistors are chosen not only to meet current density requirement, but also to achieve good matching. || in table 3.3 means the resistors are in parallel, i.e, R_1 has a design parameter of 6/5 || 3 (μm), which denotes it comprises of three resistor in parallel with length of $6\mu m$ and width of $5\mu m$.

component	design parameter	design value
M1	20×5/0.35 (μm)	
M2	8×5/0.35 (μm)	
M3	24×5/0.35 (μm)	
M4	24×5/0.35 (μm)	
M5	40×5/0.35 (μm)	
R_{b1}	7/5 || 3 (μm)	1750Ω/3
R_1	6/5 || 3 (μm)	1500Ω/3
R_2	5/5 || 3 (μm)	1250Ω/3
R_f	10/5 || 2 (μm)	2500Ω/3
L_1	3.7 (nH)	
L_2	5.0 (nH)	

Table 3.3: Optimized component values of the TIA

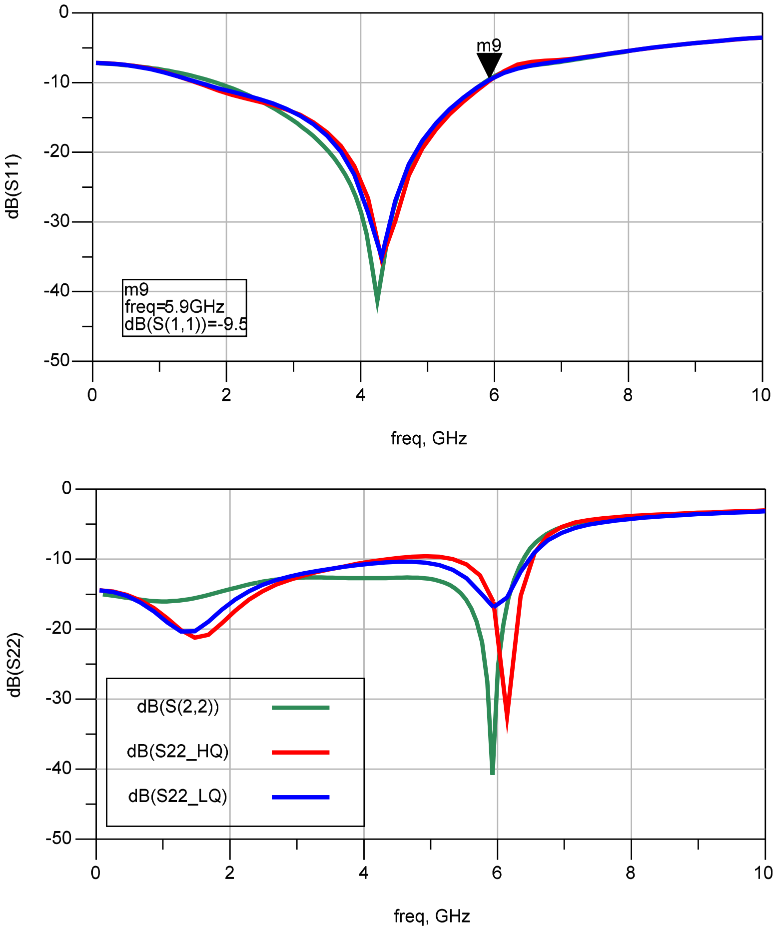


Figure 3.20: Simulation results of S_{11} and S_{22} of the TIA in AMS C35 with different inductor models

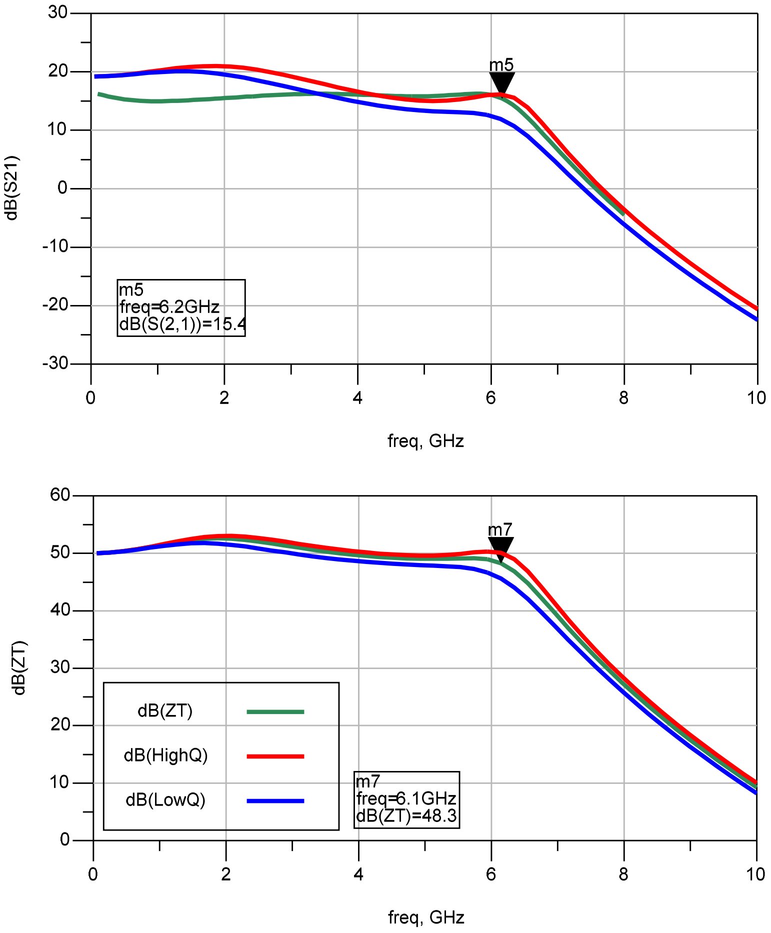


Figure 3.21: Simulation results of S_{21} and Z_T of the TIA in AMS C35 with different inductor models

Fig. 3.20 and 3.21 show simulation results of S parameters and the transimpedance gain of the TIA under different inductor models (low Q, typical means and high Q), where transimpedance gain Z_T is calculated by the method described in section 3.3.4. Fig. 3.22 and Fig. 3.23 show simulation results of S parameters and the transimpedance gain of the TIA under different resistor models (typical means, worst speed and worst power). Under all these conditions S_{12} of the amplifier is below -40dB.

Fig. 3.24 shows the simulation result of the input referred noise current of the TIA, a typical value of around 22 $pA/\sqrt{Hz}$ is achieved at the frequency of 5GHz. When frequency exceeds 5GHz, the input-referred noise current increases quickly caused by the degradation of the transimpedance gain.

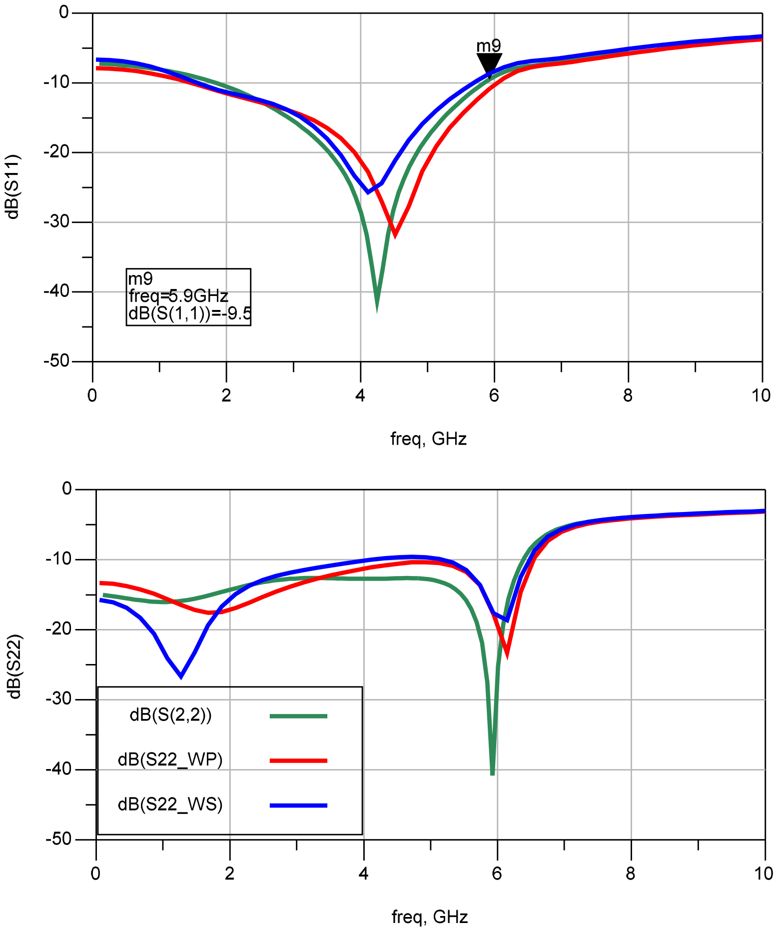


Figure 3.22: Simulation results of S_{11} and S_{22} of the TIA in AMS C35 with different resistor models

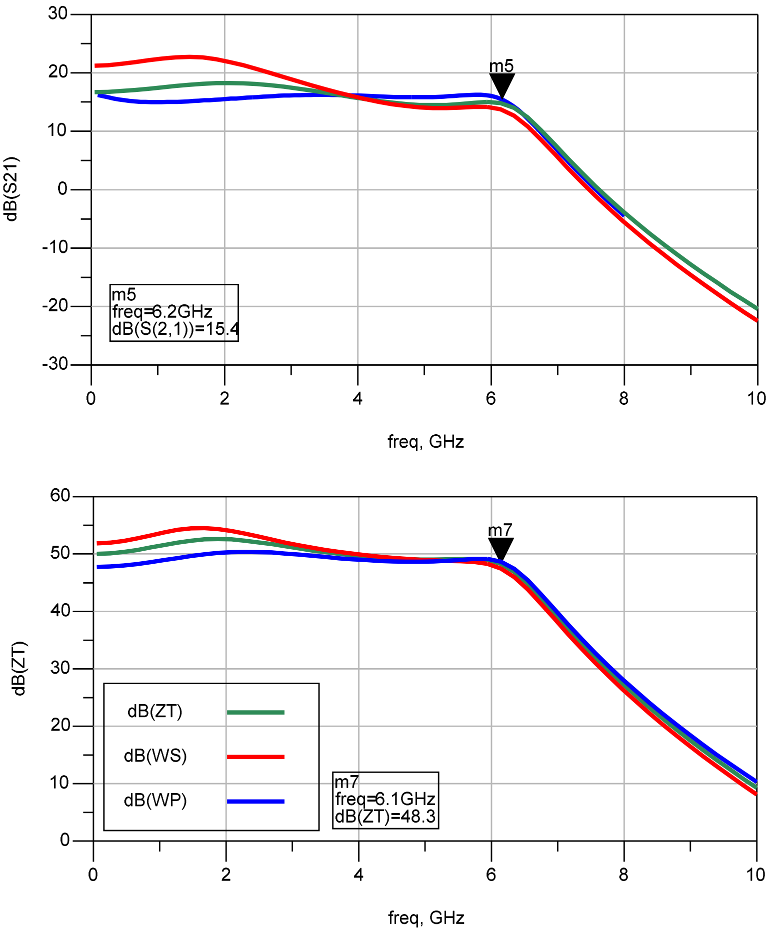


Figure 3.23: Simulation results of S_{21} and Z_T of the TIA in AMS C35 with different resistor models

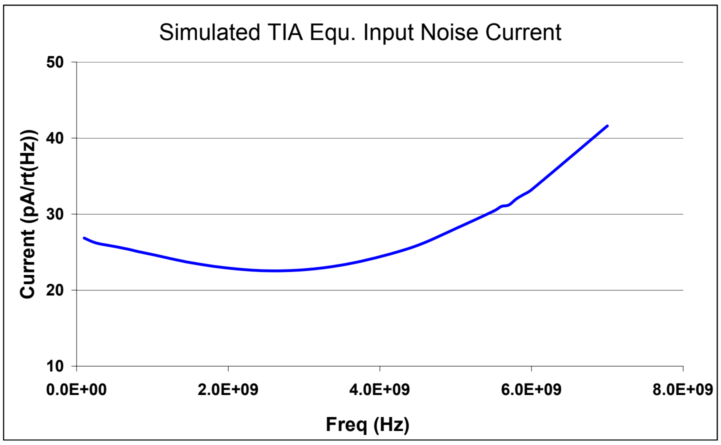


Figure 3.24: Simulated input-referred noise current of the TIA in AMS C35

3.3.5 Layout

To avoid the parasitics from the packaging and bonding, the TIA prototype circuit was designed to be tested with coplanar probes G-S-G (Ground-Signal-Ground). Correspondingly, G-S-G pads were employed for the input and output ports off-chip connection, which consist of three $95\mu m * 95\mu m$ naked pads with centre to centre separation of 100-μm. In fact, the parasitic capacitance of the pad has been absorbed into the input and output matching network, and all the previous simulation results have included the parasitic effect of the pad[8]. Thus the deembedded measurement of pad impedance is not necessary for TIA testing, which simplifies the measurement procedure. Simulation also shows that the RGC input stage can even accommodate an input capacitance up to 600 fF without significant degradation on -3 dB bandwith.

Layout of the TIA is shown in Fig. 3.25, the chip area is $500\mu m * 775\mu m$ (pads included). As can be seen from the diagram, passive components, i.e., inductors L_1, L_3 and pads, occupy most of the space of the amplifier, causing unavoidable

---

[8] The parasitic capacitance has a typical value of $306 fF$ (maximum of $400 fF$), caused mainly by the capacitance between metal 1 and substrate.

long inter-connections between different cells. To minimize the parasitic capacitance and resistance of wiring, metal 4 is used as far as possible for inter-connections in signal path. For the TIA in Fig. 3.18, node X and A are most sensitive to parasitic capacitance whilst node B is alleviated partially by inductor-series peaking of L_3. Simulation shows that additional 60 fF at node X or 30 fF at node A would shrink the -3 dB bandwidth of the amplifier to be lower than 6-GHz. For this reason, transistors M1, M2, M3 and resistors R_1, R_f are put as close as possible to minimize the inter-connection capacitance. Moreover, since the output port is more sensitive to parasitic capacitance than the input port, active components (M1-5) are placed near the output pad whilst long wiring is used from input pad to the active region.

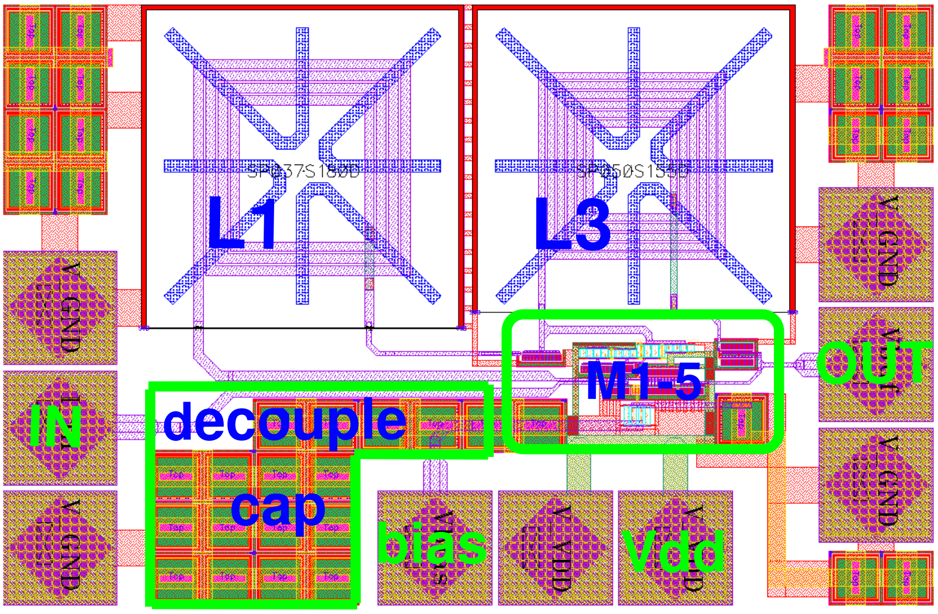


Figure 3.25: Layout of the TIA in AMS C35

The components are laid out in one direction (from left to right in Fig. 3.25) in order to avoid signal feedback caused by crossover parasitics. This technique also aids stability of the amplifier. Decoupling capacitors are filled in the corners and spare space, which helps shunt high frequency noise in power supply and bias voltage. With Cadence Assura parasitic extraction tool, post layout simulation was done with all the parasitic resistors and capacitors taken into account (parasitic inductors are ignored here for its tiny effect). Post layout simulations show little differences from pre-layout simulation results.

3.4 Measurement

3.4.1 Calibration for on wafer measurement

For on wafer measurements, The reference plane of calibration must be set at the tips of probes. SOLT (Short Open Load Through) is the mostly used calibration method with fair accuracy for measurement frequency up to 10 GHz [86]. LRM (Line Reflection Matched) can be used if further accuracy is needed. Standard SOLT calibration kit can be used if the DUT has standard coaxial connectors. For on wafer measurement, the calibration circuits, i.e., short, open, load, through can be made on wafer for calibration test structure, but it is difficult to ensure the electrical parameters of these circuits , i.e., the load impedance, the delay of the transmission line, to be accurate, fixed and not dependant on process, temperature or environment variations. Hence, generally a standard calibration substrate is provided together with the probes, which includes the standard calibration devices, and verification devices on it.

For this TIA measurement, G-S-G probes from Cascade Microtech Ltd were used for the measurement of TIA and the optical front end, which features a $100\mu m$ pitch separation. The calibration substrate used is 101-190B, also from Cascade Microtech Ltd , with the parasitic coefficients listed in table 3.4:

C_{open} (fF)	L_{short} (pH)	L_{term} (pH)
-9.3	2.4	-3.5

Table 3.4: The calibration coefficients for $100\mu m$ G-S-G probe from Cascade Microtech

Generally each time prior to the measurement, the calibration procedure should be conducted before the test, Fig. 3.26 shows a typical calibration verification result with a standard delay of 14 ps for the verification transmission line.

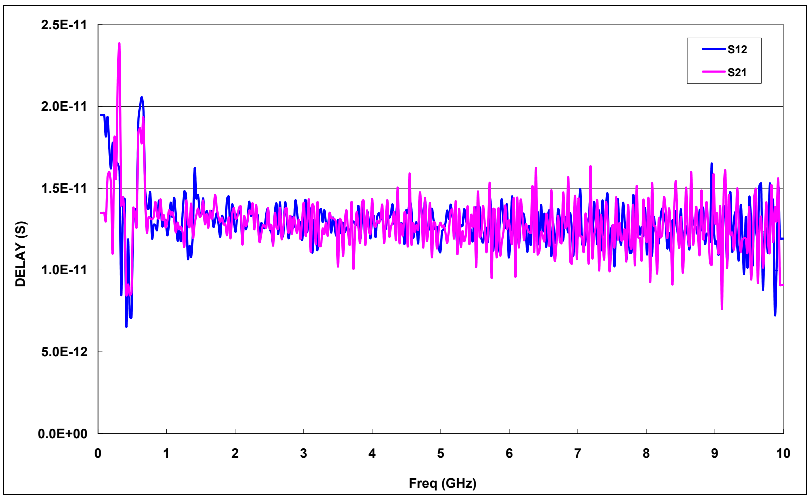


Figure 3.26: The verification delay time after calibration from 40 MHz to 10 GHz (standard 14 ps)

3.4.2 Frequency response measurement of the TIA

The frequency measurement set up of the TIA is shown in Fig. 3.27, while the micrograph of the TIA is shown in Fig. 3.28. Compared to the optical measurement set up, less instruments are used here. The input and output of the TIA are connected to the 2 ports of VNA (Anritsu 37397D) via G-S-G probes which provide 50Ω match. No bias Tee is needed here since VNA 37397D integrates bias Tee inside for port 1 and port 2. It greatly simplifies the cable and adapter connections. A needle probe was used to give power supply for the TIA since enough decoupling capacitors were integrated on chip, and the inductance of the needle can be ignored. Simulation further verifies that even with around 100 nH parasitic inductance from the needle probe, the TIA has no significant loss on gain or bandwidth. To monitor the power consumption of the TIA, a multimeter was connected in series with the voltage source Keithley 230.

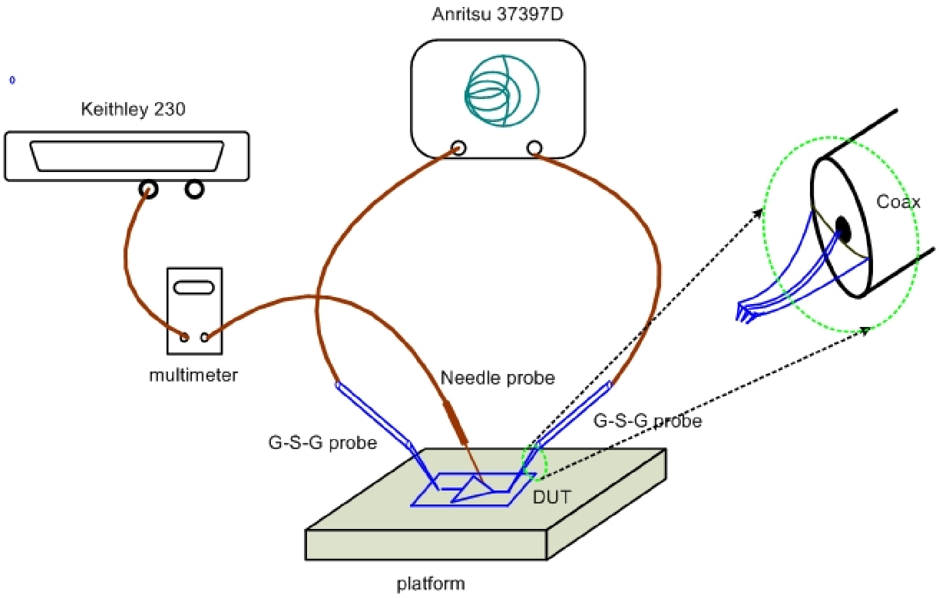


Figure 3.27: TIA frequency measurement set up

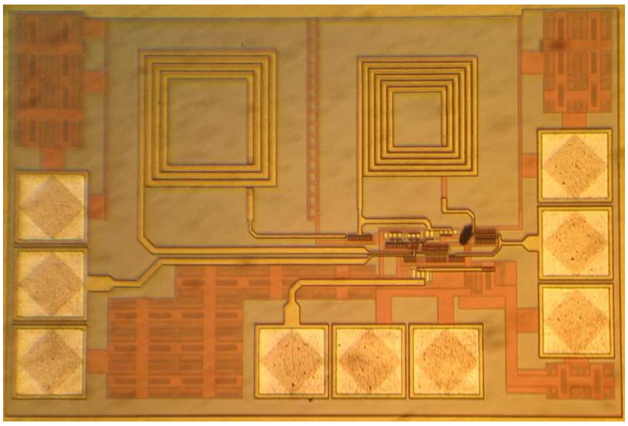


Figure 3.28: TIA die micrograph

The TIA was fully characterized under different power supply voltage ranging from 2.5 V to 3.6 V. With the S parameters recorded, the frequency response of the TIA can be post processed with the following equation (see section **3.3.4**):

$$Z_T = Z_0 \cdot \frac{S_{21}}{1 - S_{11}} \tag{3.109}$$

It should be noted here that the recorded file on the VNA for this TIA measurement must have S2P format to reserve both the magnitude and phase information since in equation (3.109), S21 and S11 are complex. If TXT file format was chosen for data recording, only the on screen displayed data will be saved. S11 here is only meaningful for the calculation of transimpedance gain, there is no design specs or constraints for S11 since the input impedance of the TIA was intended to be minimized (in this design is much smaller than 50Ω) to conduct more current from the photodiode to the amplifier. On the other hand, a good match is desired at the output of the TIA for the consideration of driving the mixer. For this reason, the following measurement results of the TIA only include S22 but not S11.

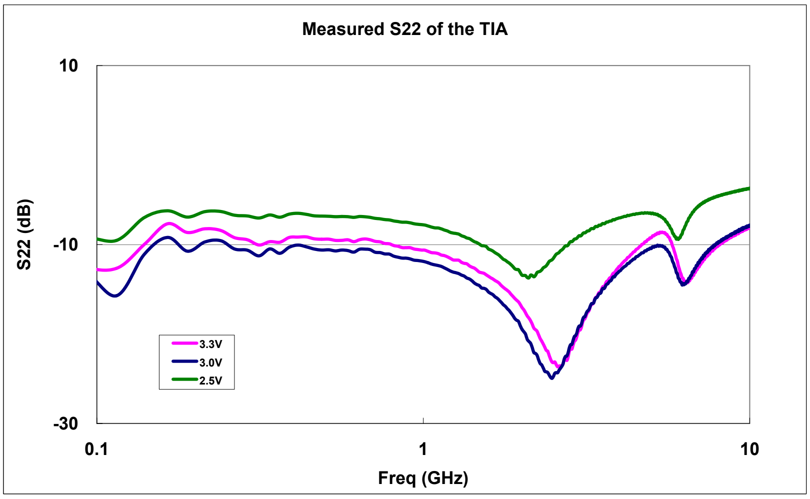


Figure 3.29: Measured S22 of the TIA under different power supply

Fig. 3.29 and Fig. 3.30 give the measured S22 and S21 of the TIA under different power supply. Fig. 3.31 shows the transimpedance gain of the TIA when power supply varies from 2.5 V to 3.3 V. In Fig. 3.32, the TIA's performance was summarized with a typical transimpedance gain of 51 dB, -3 dB bandwidth of 6.02 GHz. The S22 is lower than -8 dB from DC to 7 GHz. The TIA still presents a -3 dB bandwidth of 3.2 GHz at 2.5 V power supply. It achieves a fairly good gain flatness with gain variation smaller than ±2 dB within its -3 dB bandwidth.

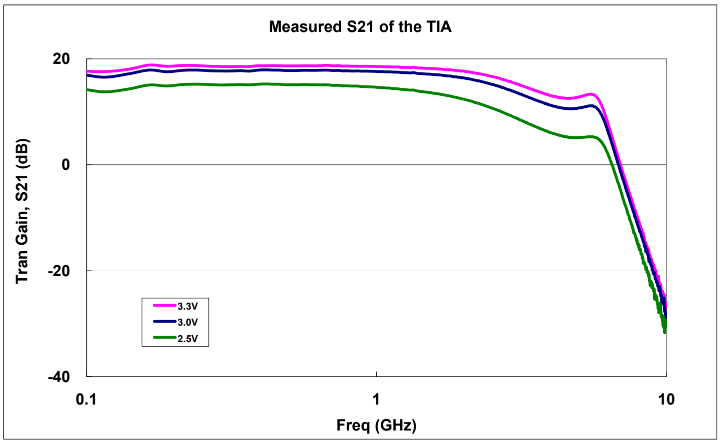


Figure 3.30: Measured S21 of the TIA under different power supply

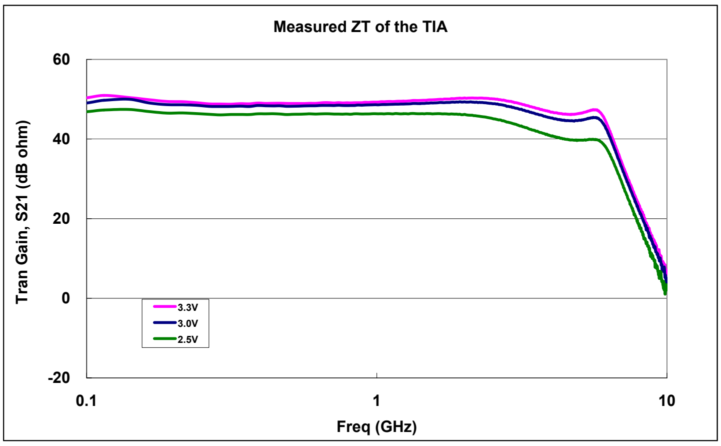


Figure 3.31: Measured transimpedance gain (ZT) of the TIA under different power supply

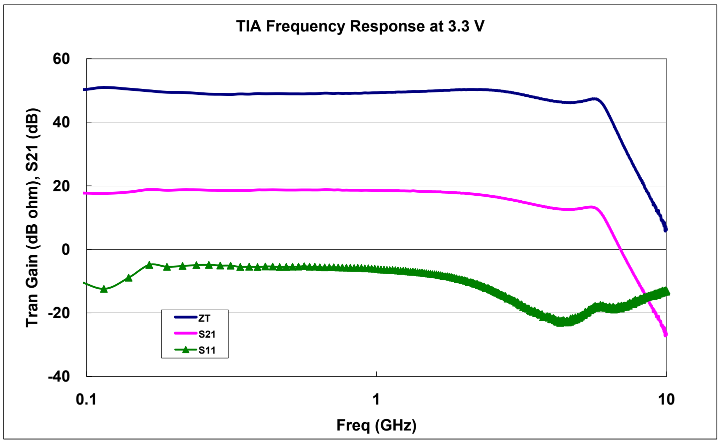


Figure 3.32: Measured transimpedance gain (ZT), S21 and S22 of the TIA at power supply of 3.3 V

For comparison, Fig. 3.33 shows the measured results and simulation results of the TIA for both S11 and transimpedance gain ZT at normal power supply of 3.3V. Very good agreement between the measured results and simulated has been achieved.

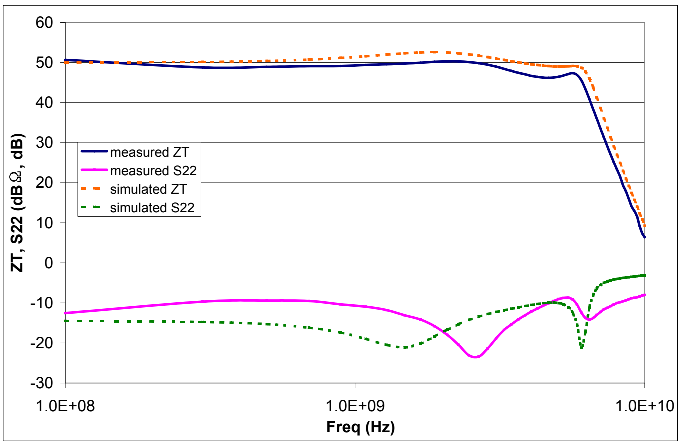


Figure 3.33: Comparison between the measured and simulated results of the TIA at 3.3 V power supply

3.4.3 Noise measurement of the TIA

The test setup for the noise measurement is shown in Fig. 3.34. The differences between noise test set up and frequency response test set up for the TIA are the spectrum analyser and the floating of TIA's input port. For noise measurement, it is important to set up the spectrum analyser correctly, particularly the set up of RBW (Resolution bandwidth) and VBW (video bandwidth). Generally for a spectrum analyser, RBW is determined by its intermediate frequency filter bandwidth, since noise is often broadband in nature, the broader the RBW, the higher the noise floor; the narrower the RBW, the lower the noise floor. The choice of RBW depends on the measurement frequency accuracy requirement, and signal or noise power level.

VBW is the bandwidth of the spectrum analyser's internal video filter, which is placed after the detector. In video filtering, the average level of the noise remains the same but the variation in the noise is reduced. Hence, the effect of video filtering is a "smoothing" of the noise. This is an important parameter that must be set appropriate when carrying the noise measurement.

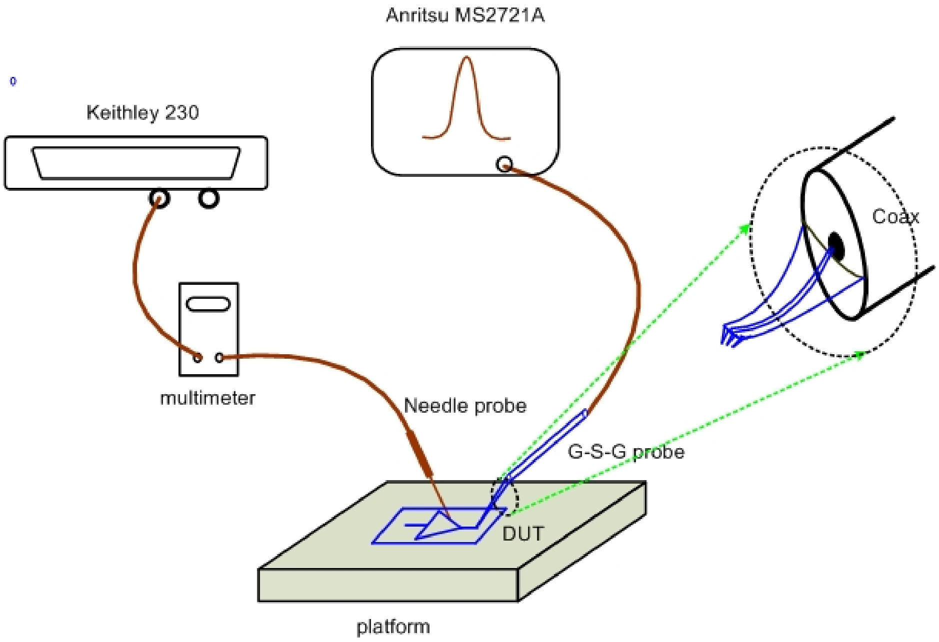


Figure 3.34: TIA noise measurement set up

Besides correct settings for RBW and VBW, low noise level of spectrum analyser is needed for the TIA noise measurement. Simulation predicts an input referred noise current of around 20 $pA/\sqrt{Hz}$ at 5GHz for the TIA, with 51 $dB\Omega$ transimpedance gain and 50Ω input impedance of the spectrum analyser, a rough estimate of the noise power density at the output port of the TIA can be written as

$$\begin{aligned} P_{nd} &= 10 * log[(20\, pA/\sqrt{Hz} * 10^{\frac{51dB\Omega}{20}}\,\Omega)^2/50\Omega * \frac{1}{10^{-3}}] \\ &\approx -150\, dBm/Hz \end{aligned} \tag{3.110}$$

If the RBW is set to 10 Hz, the output noise level should be around -140 dBm, which is a reference parameter for choosing the appropriate spectrum analyser. The chosen spectrum analyser used to measure the output noise power of the TIA was Anritsu MS2721A, which has a measurement frequency ranged from 100 KHz to 7.1 GHz, with tuning resolution of 1 Hz. Its SSB phase noise is -100 dBc/Hz at 10, 20, 30 KHz offset. MS2721A features a low noise floor when its embedded preamplifier is set on. Table 3.5 gives the displayed average noise level for the spectrum analyser. The gain of the pre amplifier is 25 dB and the RBW is set to 10 Hz.

Freq	noise level (dBm, RBW = 10 Hz)	NF (dB, preamp on)
10 MHz ~ 1 GHz	-153	11
1 GHz ~ 2.2 GHz	-150	14
2.2 GHz ~ 2.8 GHz	-146	18
2.8 GHz ~ 4 GHz	-150	14
4 GHz ~ 7.1 GHz	-148	16

Table 3.5: The noise level of MS2721A at different frequencies with RBW=10 Hz.

Sometimes to reduce the noise floor of the spectrum analyser, an additional amplifier can be inserted in the test path just before the spectrum analyser. However there is a stringent demand on the noise figure and gain flatness for this amplifier, otherwise it introduces additional uncertainty and variation to the measurement. For TIA noise measurement, amplifier ZX60-6013E from minicircuits was used to lower the noise level of the spectrum analyser, while giving insignificant differences on the measurement results.

The input referred noise current of the TIA was derived from the measured output noise power density divided by it transimpedance gain at that frequency spot. Fig. 3.35 shows the measured TIA input referred noise current with 3.3 V power supply and 3.0 V power supply. Fig. 3.36 compares the measured TIA input referred noise current and simulated one.

In summary, the TIA consumes 16 mA current from 3.3 V power supply, while the current drops to 9 mA at 2.5 V power supply. Typically it presents a transimpedance gain of 51 $dB\Omega$, with a -3 dB bandwidth of 6.02 GHz. The input referred noise current is about 25 $pA/\sqrt{Hz}$ at 5 GHz. This is the fastest TIA ever reported in 0.35 μm CMOS technology according to author's knowledge. Table 3.6 lists the recently published papers on high speed CMOS transimpedance amplifier.

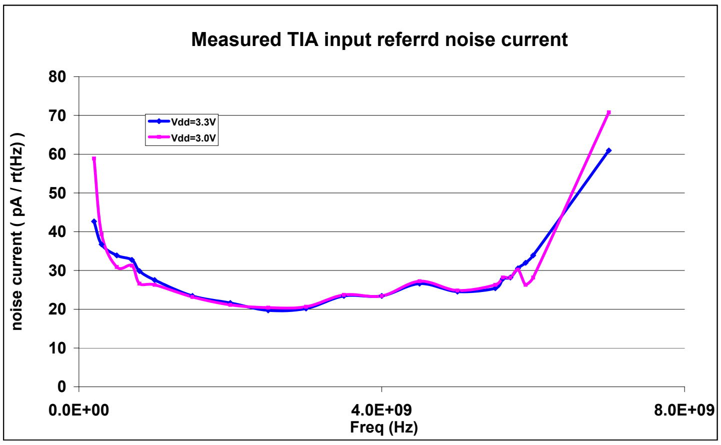


Figure 3.35: Measured TIA input referred noise current at 3.3 V and 3.0 V power supply

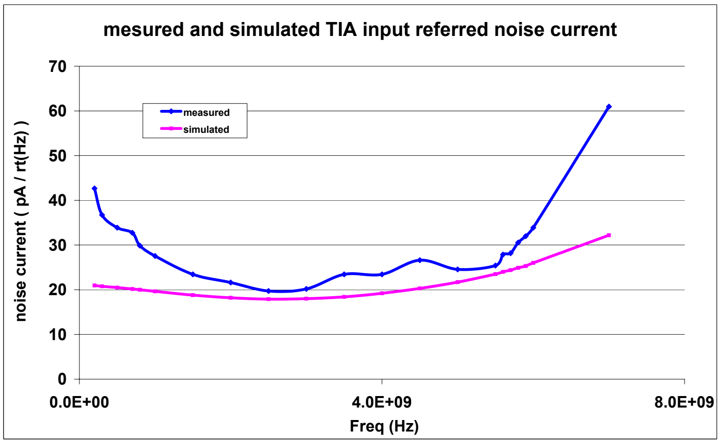


Figure 3.36: Measured and simulated TIA input referred noise current

Author	Technology	Gain ($d B\Omega$)	B.W. (GHz)	PD. Cap. (pF)	input-ref. Noise ($pA/\sqrt{Hz}$)
2001, Kuo [89]	0.35μm CMOS	54	1.35	–	–
2002, Schrodinger [88]	0.35μm CMOS	–	1.25	1.0	16
2003, Chiang [87]	0.35μm CMOS	52	2.1	–	–
2004, Park [77]	0.25μm CMOS	80	0.67	1.0	21
2006, Chen [105]	0.35μm CMOS	55	2.2	0.3	16
This work [91]	0.35μm CMOS	**51**	**6**	**0.6**	**21**

Table 3.6: Comparison of recently published high speed TIA in CMOS technology

Chapter 4

Mixer

4.1 Introduction

In addition to the proposed high speed photodiode and TIA implemented in CMOS, mixer is another important circuit block in the optical front end. Fig. 4.1 shows the block diagram of the proposed CMOS optical front end. Different from general optical receivers for high speed communication systems where signals are broadband and non-modulated, an acoustic wave imaging system has useful signals only located at the mode lock frequency and its harmonics. Therefore a broadband mixer is placed after the TIA for downconversion of the discrete high frequency harmonics. The IF amplifier and output buffer follow the mixer to give further amplification and provide 50 Ω matching to the measurement equipment.

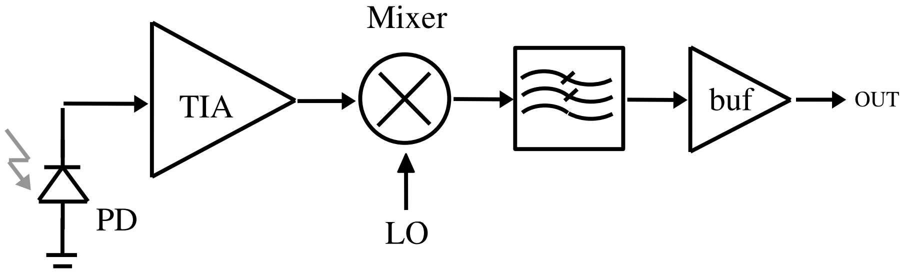


Figure 4.1: Block diagram of the optical front-end

Mixers perform frequency translation by beating two signals in a nonlinear element to produce sum and difference frequencies. They are critical components in receivers and transmitters to attach or remove the baseband signal to or away from the carrier. Mixer was first utilized in the superheterodyne receiver, reported by Armstrong [90] and called as the “first detector”, to convert the incoming high frequency signal to a lower intermediate frequency. By lowering the signal frequency with mixers, it is much more easier to obtain requisite gain and signal filtering in the receiver chain. Additionally, the potential oscillation problem due to the high receiver gain does not exist because the overall gain is distributed over different frequency bands [83]. Basically mixers are nonlinear time variant circuits, their operation is based on either switching of the signal path, or the modulation of a nonlinear component by a large signal tone, with the presence of the information-contenting signal to be converted in frequency. The nonlinearity caused mixing property is discussed in following section.

4.1.1 Mixing phenomena

The simplified nonlinear transfer function of a circuit can be described as equation (4.1) :

$$V_o = \alpha_1 V_s + \alpha_2 V_s^2 + \alpha_3 V_s^3 + ... + \alpha_n V_s^n + ... \tag{4.1}$$

where V_s and V_o are the input and output signals respectively. The n'th harmonic of V_s is generated by the power harmonics of V_s in equation (4.1). For instance, if V_s is a single-tone signal represented by $Acos(\omega t)$, where A is the amplitude and ω is the fundamental frequency, the square of V_s generates the second harmonic $cos(2\omega t)$ while the cube of V_s generates the third harmonic $cos(3\omega t)$ as :

$$V_s^2 = (Acos(\omega t))^2 = \frac{A^2}{2}(1 + cos(2\omega t)) \tag{4.2}$$

$$V_s^3 = (Acos(\omega t))^3 = \frac{3}{4}A^3 cos(\omega t) + \frac{1}{4}A^3 cos(3\omega t) \tag{4.3}$$

It is also noted that the n'th even power of V_s also generates other even harmonics (including DC component) with lower order than n whilst the n'th odd power of V_s also generates other odd harmonics (including fundamental component) with lower order than n.

The nonlinear terms in equation (4.1) generates frequency-mixing products if V_s comprises more than one signal with different frequencies. This phenomena is known as intermodulation. For instance, if V_s is the sum of two signals of different frequencies ω_a and ω_b :

$$V_s = A_a cos(\omega_a t) + A_b cos(\omega_b t) \tag{4.4}$$

the square of V_s generates

$$\begin{aligned} V_s^2 &= (A_a cos(\omega_a t) + A_b cos(\omega_b t))^2 \\ &= \frac{A_a}{2}(1 + cos(2\omega_a t)) + \frac{A_b}{2}(1 + cos(2\omega_b t) + \\ &\quad + A_a A_b [cos(\omega_a + \omega_b)t + cos(\omega_a - \omega_b)t] \end{aligned} \tag{4.5}$$

where $cos(2\omega_a t)$ and $cos(2\omega_b t)$ terms are the second harmonics of the two signals respectively. In addition, the $cos(\omega_a + \omega_a)t$ and $cos(\omega_b - \omega_b)t$ terms are the second-order intermodulation products which are located at the sum and difference frequencies of ω_a and ω_b respectively.

Similarly, the cube of V_s generates intermodulation products at frequencies $(2\omega_a \pm \omega_b)$ and $(2\omega_b \pm \omega_a)$ where the second harmonic of one signal mixes with the fundamental harmonic of the other signal. This phenomena is known as third-order intermodulation. Unlike the second-order intermodulation products, the third-order intermodulation products at frequencies $(2\omega_a \pm \omega_b)$ and $(2\omega_b \pm \omega_a)$ are located close to the fundamental signals. This may result in undesirable consequences, for instance, corruption of the information-contenting signal at frequency $(\omega_a - \omega_b)$.

Generally, if V_s consists of several frequency tones or a modulated channel, the separate frequency tones intermodulate each other, thus influencing linear combinations of all excitation frequencies. Their harmonics producing a response

$$f_{n,m,k} = nf_1 + mf_2 + kf_3 + ..., \qquad n,m,k \in Z. \tag{4.6}$$

4.1.2 Mixer topologies

In terms of frequency conversion direction, there are basically two kinds of mixers, i.e., down-conversion mixers in receiver which convert the RF signal to an IF (Intermediate Frequency) and up-conversion mixers in transmitter which convert IF to RF. From the equation (4.5), the mutual dependencies of the IF, RF, and local oscillator (LO) frequencies are simplified as :

$$f_{IF} = |f_{RF} \pm f_{LO}| \tag{4.7}$$

This chapter focus on the down conversion mixer since we are trying to demodulate the high frequency signal from the laser for post processing. A typical RF receiver architecture is illustrated in Fig. 4.2. RF filter and low noise amplifier perform the pre-selection and amplification, which are necessary for the specified selectivity and sensitivity of the system. Mixer then converts the RF signal to IF by mixing the RF signal from the preamplifier with LO (Local Oscillator), which is a very important building block within the receiver front-end as its performance affects the system performance and the performance requirements of adjacent building blocks. For instance, a mixer with low noise reduces the gain requirement on the previous amplifier; a mixer with high gain suppresses the noise contribution from the following stages; a mixer with high IP3 improves the system's linearity and selectivity.

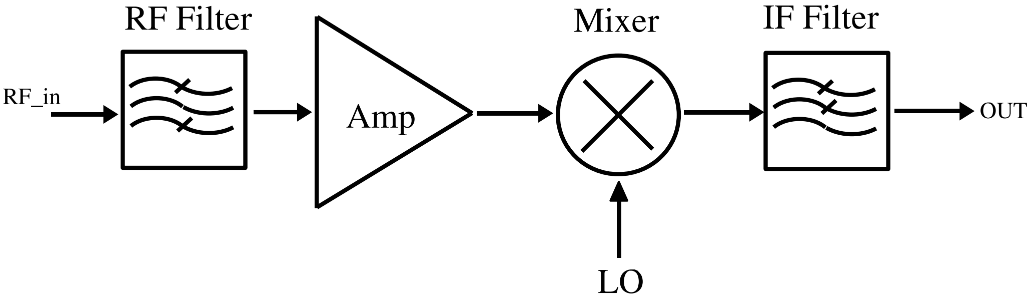


Figure 4.2: A typical RF receiver architecture

In terms of conversion gain, downconversion mixers can be categorized into passive and active mixers. Passive mixers, i.e., diode mixers and passive field effective transistor (FET) mixers generally introduce conversion loss although the circuits are very linear and can operate at very high frequencies. On the other hand, active mixers provide conversion gain to reduce the noise contribution from the following stages. For the application in this thesis, active mixers with conversion gain are explored aiming to suppress the noise from the off-chip circuits.

A typical single-balanced active mixer shown in Fig. 4.3 is a common kind of balanced active mixer. By means of balanced, it implies that there is no LO-to-RF feedthrough and LO-to-IF feedthrough since the LO signal has no DC component. The mixer comprises of a driving stage (a common-source transconductance stage in Fig. 4.3) and a differential switching pair. The common-source transconductance stage amplifies the RF signal to compensate for the attenuation caused by the switching operation, and to reduce the noise contribution from the switching stage. The amplified RF signal is converted down to IF (or up in upcovnersion mixers) by the LO switching pair. Single-balanced mixer is advantageous for the simple architecture, but it suffers from the LO feedthrough at the IF output port caused by the DC component in the RF signal.

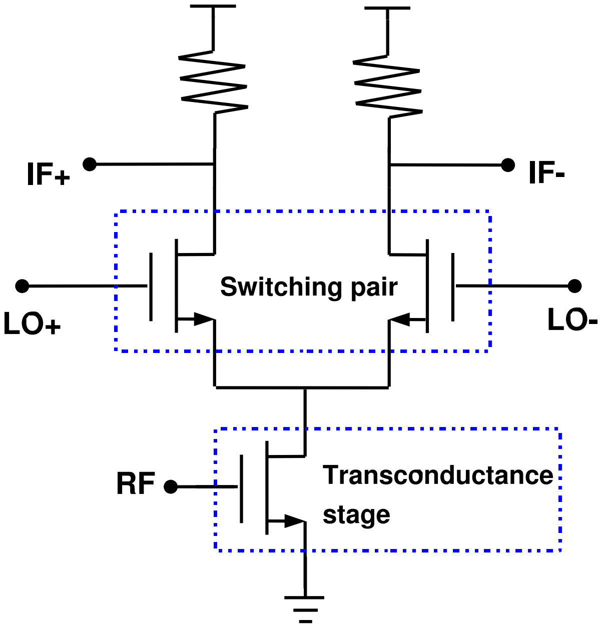


Figure 4.3: Single-Balanced Active mixer

Double-balanced active mixer shown in Fig. 4.4 is another most frequently used balanced mixer. The core is a Gilbert cell which comprises a differential pair driving stage and a switching quad. Resistors instead of PMOS transistors are commonly employed as load of the mixer for lower parasitic capacitances. Degeneration components (Z_e) sometimes are added in series with the source of the transconductance transistor in order to increase the linearity of the driver stage, but at the cost of loss of conversion gain.

The operation of Double balanced mixers is similar to that of single-balanced mixers. However, it has the advantage of rejecting the strong LO component and even-order distortion products due to the double-balanced structure [92]. Moreover, it offers high port to port isolation, greater immunity from certain interference mechanisms such as supply voltage noise, substrate noise compared with other kinds of mixers.

Besides single-balanced and double-balanced mixers, there is a third group of active mixers, namely the unbalanced mixers. Fig.4.5 and Fig. 4.6 show two different circuit topologies of unbalanced mixers. In both topologies, the mixing operation is performed by modulating transconductance of the driving stages with

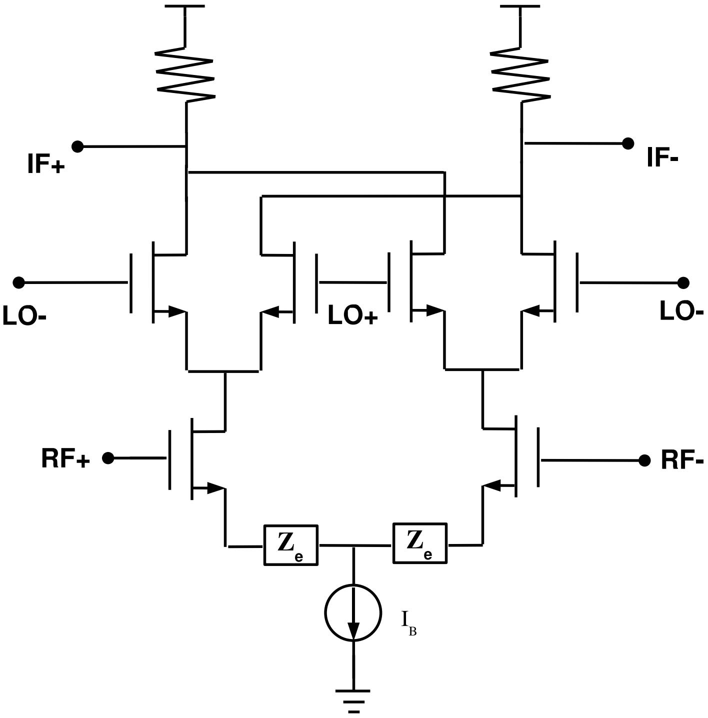


Figure 4.4: Double-Balanced Active mixer

the LO signals. In the single-transistor active mixer shown in Fig. 4.5, the LO signal modulates the transconductance of the common-source driving stage by varying the gate-source voltage (V_{gs}) of the NMOS transistor. In the dual-gate FET mixer shown in Fig. 4.6, the LO signal modulates the transconductance of the common-source driving stage by varying the drain-source voltage (V_{DS}) of the lower transistor. Since there are DC components in both RF and LO signals, unbalanced mixers do not reject LO-to-IF and RF-to-IF feedthrough signals.

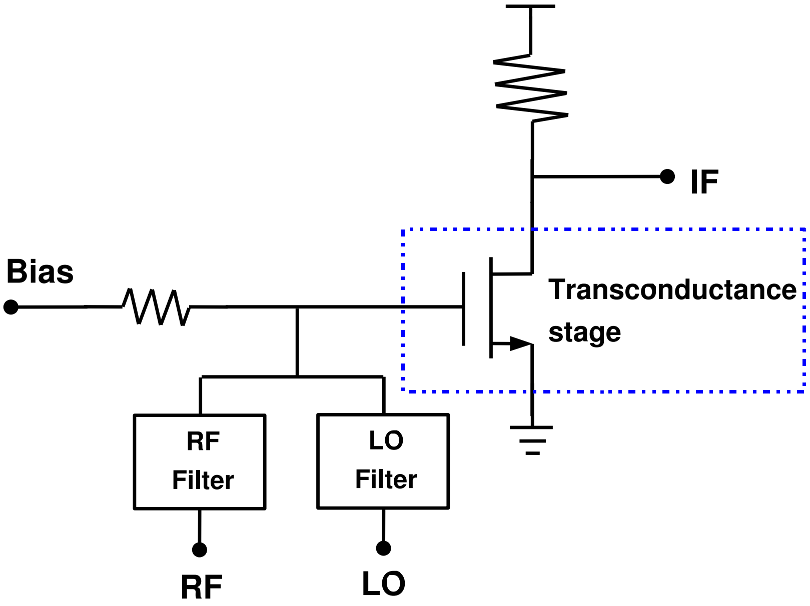


Figure 4.5: Single -Transistor Active mixer

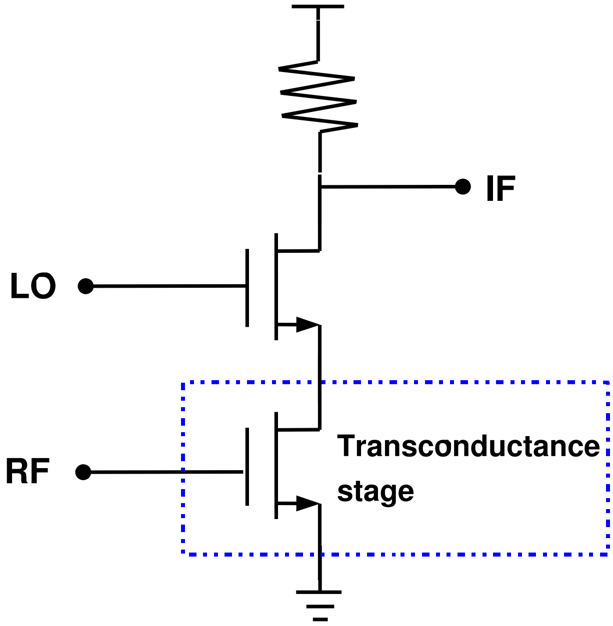


Figure 4.6: Dual-Gate FET mixer

Both single-balanced and double-balanced mixers reject LO-to-RF feedthrough if the differential switching stages are driven differentially. However, in the unbalanced mixers, LO-to-RF feedthrough is not rejected because the LO signal is unbalanced. In the single-transistor active mixer shown in Fig. 4.5, the LO signal is injected into the RF port through the RF filter. In the dual-gate FET mixer shown in Fig. 4.6, the LO signal is injected into the RF port through the gate-to-drain capacitance of the lower transistor.

The unbalanced mixers have the best noise performance due to the simplicity of the circuitry. In other words, there are fewer noise contributors, compared to both single-balanced and double-balanced designs. However, the unbalanced mixers suffer from their unbalanced properties. For instance, the RF and LO feedthrough cause self mixing problem and interferences to adjacent circuit blocks. Moreover, the stringent requirements on RF filter and LO filter limit their applications.

4.1.3 Mixer performance parameters

Noise Figure The noise performance of the mixer is expressed usually by the noise factor or noise figure. The noise factor is defined as degradation of signal-to-noise ratio (SNR) caused by the circuit [93]. The noise figure is the noise factor in decibel scale. There are two types of noise figure measures for downconversion mixers, namely, single-sideband (SSB) noise figure and double-sideband (DSB) noise figure. The difference is, in the definition of DSB noise figure, both sidebands of LO signal (the RF signal and its image signal) are treated as desired input signal, while only one side of the LO frequency is retrieved in the definition of SSB noise figure (as shown in Fig. 4.7). Therefore, the downconversion of a DSB signal can be considered to have twice as much signal power as the downconversion of a SSB signal. With the same noise power, the SSB noise figure is about 2 times (3-dB) higher than DSB noise figure for a mixer.

Generally, the switching operation in mixers can be viewed as multiplication of the RF input by a rectangular waveform. Consequently, if the LO signal has a 50% duty cycle, the RF input is multiplied by all the odd harmonics of the LO signal. This suggests that RF noise components around $3\omega_{LO}$, $5\omega_{LO}$, etc., are downconverted to the IF band, further increasing the output noise (shown in Fig. 4.7). Nonetheless, this effect is suppressed by low conversion gain (even conversion loss) at high frequencies and small magnitude of higher order harmonics.

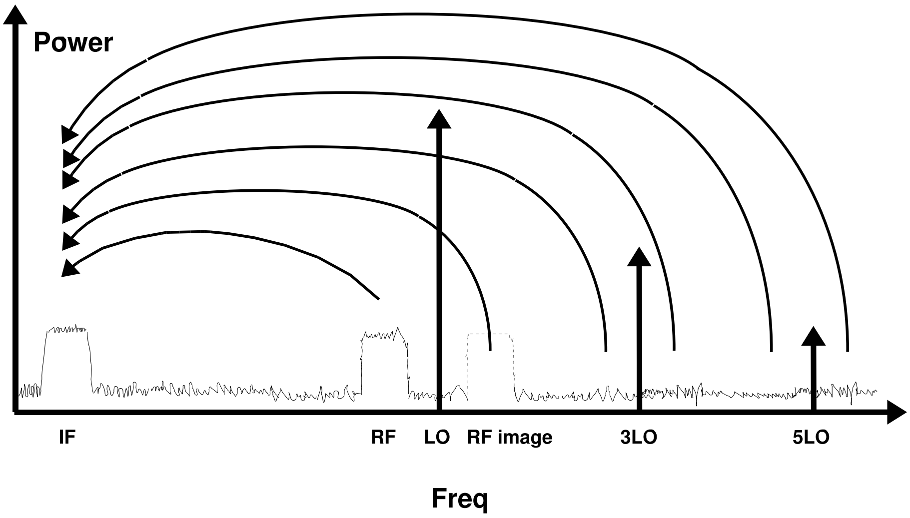


Figure 4.7: Noise aliasing from harmonic LO sidebands

Conversion Gain A downconversion mixer should provide sufficient power gain to compensate for IF filter loss, and to reduce the noise contribution from the following stages. However, this gain is not expected to be too large, as a strong signal may saturate the output of the mixer. Two kinds of gain, namely, power gain (G_p) or voltage gain (G_v) are typically referred in mixer specification. They can be related as

$$G_p = (\frac{V_o}{V_i})^2 \cdot \frac{R_S}{R_L} = (G_v)^2 \cdot \frac{R_S}{R_L} \tag{4.8}$$

where V_o and V_i are output and input voltages respectively; R_L and R_S are load and source impedance respectively. If the load impedance and source impedance of the mixer are equal, its power gain and voltage gain would have the same values in dB (decibel scale).

Nonlinearity Input referred intercept points are used as figures-of-merits to characterize the linearity performance of a receiver or a separate block like a mixer. They are extrapolated input amplitudes at which the desired signal becomes equal to the spectral component, which is generated by the respective intermodulating signals, as described in section 4.1.1. Fig. 4.8 shows conceptions of the 3-order intercept point[1], which is extrapolated input amplitudes at which the desired signal becomes equal to the third order harmonics.

Also shown in Fig. 4.8 is another widely used linearity performance parameter, the input 1-dB compression point (ICP), which is used to describe the large signal handling capabilities of the circuit. The gain compression occurs because of either the odd-order nonlinearity or voltage clipping and current limiting at high input signal power. The point where the large signal is 1-dB below the small-signal gain is defined as 1-dB compression point.

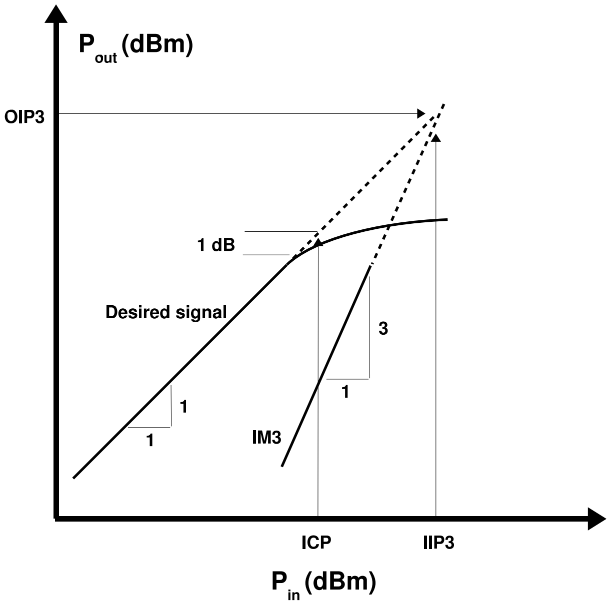


Figure 4.8: 3-order intercept- and compression points

Port Isolation The isolation between LO and RF ports of the mixer is important, as LO-to-RF feedthrough causes the self mixing from LO and its leakage. LO-to-IF and RF-to-IF feedthrough may saturate the IF output and decrease the 1-dB compression point of the mixer.

---

[1] It is noted that the intercept points are determined by a two-tone test at small input signal levels, when the nonlinear response is linearly dependent on the stimulus, as illustrated in Fig. 4.8.

Port Return Loss The impedance of the RF and LO ports are typically matched to 50Ω, while the impedance of IF output port is matched to that of the following stage, i. e., the IF filter. Impedance matching is necessary to avoid signal reflection and power loss in the mixer. Typically, return losses of less than -10 dB are required at RF port. At LO port, the return loss specification can be more relaxed since the amplitude of LO can be increased by delivering more power to the LO port, but at the expense of increased power consumption of the overall system.

4.2 Gilbert Cell Mixer Analysis

The principle of Gilbert cell has been briefed in section 4.1.2. The detailed operation of MOS FET Gilbert cell can be divided into two modes in terms of different LO amplitude : analog multiplier mode at small LO amplitude and mixing mode at large LO amplitude.

Small Signal Mode If a voltage with small differential amplitude of $2x$ (DC set at X) is applied to a MOS differential pair, as illustrated in Fig. 4.9(a), the output currents can be given as below by assuming all the transistors are in saturation region

$$I_1 = \frac{1}{2} K(V_{gs1} - V_T)^2 \tag{4.9}$$

$$I_2 = \frac{1}{2} K(V_{gs2} - V_T)^2 \tag{4.10}$$

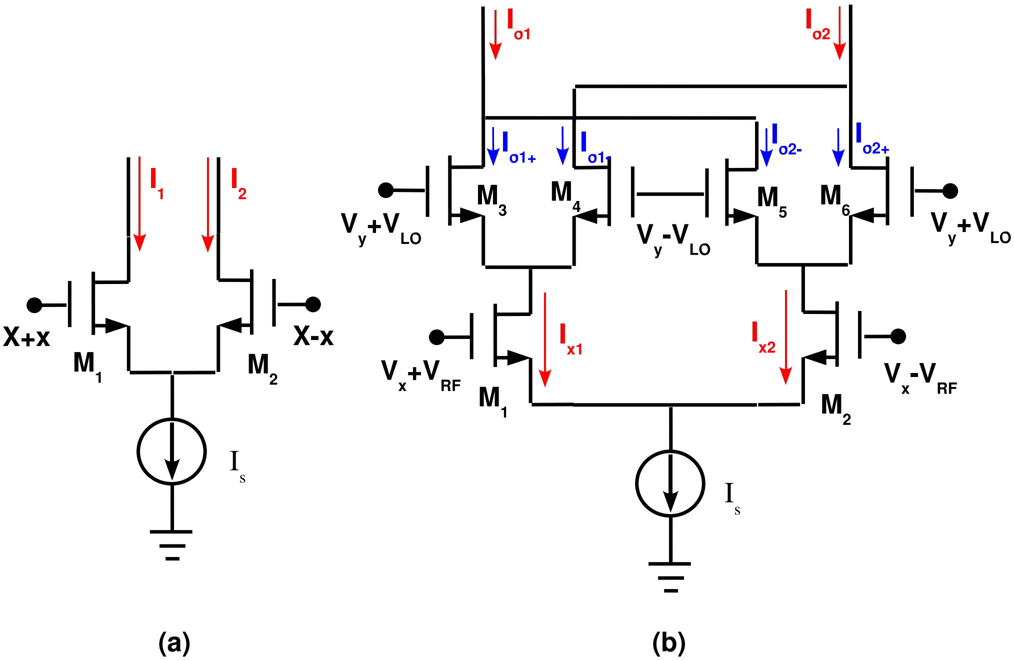


Figure 4.9: (a) Small signal analysis for differential pair (b) Gilbert cell as multiplier

where $K = \mu C_{ox} W/L$ and V_T is the conventional notation for the transconductance parameter and the threshold voltage of the MOS transistor, respectively. V_{gs1} and V_{gs2} are the gate source voltages of M_1 and M_2, respectively. it is assumed that the differential transistors (M_1 and M_2) have the same size ($W_{1,2}/L_{1,2}$) and V_T. Since the differential pair has the common source and the tail current is fixed as I_s, we have

$$I_s = I_1 + I_2 \tag{4.11}$$

$$2x = V_{gs1} - V_{gs2} \tag{4.12}$$

solving for V_{gs1} and V_{gs2} from equations (4.9) and (4.10)

$$\begin{aligned} V_{gs1} &= V_T + \sqrt{2I_1/K} && (4.13) \\ V_{gs2} &= V_T + \sqrt{2I_2/K} && (4.14) \end{aligned}$$

replacing the V_{gs1} and V_{gs2} in equation (4.12), we have

$$\sqrt{I_1} - \sqrt{I_2} = \sqrt{2K} \cdot x \quad (4.15)$$

with $(I_1 + I_2)$ and $(\sqrt{I_1} - \sqrt{I_2})$ known, it can be derived that $I_1 \cdot I_2 = (I_s/2 - Kx^2)^2$. so the differential output current is

$$\begin{aligned} I_1 - I_2 &= \sqrt{(I_1 + I_2)^2 - 4I_1 \cdot I_2} \\ &= \sqrt{KI_s} \cdot 2x \cdot \sqrt{1 - \frac{Kx^2}{I_s}} && (4.16) \end{aligned}$$

If the differential input voltage amplitude ($2x$) is small, we have

$$\begin{aligned} I_1 - I_2 &= \sqrt{KI_s} \cdot 2x \\ &= g_m \cdot 2x && (4.17) \end{aligned}$$

where g_m is the transconductance of transistor M_1 and M_2; $2x$ is the differential input voltage amplitude. As expected, the above derivation leads to the same result as the simple equation $\triangle I = g_m \triangle V$ gives with the small signal mode assumption, $Kx^2 \ll I_s^2$.

Based on the above conclusion, we can have the quantitative analysis of Gilbert cell in small signal mode. In Fig. 4.9 (b), the output differential current is

$$\begin{aligned}
I_{o1} - I_{o2} &= (I_{o1+} + I_{o2-}) - (I_{o2+} + I_{o1-}) \\
&= (I_{o1+} - I_{o1-}) - (I_{o2+} - I_{o2-}) \\
&= \sqrt{K_{LO} I_{x1}} \cdot 2V_{LO} - \sqrt{K_{LO} I_{x2}} \cdot 2V_{LO} \\
&= 2V_{LO}\sqrt{K_{LO}} \cdot (\sqrt{I_{x1}} - \sqrt{I_{x2}}) \qquad (4.18)
\end{aligned}$$

where $K_{LO} = \mu C_{ox} W_3/L_3$ represents the conventional notation for the transconductance parameter and it is assumed that the quad transistors have the same size. From the equation (4.15), we have $(\sqrt{I_{x1}} - \sqrt{I_{x2}}) = \sqrt{2K_{RF}} \cdot V_{RF}$. Thus, the MOS Gilbert multiplier shown in Fig. 4.9 (b) yields

$$I_{o1} - I_{o2} = 2\sqrt{2K_{RF}K_{LO}} \cdot V_{RF} V_{LO} \qquad (4.19)$$

where V_{RF} and V_{LO} are the half voltage amplitude of the input RF port and LO port of the Gilbert cell. With a load resistance of R, the gain of the Gilbert cell in small signal mode is $(2\sqrt{2K_{RF}K_{LO}} \cdot V_{LO} \cdot R)$. It is noted that the voltage gain is not effected by the tail current I_s under the small signal assumption $Kx^2 \ll I_s^2$.

Large Signal Mode For general mixer applications, the quad transistors in Gilbert cell as shown in Fig. 4.9 (b) are operating in nonlinear, large signal mode (with large V_{LO}) for gain and noise considerations. In equation (4.19) , the voltage gain of the Gilbert cell increases as V_{LO} increases under the assumption that V_{LO} is small and all transistors remain in saturation region. However, further increment of V_{LO} pulls the quad transistors out of saturation region in the peaking phase. The transistors are in fact in a “switching” mode when the V_{LO} is large enough. For simplicity, assuming an ideal square wave is applied to the LO ports. Mathematically, the square wave with a frequency of ω_{LO} can be described as

$$\begin{aligned}
S(t) &= \frac{1}{2} + \frac{2}{\pi}\sum \frac{(-1)^{n+1}}{n} \sin(\frac{n\pi}{2}) \cos(n\omega_{LO}t) \qquad n = 1, 2, 3, ... \\
&= \frac{1}{2} + \frac{2}{\pi}\cos(\omega_{LO}t) - \frac{2}{3\pi}\cos(3\omega_{LO}t) + \frac{2}{5\pi}\cos(5\omega_{LO}t) + ... \qquad (4.20)
\end{aligned}$$

and its complementary square wave $\overline{S(t)}$ is written as

$$S(t) = \frac{1}{2} - \frac{2}{\pi}\cos(\omega_{LO}t) + \frac{2}{3\pi}\cos(3\omega_{LO}t) - \frac{2}{5\pi}\cos(5\omega_{LO}t) + ... \tag{4.21}$$

Although the quad transistors are in switching mode, the transconductance transistors $M_{1,2}$ stay in small signal mode because of the small RF input power, and so the output currents are given by

$$\begin{aligned} I_{x1} &= \frac{I_s}{2} + g_m V_{RF}\cos(\omega_{RF}t) \\ I_{x2} &= \frac{I_s}{2} - g_m V_{RF}\cos(\omega_{RF}t) \end{aligned}$$

where g_m is the transconductance of the driving stage transistors $M_{1,2}$. Assuming the quad transistors $M_{3,4,5,6}$ are in ideal switching mode, the circuit can be simplified as in Fig. 4.10 (a). An expression for the output currents can be found to be

$$\begin{aligned} I_{o1} &= I_{o1+} + I_{o2-} \\ &= I_{x1}\cdot S(t) + I_{x2}\cdot \overline{S(t)} \\ &= [\frac{I_s}{2} + g_m V_{RF}\cos(\omega_{RF}t)]\cdot[\frac{1}{2} + \frac{2}{\pi}\cos(\omega_{LO}t) - \frac{2}{3\pi}\cos(3\omega_{LO}t) + ...] \\ &\quad + [\frac{I_s}{2} - g_m V_{RF}\cos(\omega_{RF}t)]\cdot[\frac{1}{2} - \frac{2}{\pi}\cos(\omega_{LO}t) + \frac{2}{3\pi}\cos(3\omega_{LO}t) + ...] \\ &= I_s + \frac{2}{\pi} g_m V_{RF}[\cos(\omega_{LO} - \omega_{RF})t + \cos(\omega_{LO} + \omega_{RF})t + ...] \end{aligned} \tag{4.22}$$

Similarly, I_{o2} can be derived as

$$I_{o2} = I_s - \frac{2}{\pi} g_m V_{RF}[\cos(\omega_{LO} - \omega_{RF})t - \cos(\omega_{LO} + \omega_{RF})t + ...] \tag{4.23}$$

finally, the differential output current I_o is

$$I_o = \frac{4}{\pi} g_m V_{RF}[\cos(\omega_{LO} - \omega_{RF})t + \cos(\omega_{LO} + \omega_{RF})t + ...] \tag{4.24}$$

Noting that V_{RF} is the half amplitude of the RF input, the conversion gain of the mixer with a load resistance of R in large signal mode is

$$G_{conv} = \frac{2}{\pi} g_m R \tag{4.25}$$

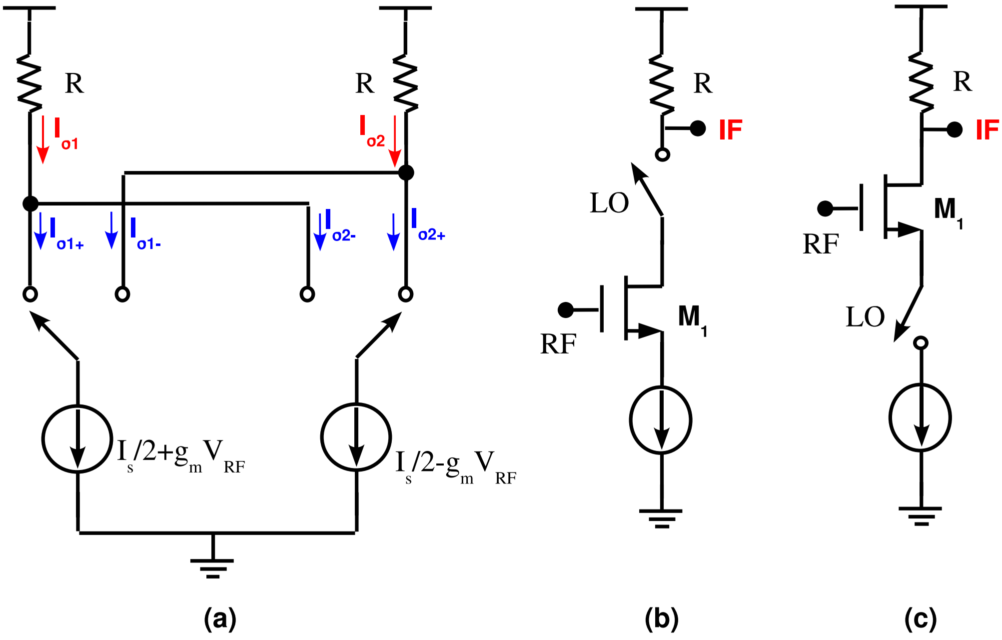


Figure 4.10: (a) Gilbert cell in switching mode (b) mixer simplified circuit with LO between RF and IF (c) mixer simplified circuit with RF between LO and IF

Ports Assignment It is interesting that most of the Gilbert cells employed as mixers apply the LO signals between the driving stage (where the RF signal is applied) and the IF output ports without explanation. For instance, LO signals are applied to the gates of transistors $M_{3,4,5,6}$ while RF signals are applied to transistors $M_{1,2}$ in Fig. 4.9 (b). Simplified circuits for this comparison are shown in Fig. 4.10 (b) and Fig. 4.10 (c), where LO is placed between RF and IF in the former whereas RF is placed between LO and IF in the later. The difference lies in the fact that, the DC operation of the driving stage (M_1) is greatly disturbed in the mixer circuit of Fig. 4.10 (c) because of the LO switching. For most mixer applications, RF input has a small power (amplitude) and the driving stage is used to amplify and convert the small input signal into a current one, which declares the impracticality of the configuration in 4.10 (b).

Noise Noise analysis of current-commutating CMOS mixers, i.e., Gilbert cell are reported in papers [95, 97, 96]. An active mixer comprises of an input transconductance, switches, and an output load. Noise is present in all the transistors making up these functions. Frequency translation of noise is dealt with approximately, using a simple analysis and qualitative reasoning. In [96], the output noise of a Gilbert mixer can be written as:

$$V_{o,n}^2 = 8kTR_L(1+\gamma\frac{R_L I}{\pi A}+\gamma g_m R_L) \tag{4.26}$$

where R_L is the resistance of load resistors, I is the total bias current of the mixer, A is the LO amplitude. g_m is the transconductance of the transistors from the transconductance stage. The first term in equation (4.26) is the noise due to the load resistor; second term comes from the switches (here we assume the square wave applied to the quad switches of the Gilbert cell); the third term reveals the thermal noise of the transconductance stage transferred to the mixer output. Note that in Gilbert cell there are two load resistors and two transconductance transistors, corresponding to the superposed thermal noise of $(2\cdot 4kTR_L)$ and $(2\cdot 4kT\cdot\gamma g_m\cdot R_L)$. In [96], switches are assumed to contribute noise to the mixer output over the time when they are both ON. If a switch is OFF, it obviously contributes no noise, and neither does the other switch that is ON because it acts as a cascode transistor whose tail current is fixed to I by the RF input transconductance stage. This is further illustrated in Fig. 4.11. If the channel modulation effect of M_1 is ignored, the impedance of node X is infinite, which means that all the current flowing through load resistor R will flow through M_1. The noise current of i_{n2} then has no effect on the output. Here the noise current is offset by M_2 itself, $i_{n2}+i_{D2}=0$. In practical design, the parasitic capacitances at node X magnify the noise from M_2 since which bypass the M_1 and conduct the current to ground.

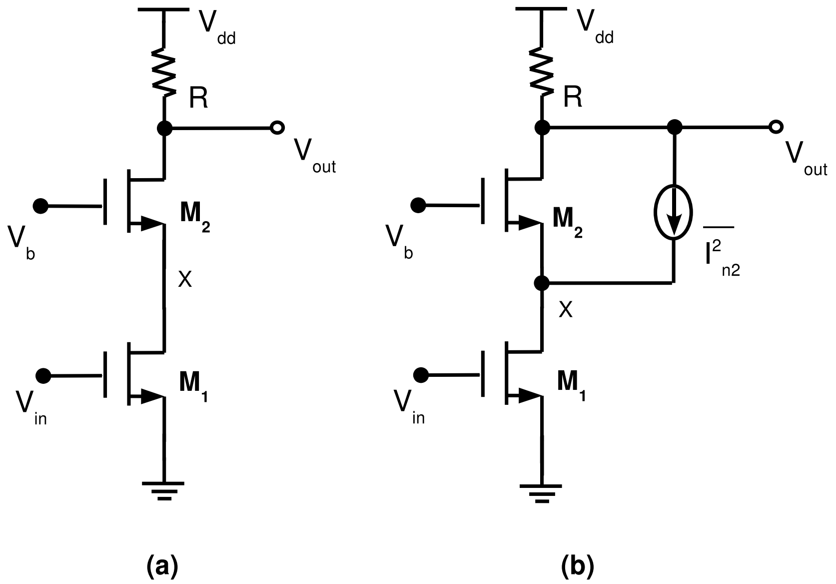


Figure 4.11: (a) Cascode stage; (b) noise of M_2 modeled by a current source

4.3 5 GHz Mixer in AMS 0.35μm CMOS

4.3.1 Optimization techniques in AMS 0.35μm CMOS

Austria Microsystems (AMS) 0.35-μm CMOS (C35B4C3) has been chosen as the prototype fabrication technology for the mixer fabrication. The double poly, four-metal, n-well process has a typical transit frequency f_T of about 25 GHz. RF models for transistors, resistors, capacitors, varactors, and inductors are also provided from the foundry, which cover the frequency range from DC to 6 GHz. AMS provides three inductor libraries, but only fourth-metal module is valid for C35B4C3 which contains 7 square inductors with values range from 1.4 nH to 9.1 nH. All these 7 inductors have a single layer of metal 4 and are modeled with a lumped RLC equivalent circuit.

It is difficult to operate the mixer at 5 GHz in 0.35-μm CMOS process, which demonstrates a typical f_T of about 25 GHz and f_{max} of 40 GHz. To minimize the gate series resistance, all the RF transistors are fingered with a fixed finger width of 5 μm. Referred to the RF models for NMOS transistors, the current density should be around $10^{-4}\ A/\mu m$ to achieve a maximum f_{max}, while V_{ds} should be as large as possible. In practical design, there are trade-offs between the circuit operation speed and power consumption since the current of the mixer is limited by the system specification and V_{ds} can not exceed the power supply in a Gilbert cell[2].

Fig. 4.12 shows the maximum oscillation frequency (f_{max}) versus I_{ds} for a typical NMOS transistor in AMS C35 with different V_{ds}. When the NMOS transistor is biased at $V_{ds} = 1.5\ V$, the maximum f_{max} of 43 GHz is achieved at an optimum I_{ds} of

$$I_{ds} = 10^{-4}\, A/\mu m \times 24 \times 5\ \mu m\ = 12\, mA \tag{4.27}$$

For a double balanced mixer with two RF transconductance transistors, it means the total bias current consumed is $2 \times 12\, mA = 24\, mA$. This is unacceptable for the optical front-end. In fact, the target power consumption for the mixer is $10\ mA$, which means that I_{ds} is equal to $5\ mA$ for each RF transconductance transistor. For NMOS transistor of size $24\times5\times0.35\ \mu m$, a f_{max} of 40 GHz is available from Fig. 4.12. It can be seen that for only 8% reduction in f_{max}, a saving of 60% in power consumption is achieved! For NMOS transistor with size of $40\times5\times0.35\ \mu m^3$, a f_{max} of 37 GHz is available based on simulation results. Still there is a significant power consumption cut with acceptable f_{max} for the NMOS transistors.

---

[2] In fact, V_{ds} is smaller than the power supply V_{dd} , which can be expressed as $(V_{dd} - 1/2 \cdot I \cdot R_L - V_{sat})$ if ideal switching is assumed.

[3] $40\times5\times0.35\ \mu m$ is the maximum size of NMOS RF transistors allowed by the RF models in AMS C35.

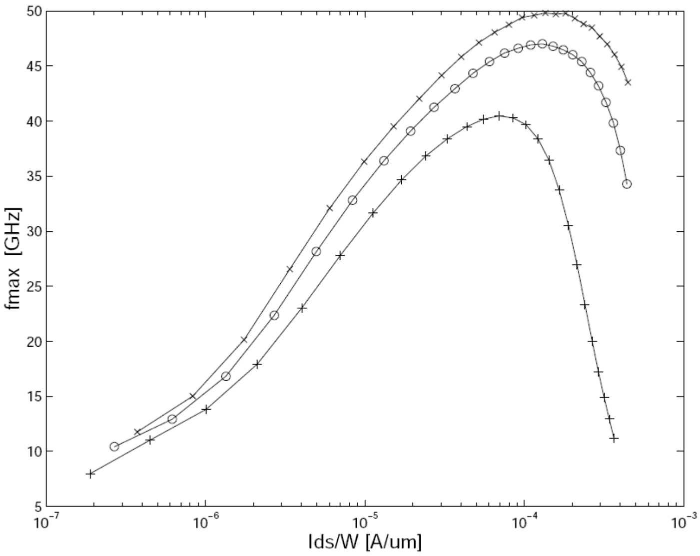


Figure 4.12: f_{max} of NMOS transistor 24×5×0.35 μm in AMS C35, V_{ds}=1 V (+), 2 V (○) and 3 V (×).

Because of the slow mobility of the holes in silicon, PMOS transistors are avoided to be used in the signal path when speed of the circuits are of great importance. Besides the transistors, the RF models for passive components have a big impact on the RF circuit performance. The accuracy of RF models of the passive components sometimes imposes as the bottleneck for RF design [83]. Fortunately, the RF models provided from AMS predict well of the high frequency performances of the inductors, capacitors and resistors in the libraries. Fig. 4.13 and Fig. 4.14 show the simulation results of Q factors for inductor SP037S180D and poly-poly capacitor (1 pF) with lumped RLC models from AMS C35 and ADS (Advanced Design System from Agilent) was used to do the simulation work for the mixer.

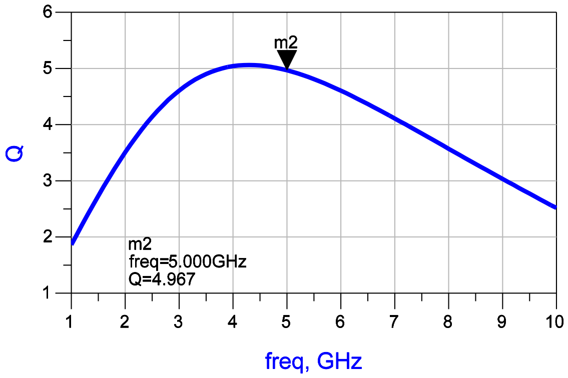


Figure 4.13: Q factor of the inductor SP037S180D (3.7nH) in AMS C35

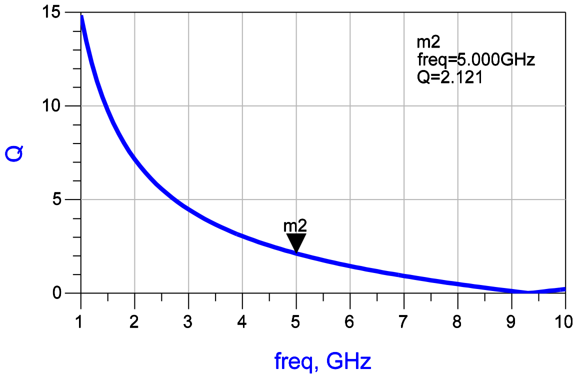


Figure 4.14: Q factor of the poly-poly capacitor (1 pF) in AMS C35

4.3.2 Matching network

Impedance matching for high frequency circuit design is important to avoid signal reflection and excessive ripples in the frequency responses. The impedance of the RF and LO input ports are typically matched to 50 Ω, while the impedance of IF port does not need to be 50 Ω since the signal frequency has been downconverted to a low intermediate frequency and it is generally followed by high impedance IF amplifier or filter, i. e., D2S in this design. But for the consideration of evaluating the mixer, D2S output has to be matched to 50 Ω, the typical impedance of a spectrum analyser. Basically, return losses of less than -10 dB (voltage wave standing ratio of less than 2) are required. On the other hand, the return loss specification on the LO port can be more relaxed because it can be compensated by higher LO power. However, excessive return loss would result in increased power consumption of the overall system.

Any arbitrary port resistance R_p can be matched to the source resistance R_s (for input port) or load resistance R_L (for output port) using a two-element impedance-matching network. If R_p is larger than R_s (or R_L), the impedance-matching network shown in Fig. 4.15(a) can be used. If the R_p is smaller than the R_s (or R_L), we can have the two elements matching network by just exchanging R_s and R_p, as shown in Fig. 4.15(b).

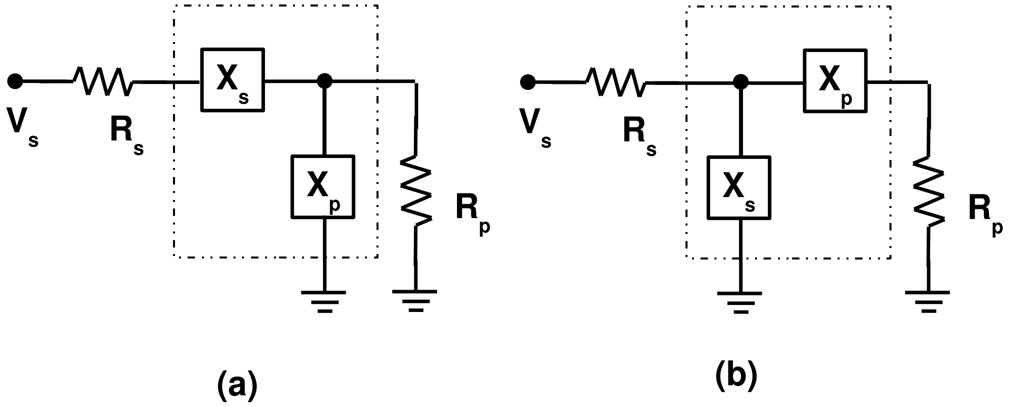


Figure 4.15: Matching network for (a) R_p larger than R_s; (b) R_p smaller than R_s.

The reactive elements X_s and X_p are used for impedance transformation. The quality factor (Q) of the network is given by

$$Q_N = \sqrt{\frac{R_p}{R_s} - 1} \tag{4.28}$$

The value of Q_N specifies the Q of X_p in parallel with R_p and the Q of X_s in series with R_s. In other words,

$$Q_N = \frac{R_p}{|X_p|} = \frac{|X_s|}{R_s} \tag{4.29}$$

The reactive elements X_s and X_p can be implemented by either inductors or capacitors. If X_s is capacitive, X_p has to be inductive. Similarly, if X_s is inductive, X_p has to be capacitive. The difference between these two configuration is that one is a lowpass network while the other is a highpass network. If the port impedance is not purely resistive, the reactive part of the port impedance can be tuned out by a reactive element, or absorbed into the matching network.

In practical design, the reactive elements (X_s and X_p) of the impedance-matching network can be implemented by either on-chip or external components. Generally on-chip components are preferred for system integration, although they have lower Q (due to the parasitic series resistance and substrate loss) than external components. Sometimes bond wires are utilized as the part of the matching network because of their parasitic inductance.

If the ratio of R_s to R_p is very large, the matching network is very sensitive to component variations. Correspondingly, S_{11} demonstrates a sharp notch to the frequency response. In this case, multi-stage matching network could be used to transform the impedance of R_p to an intermediate value, i. e., R_m before it is transformed to match the value of R_s. The sensitivity of the network can therefore be degraded.

For this design, the source impedance is 50 Ω, the input impedance of the RF port and LO port are capacitive and the resistance is much smaller than 50 Ω. On-chip matching-network is chosen for this design because the prototype chip was not to be packaged and all the measurements were to be based on-wafer. ADS design guide was chosen to do the matching-network, which provides different type of matching networks and gives the synthesized component values.

Because of the low Q of on chip inductors (Fig. 4.13 and 4.14), the synthesized components values are only approximated, the final matching-network was optimized as a simple inductor for both RF port and LO port, with values of 3.7 nH and 4.7 nH, respectively. The simulated reflection coefficients for RF port, LO port and D2S output port are shown in Fig. 4.16.

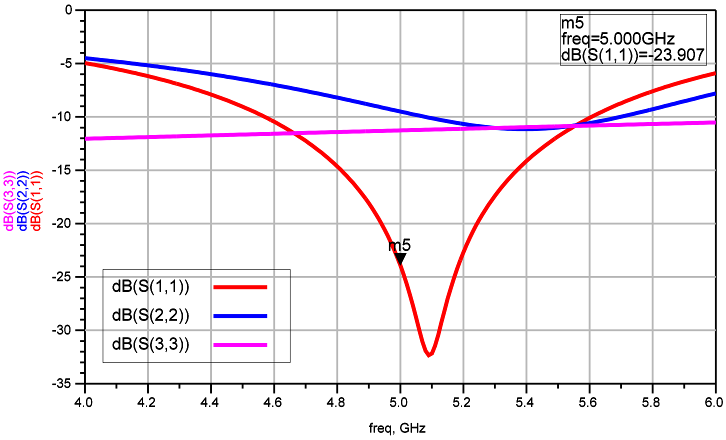


Figure 4.16: The reflection coefficients at the RF port (S_{11}), LO port (S_{22}) and D2S output port (S_{33})

4.3.3 LO driver and D2S

For the purpose of saving RF pads and for the convenience of measurement, LO driver and Differential to Single (D2S) circuits were designed and fabricated on chip together with the mixer. Besides, these auxiliary circuits are necessary for a realistic optical front-end system which intends to integrate more function blocks on chip. The LO driver schematic is shown in Fig. 4.17. M_1 and M_2 form the differential pair to amplify the input LO from off-chip. The gate of M_2 is AC grounded by a 7 pF capacitor. Resistor R_b provides DC operating point of M_2 by shorting the gates of M_1 and M_2. In this design, R_b is implemented by poly2 resistor and has a value of around 4 $K\Omega$ which together with the AC grounded capacitor form a low pass filter with a cut-off frequency of 6 MHz. For measurement set up of the mixer, DC bias voltage of the LO driver has been chosen to be around 1.4 V. It guarantees a reasonable V_{ds} for M_1 and M_2, at a tail current of 5 mA (2.5 mA for each branch of the differential pair), which is necessary for a f_{max}. On the other hand, it leaves enough headroom for the load resistors, for the consideration of the gain.

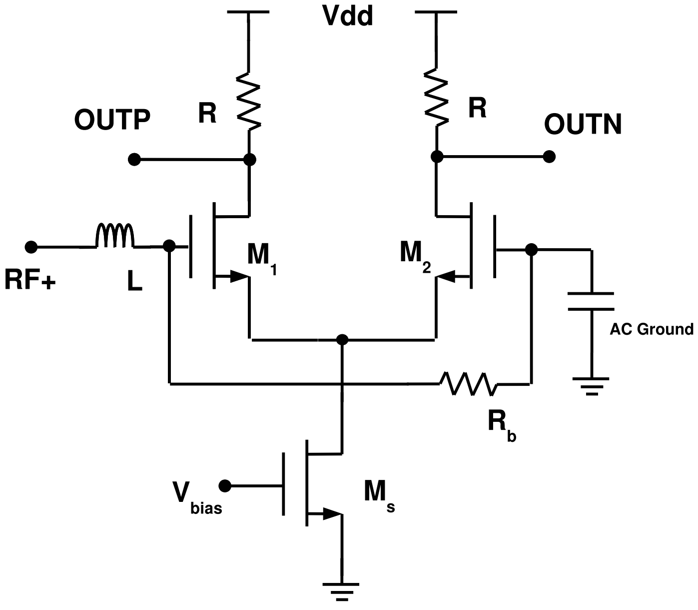


Figure 4.17: Schematic of the LO driver

To achieve the maximum gain for the circuit, an inductive load would have been preferable for the LO driver since it give better impedance and gain at high frequencies, but this comes with the cost of large chip area, which is unbearable because of the unavoidable inductors in TIA and input matching circuits of the mixer. Besides, another inductor is reserved for the LO driver which provides matching from off-chip to the input of the circuit. Simulation shows that at 5 GHz, appropriate resistor load still provides gain for the LO driver. With a tail current of 5 mA, the load resistor is chosen to be 300 Ω which is optimized for bandwidth, gain and the DC coupling to the LO ports of the Gilbert mixer.

It is difficult to have the two outputs of the LO driver with the same amplitude because of unmatched input to the differential pair, and the gate to drain parasitic capacitances C_{gd} of M_1 and M_2. In fact, V_{outp} has a larger amplitude than V_{outn}. Fortunately it does not impose a big problem so long as the output amplitude is large enough and the two outputs are out of phase. The design values for the components are listed in table 4.1.

component	design parameter	design value
M_1	$30 \times 5/0.35\mu m$	
M_2	$30 \times 5/0.35\mu m$	
M_s	$30 \times 5/0.35\mu m$	
R	$6 \times 100/3\mu m$	1818Ω/6
R_b	$2 * 120/3\mu m$	2182Ω×2
L	4.7 nH	4.7 nH
$C_{ac\ grounded}$	$7 \times 33.9 \times 33.9\mu m$	7 pF

Table 4.1: Design values of the LO driver

The AC simulation results of the LO driver are shown in Fig. 4.18. Without inductive peaking, the LO driver demonstrates 7.7 dB voltage gain at 5 GHz, which greatly reduces the power requirement of the LO input.

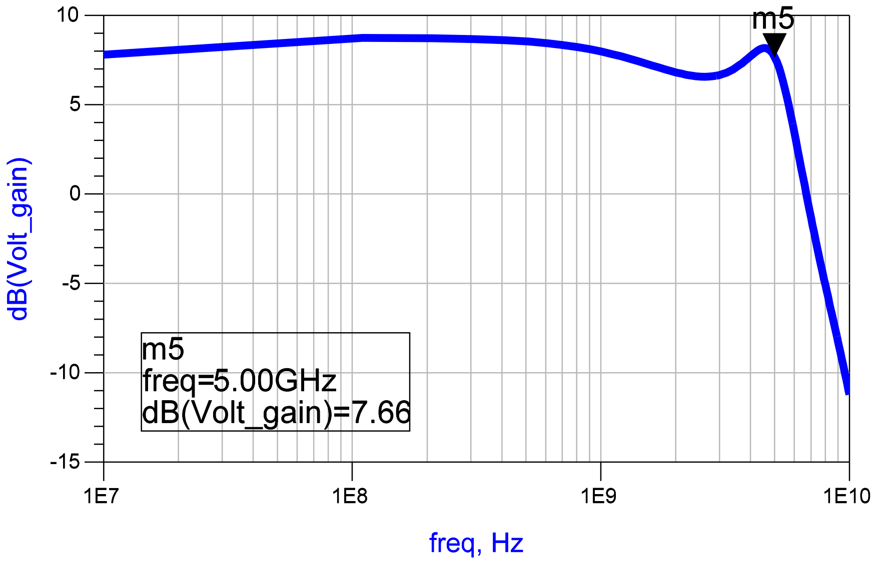


Figure 4.18: The frequency response of the LO driver

Both the prototype chips of the mixer subcircuit and the optical front-end were not packaged for this project, and all the measurements were to be conducted with probe station. G-S-G probes and needles probe are used for RF signal coupling (both input and output) and DC connection, respectively. The probes are expected to be as few as possible since their mechanical arms occupy a lot of space on the test platform[4] and the system would get congested. Therefore the differential output from the proposed Gilbert cell is preferred to be converted to single ended one. This was done by the Differential to Single (D2S) circuit, as shown in Fig. 4.19. It is basically a single stage amplifier followed by an output buffer. Since the signal has been downconverted to IF by the mixer, the design constraint on the bandwidth of D2S has been relaxed. The tail current for the differential pair M_1 and M_2 in Fig. 4.19 can be chosen smaller than that in the LO driver, 3 mA for this design. M_5 and M_6 form source follower which provide 50 Ω impedance match to the measurement instruments. Note that the source follower output impedance can be roughly written as $1/g_m$, to have more gain at the output resistor R_L, it is expected g_m to be chosen as large as possible because the source follower gives a gain of $g_m R_L/(1+g_m R_L)$. On the other hand, to avoid distortion at the output, DC operating point should be close to the mid of power supply, which results a large DC current through M_5, M_6 and R_L.

For the mixer in RF1[5], however, some design issues in the D2S circuit were discovered after the tape out. First, the load resistor R can be replaced by PMOS transistors which saves chip area; secondly, for the D2S output in Fig. 4.19 (the load R_L is 50 Ω to emulate the input impedance of the measurement instruments, i. e., spectrum analyser), it must be DC coupled to the spectrum analyser, otherwise, there is no DC current supply for M_5 and M_6. But most of the spectrum analyzers have DC block at the RF input port for DC protection. In the tape out of the optical front end (tape out series No. Atto1b), these issues were removed and the D2S circuit was re-designed.

---

[4] The probe station used for the measurement is not a commercial one, instead, it was built up manually in the EEE Department, University of Nottingham.

[5] RF1 is the MPW (Multi-Project Wafer) tape out series number, which includes prototype circuits for RF applications, i.e., 5 GHz frequency divider for frequency synthesizer; 4 GHz LNA for breast cancer imaging system, and 5 GHz mixer for the high speed optical front end.

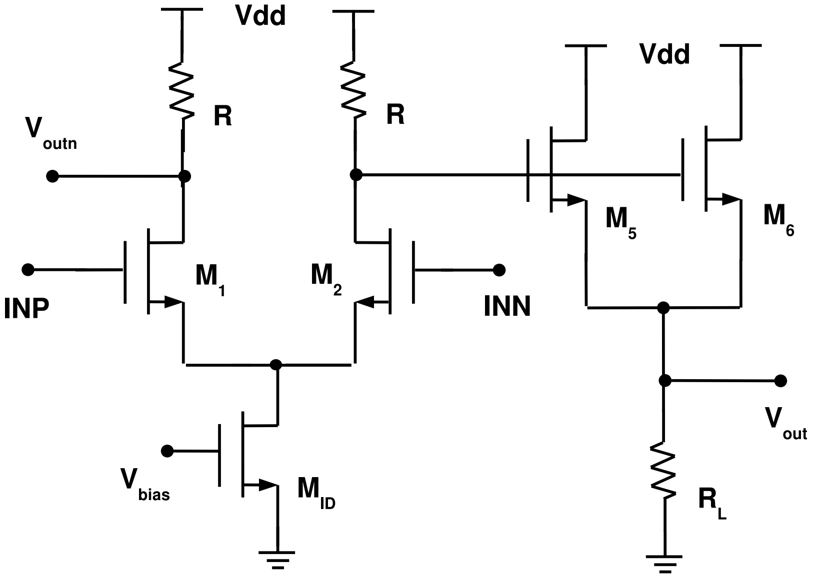


Figure 4.19: Schematic of the Differential to Single circuit

component	design parameter	Design value
$M_{1,2}$	$20 \times 5/0.35\mu m$	
$M_{5,6}$	$40 \times 5/0.35\mu m$	
M_{ID}	$12 \times 100/3\mu m$	
R	$4 \times 120/3\mu m$	2182 Ω /4

Table 4.2: Design values of the LO driver

The simulation result of D2S is shown in Fig. 4.20. As the IF frequency is around 100 MHz, D2S provides about 6.9 dB gain. These should be deducted from the mixer measurement results to get the net gain of the mixer core. The simulation results did not take into account of the effects of parasitic capacitances, i.e., that from the pads and wirings. Its - 3 dB bandwidth is estimated to be in fact about few hundred MHz, which means that if higher IF frequency is chosen, more current needs to be assigned to the D2S circuit to improve its bandwidth.

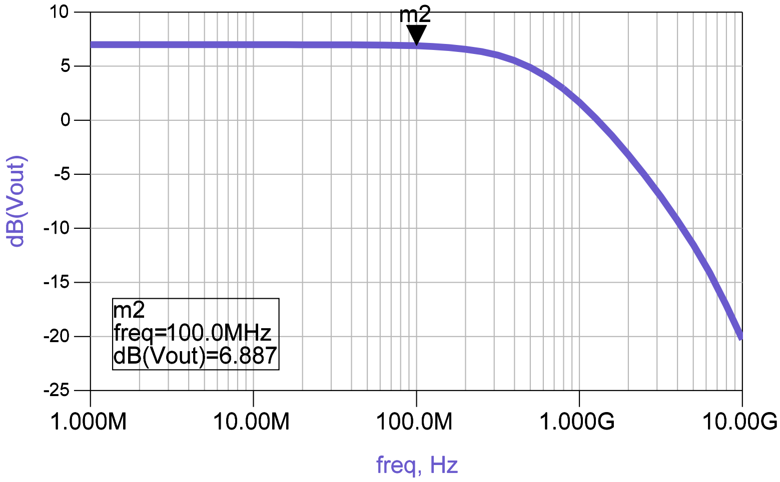


Figure 4.20: AC simulation result of the D2S

4.3.4 Mixer core

The mixer core in RF1[6] is a Gilbert cell as shown in Fig. 4.21. M_{IB} is the current source of the mixer which determines the DC current flowing through the M_1~M_6 and load resistors. By selecting appropriate V_{bias}, the current through M_1 and M_2 are chosen to be around 5 mA which balances the f_{max} requirement of M_1, M_2 and the power consumption of the mixer. It is noted that the DC operating points of the circuit are also effected by V_{bias} since load resistors consume a voltage headroom of $1/2 \cdot R \cdot I_B$.

RF transconductance stage consists of transistors M_1 and M_2. From Gilbert mixer noise expression (4.26), M_1 and M_2 are chosen to have the maximum size $40 \times 5\mu m/0.35\ \mu m$. From another point of view, large sizes of M_1 and M_2 are also beneficial for conversion gain, $2/\pi \cdot g_m \cdot R_L$. Although compared to smaller NMOS transistors, the f_{max} of M_1 and M_2 degrades because of the decreased current density, simulation shows that the mixer works well up to 5 GHz.

---

[6] RF1 is the tape out series number.

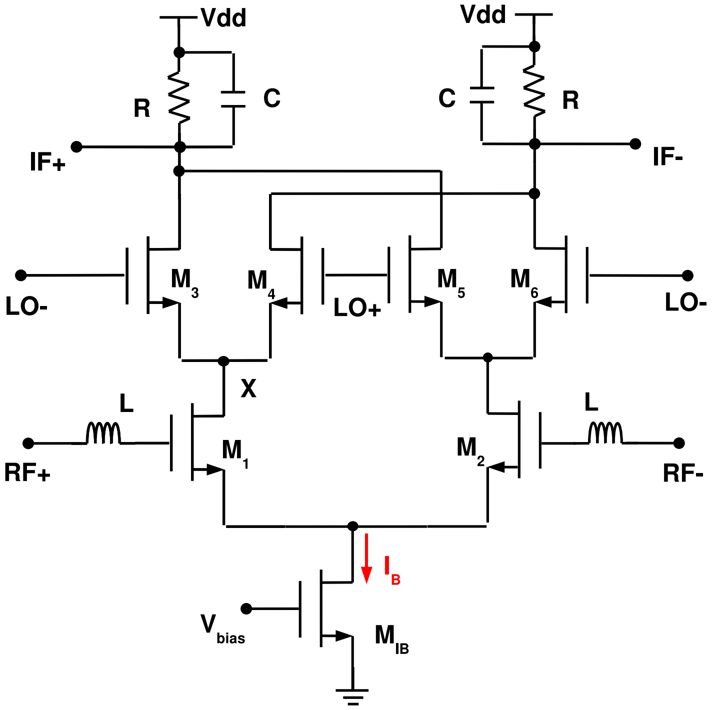


Figure 4.21: Schematic of the mixer core

Trade-offs exist for the size of switching quad $M_3 \sim M_6$. As mentioned previously about the noise of Gilbert cell in section 4.2 , short switch-on time is preferable to shorten the time of both ON state of the switching pair. This means a reasonable large size of switching quad. On the contrary, large $M_3 \sim M_6$ introduce large parasitic capacitance at the common source node (i. e., node X in Fig. 4.21), which causes a rectifying effect at this node, after downconversion by the noisy LO, it leads to a noise current at the mixer output [96]. The rough physical model demonstrates a Gilbert mixer noise expression (4.26) independent of switching quad size. Finally, from the point of LO driving, $M_3 \sim M_6$ should be sized relatively small since 4 large transistors (quad) tend to be heavy loads of the driving circuits, resulting more power consumption.

Poly2 resistors instead of PMOS transistors are employed as the load of the Gilbert mixer. This is for: 1) poly resistor demonstrates better linearity than PMOS transistor; 2) PMOS transistor has a larger resistance variation at the process corners which could be problematic to the DC operating points of the mixer.

Because of the current density limit of poly2 (0.3 $mA/\mu m$), resistor R in Fig. 4.21 was implemented with 8 poly2 resistors in parallel. This consumes more chip area and causes much larger parasitic capacitances at the output ports. But the later is not a problem because the signal frequency at output ports has been downconverted, the parasitic capacitances can be absorbed into the low pass filter, to filter out the LO feedthrough. Here capacitor C, as part of the low pass filter, was implemented by poly-poly capacitor. The design values for the components in the mixer core are listed in table 4.3.

component	design parameter	design value
$M_{1,2}$	$40 \times 5/0.35\mu m$	
$M_{3,4,5,6}$	$20 \times 5/0.35\mu m$	
M_{IB}	$40 \times 5/0.35\mu m$	
R	$8 \times 120/3\mu m$	2182Ω/8
L	3.7 nH	3.7 nH
$C_{ac\ grounded}$	$33.9 \times 33.9\mu m$	1 pF

Table 4.3: Design values of the mixer core

All the pre-layout simulations of the mixer were conducted in ADS. For large signal mode, nonlinear circuits and systems, frequency-domain based Harmonic Balance offers lot of benefits over conventional time-domain transient analysis [99]. The post-layout simulations of the mixer were done in Cadence, but only with several periods of transient simulation results which aimed to verify the functionality of the mixer[7].

Fig. 4.22 and Fig. 4.23 show the simulated conversion gain of the mixer at 5 GHz with IF at 100 MHz. The mixer in RF1 gives a total gain (including the 6.9 dB gain of D2S) around 18 dB, with 11 dB contributed from the mixer core (Gilbert cell). As expected when the input RF power increases to a certain level, the output drops and the mixer runs out of the linear region. The RF sweep simulation results indicate a 1 dB compression point of around -15 dBm for the mixer.

---

[7] It is quite time-tedious to run transient simulation for the post layout circuits which contain huge number of extracted parasitics.

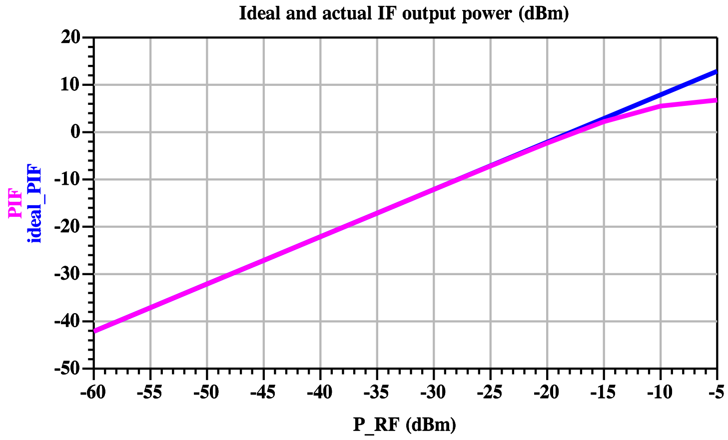


Figure 4.22: Mixer IF output vs RF input power at 5 GHz RF input and 100 MHz IF.

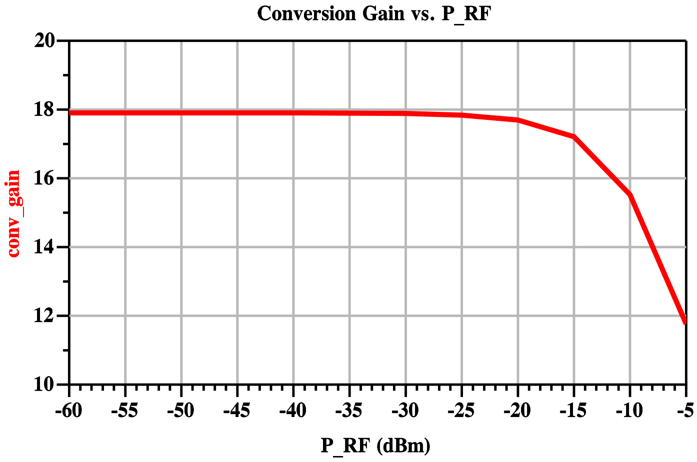


Figure 4.23: Mixer conversion gain vs RF input power at 5 GHz RF input and 100 MHz IF.

Fig. 4.24 shows the simulated noise figure and conversion gain of the mixer versus input LO power. In general, the noise figure of the mixer decreases as the LO power is increased to a certain level. This is easy to understand since large LO power reduces the simultaneous ON states of the switching quad, and thus reduces the noise contributions from them, as mentioned in section 4.2. However the further increase of the LO amplitude leads to more LO coupled noise to the common source node of the switching quad through the gate-source capacitors C_{gs} of the switching quad, which deteriorates the noise performance. Similar situation happens to the conversion gain, it follows the changes of LO amplitude initially, but gain drops after a certain value of LO power. Besides the feedthrough effect of gate-source capacitors C_{gs}, simulation shows that the increased imbalance between the two outputs from LO driver also makes things worse when LO power increases.

Because of the amplification of LO driver, small power can be applied to the mixer circuit, which saves power consumption of the system. An optimized value of -5 dBm for LO power can be found from Fig. 4.24, at which point the mixer demonstrates a maximum conversion gain of 17.8 dB and a minimum noise figure of 12.6 dB.

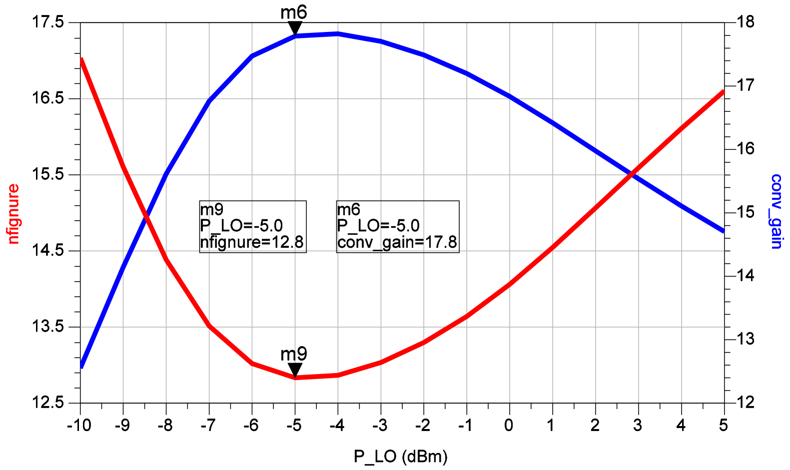


Figure 4.24: Simulated mixer noise figure and conversion gain vs LO power. RF=5 GHz, IF=100 MHz.

Fig. 4.25 shows the simulated output power of the mixer with 2 tones RF input. The frequencies of the two RF input tones are 5.001 GHz and 4.999 GHz, respectively. With 2 MHz frequency separation, the fundamental tone and the third order tone at the output port of the mixer are plotted in Fig. 4.25[8], from this diagram the output third order interception point (OIP3) and input third order interception point (IIP3) of the mixer can be calculated as

$$\begin{aligned} OIP3 &= (P_{fd} - P_{3rd})/2 + P_{fd} \\ IIP3 &= OIP3 - G_{mix} \end{aligned} \tag{4.30}$$

where P_{fd} stands for the power of the fundamental tone at the mixer output, P_{3rd} denotes the output power of the third order distortion at the mixer output, it can

---

[8] In fact, there are two sidebands for the third order harmonics, in Fig. 4.25 only the lower side band is plotted.

be lower side band or upper side band, G_{mix} is the mixer conversion gain.

Fig. 4.26 shows the simulated mixer conversion gain and IIP3 versus the LO amplitude. For mixers used in wireless communication systems where the RF signal can vary from under -100 dBm to around 0 dBm, IIP3 of the mixer is a very important performance parameter, which describes the linearity of the mixer, and generally the linearity of the system. For the proposed mixer, it is seen that -5 dBm, the optimized LO input power for mixer conversion gain, is nearly the worst case for IIP3 of the mixer. In other words, by setting LO input power to be -5 dBm, the mixer is very likely to get corrupted by harmonics at large RF input. Fortunately, it is not a problem for this application since the optical front-end would only need to handle the input optical power on the order of μW or smaller, which translates to a voltage of no more than 10 μV at the TIA output. For this input power, the mixer is still far away from saturated. Moreover, there is no in band or out band inter modulation for our application, since the desired input signal frequencies are discrete and well separated.

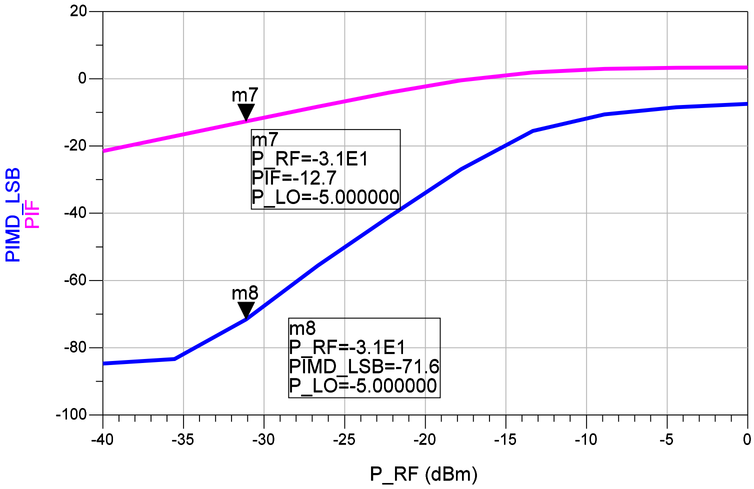


Figure 4.25: Simulation of Input IP3 of the mixer vs input RF power

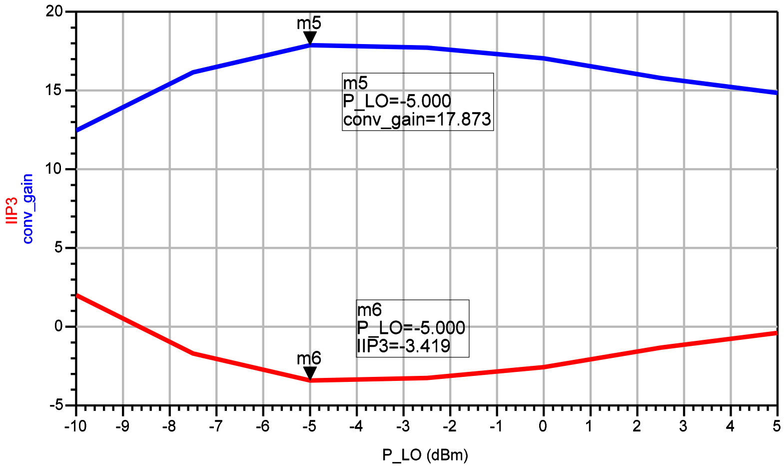


Figure 4.26: Simulation result of input IP3 and conversion gain of the mixer vs LO power

Monolithic integration of analog circuits provides opportunities for good device matching and isothermal operation, but it also comes with the coupling and cross talk problems. The capacitive coupling to the substrate and resistive coupling through the substrate is the dominant on-chip effect of the LO-to-RF feedthrough because bonding wires and packet leads are excluded in the design. Layout issues play an important role in these considerations. Generally the RF and LO signal routings should be kept well apart and should be orthogonal to each other in order to prevent any on-chip magnetic coupling between the lines. For this design, the magnetic on-chip coupling can be considered almost insignificant, the RF and LO signals were not kept orthogonal but kept well separated. The LO switching devices were placed into their own well surrounded by a grounded guard-ring, and the transistors of the driver stage into another. The on-chip LO signal routing was also shielded from the substrate. This can be done in P^{-} substrate by placing a n-well underneath the LO wires. By connecting the n-well to the positive supply-voltage, the reverse biased PN junction provides a proper shield.

The layout of the mixer is shown in Fig. 4.27. The chip has a size of $1127\mu m \times 915\mu m$, while most of the chip area are occupied by pads, decoupling capacitors, and inductors.

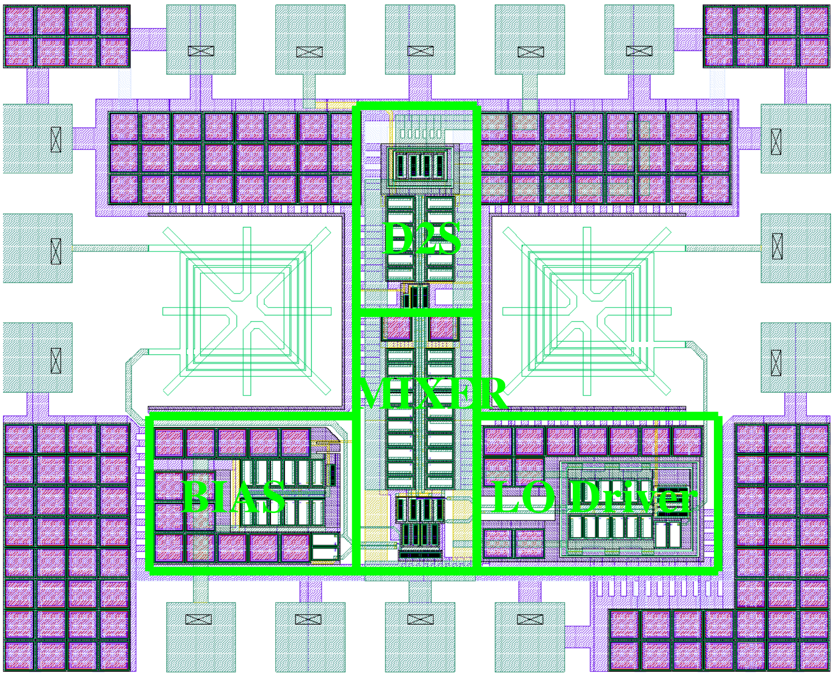


Figure 4.27: Layout of the mixer in RF1

4.4 Measurement

4.4.1 Mixer measurement set up

The mixer measurement setup is shown in Fig. 4.28. The die micrograph is shown in Fig. 4.29. Anritsu 37397D VNA is used as RF signal generator while Agilent 40 GHz signal generator E8257C generates 5 GHz LO signal for the mixer. A bias Tee was inserted between Agilent signal generator and mixer LO input to set the DC for LO ports. In the experiment, two needle probes were used to supply the power of the chip since the whole chip consumes a total current of over 40 mA (10 mA for mixer core), including the power consumption of auxiliary circuits such as biasing, LO driver, D2S, and output buffer. Again, spectrum analyser MS2721A was used to measure the output power from the mixer.

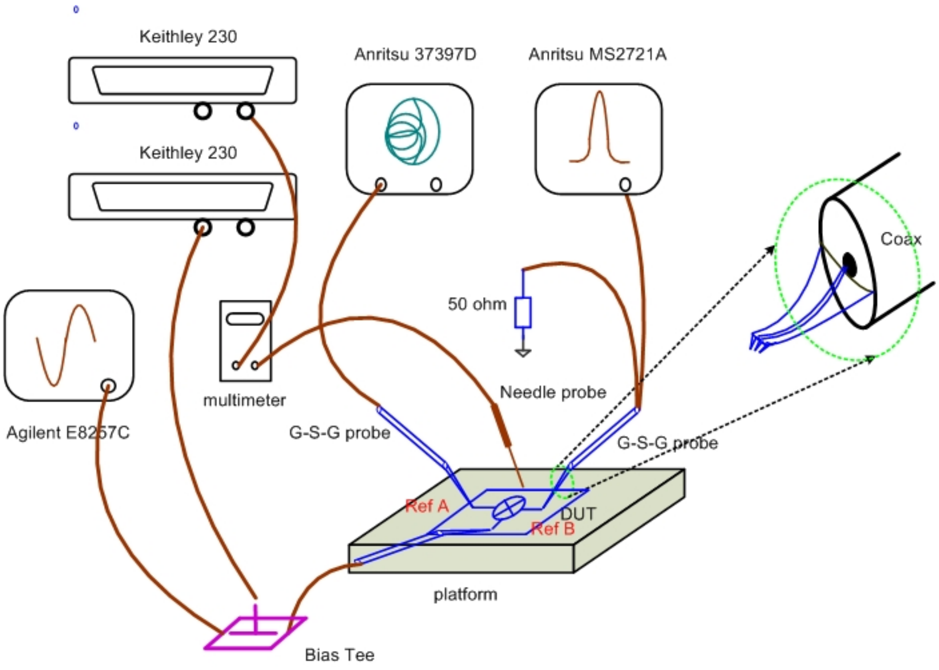


Figure 4.28: Mixer measurement setup

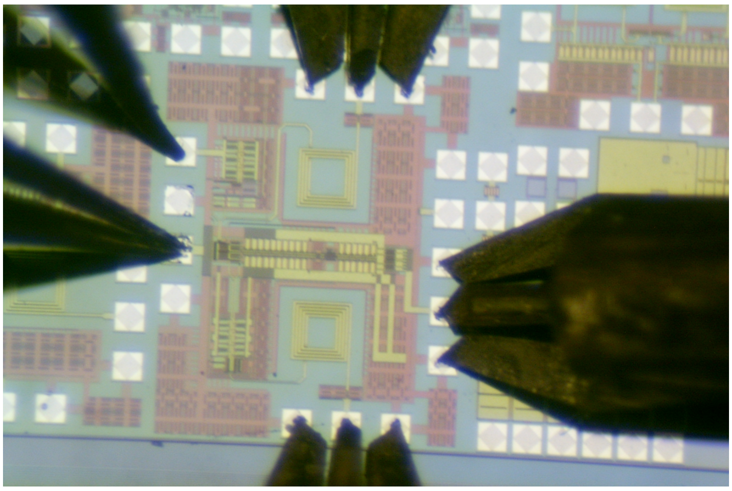


Figure 4.29: RF1 mixer die micro graph

It should be noted here that because of the mentioned DC coupling problem of

D2S circuit (section 4.3.3) , a 50 Ω terminator was connected to the output of the D2S (which is also the output of the mixer sub circuit) before it was input to the spectrum analyser. With this redeemed setup, calibration has to be made to compensate the loss power in the 50 Ω terminator. Notice that the source follower gives a voltage gain of $g_m R_L/(1+g_m R_L)$. Now R_L changes from 50 Ω to $50\Omega/2 = 25\Omega$, if the output impedance of source follower $1/g_m$ is 50 Ω, we can compute the power loss of this setup is

$$\begin{aligned} P_{loss} &= 10 \cdot log[(\frac{25}{50+25})^2/(\frac{50}{50+50})^2] \\ &= 3.5\, dB \end{aligned} \tag{4.31}$$

Simulation results gives a power loss of around 3 dB for the redeemed 50 Ω terminator setup, which is close to the above hand calculation. We conclude that the measured gain of the mixer should be increased by 3 dB to compensate the loss of the terminator[9].

Calibration of the mixer measurement is different from that of TIA measurement as auto calibration can be applied to the later. For mixer test set up, besides the mentioned power loss on the 50 Ω terminator, the power loss of cables, adapters and probes must be taken into account. In Fig. 4.28, the RF input power calibration reference plane should be set at the end of RF input probe (Ref A in Fig. 4.28), while the output power calibration reference plane should be set at Ref B. No calibration was applied to the LO input power, the LO input power and DC were set to give an optimized mixer performance, as described in section .

4.4.2 Conversion gain measurement

As mentioned earlier, the prototype mixer consumes over 40 mA current because of the auxiliary circuits inside the chip. From simulation results, a DC current of

---

[9] The simulation results in section 4.4.1 give the gain of the mixer core and mixer subcircuit which assumes a DC 50 Ω load. They do not count in the 50 Ω terminator loss when the output is AC coupled to the spectrum analyser.

10 mA comes from the mixer core; 5mA from LO driver; 3 mA from D2S and 25 mA from output buffer, which gives 43 mA in total for typical process parameters. From the measurement, however, the chip consumes a DC current ranging from 40 mA to 42 mA depending on the needle probe position and stress. This could be caused by the probe contact resistance for the power supply, especially when a large current like 40 mA flows through the needle probe. During the measurement, we have managed to land two needle probes onto a single pad (vdd) to partially solve the large current problem. Because of the same problem, all the measurement results shown in this section are taken under the condition of a power supply of 3.3 V and no measurement results dependent on power supply variation were conducted for the mixer. Without notification, the LO DC is set to 2.5 V and LO power is 0 dBm, RF DC is equal to 1.5 V.

Fig. 4.30 shows the measured mixer input reflection coefficient S11. This was done when LO was applied with an input power of 0 dBm at 5 GHz and the DC equal to 2.5 V. The point here is the mixer input S11 needs be measured under active switching state, which reflects the real operation condition of the mixer. It can be seen from Fig. 4.30 that the RF port of mixer has a S11 lower than - 10 dB at the frequency range of 4.5 GHz to 5.5 GHz.

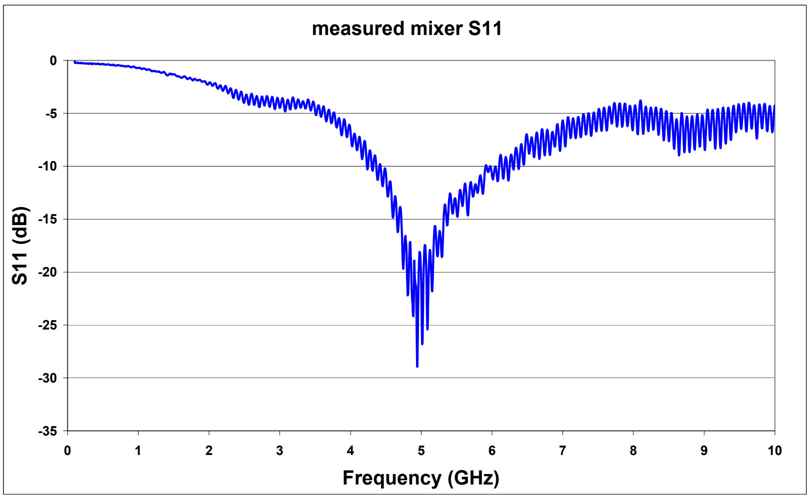


Figure 4.30: Measured mixer S11 at RF input port

Fig. 4.31 shows the measured mixer conversion gain at a fixed IF=100 MHz while LO is equal to 4.9 GHz and 5 GHz, respectively. The measured gain is about 16 dB including the gain from the D2S stage, and the power loss on the 50 Ω DC terminator was taken into account as well. No significant gain difference was observed between LO equal to 4.9 GHz and 5 GHz, which means both the RF port and LO port have a tolerable matching sensitivity. Moreover, it can be seen from Fig. 4.31 that the input 1 dB compression point of the mixer is about -16 dBm, very close to -15 dBm from the simulation results.

Roughly, the measured mixer conversion gain is 1.8 dB lower than simulation simulation results, as shown in Fig. 4.32. One reason for the gain loss could come from the output probe in the measurement. All the on chip G-S-G pads, including at RF port, LO port, IF output port, have a separation distance of 150 μm, while only two 150 μm separation G-S-G probes from picoprobe Ltd were available in the lab. The compromise is using a 100 μm separation G-S-G probe from cascade microtech Ltd instead at the IF output port during the measurement, which just touches the edge of the Ground pad. Another reason for that is the IF output port is less sensitive to the impedance matching comparing to RF port or LO port.

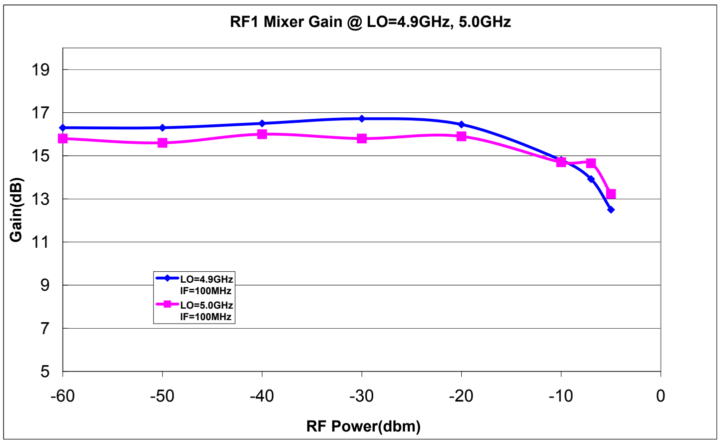


Figure 4.31: Measured mixer gain at IF=100MHz, LO=4.9GHz and 5.0GHz

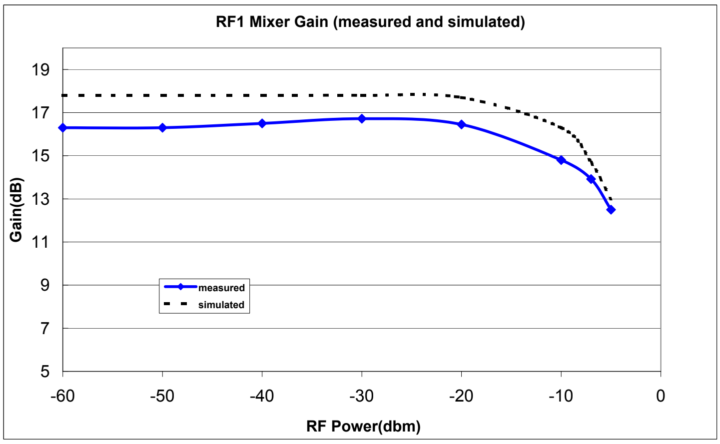


Figure 4.32: Measured and simulated mixer gain at IF=100MHz, LO=4.9GHz

Fig. 4.33 gives the mixer conversion gain versus the IF frequency while LO frequency is fixed at 4.9 GHz. For both cases of IF equal to 50 MHz and 100

MHz, The conversion gain is about 16 dB, but it drops to around 13 dB when IF frequency equals to 200 MHz. This is due to the bandwidth limitation from D2S circuit. With 3 mA current consumption, the D2S only has a -3 dB bandwidth of a few hundred MHz, as mention in section 4.3.3.

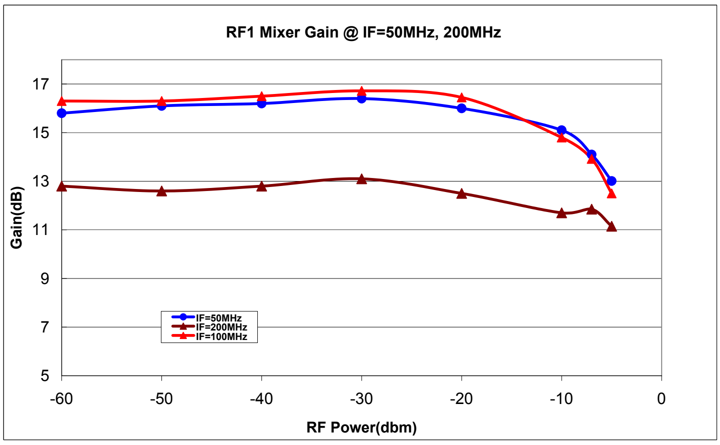


Figure 4.33: Measured mixer gain at IF=50MHz, 100MHz, and 200MHz (LO=4.9GHz)

4.4.3 Linearity and noise measurement

A mixer's linearity is generally characterized by two parameters, namely, P_{1dB}, -1 dB compression point and IIP_3, Input 3rd order Interception Point (IIP3). The former is quoted at large signal mode while the later is used at small signal level. The measurement set up for IIP_3 is almost the same as the setup of gain measurement, the only difference is a power combiner is introduced at the RF input which synthesize the two tone input signal for the mixer RF input. Then at the output port, both the output power for the fundamental tone and third order harmonics were recorded. The mixer's IIP_3 can then be extrapolated from the diagram of input-output power curves, as shown in Fig. 4.34. The fundamental output power increases proportionally with the input power, while the third order harmonics increases 3 times faster than the fundamental one. By extrapolating the two lines the mixer's IIP_3 can be found.

Two-tone input signal to the mixer's RF port was set to have the discrete frequencies of 4.8 GHz and 4.805 GHz, with equal RF power sweeping from -50 dBm to around -10 dBm. The LO frequency is fixed at 4.9 GHz. Fig. 4.34 shows a measured IIP_3 of 1.5 dBm for the RF1 mixer. Compared to a simulated value of 0 dBm, it is higher and better. One of the important reason causing the improved IIP_3 of the mixer is of course the measured gain degradation compared to simulation results.

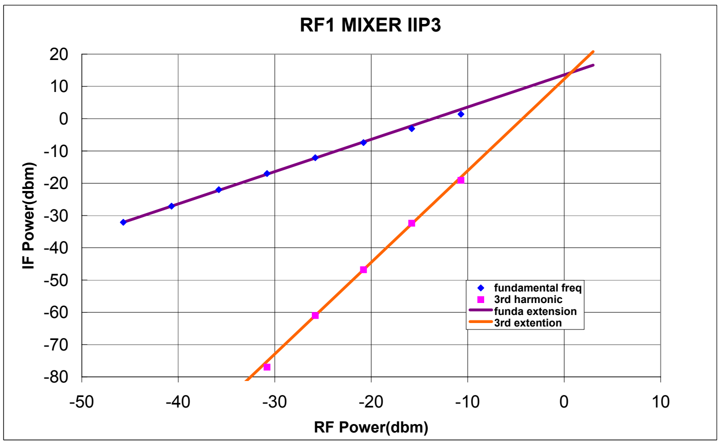


Figure 4.34: Measured mixer's fundamental and 3rd harmonics output power at two-tone RF input at 4.8 GHz and 4.805 GHz while LO=4.9 GHz. IIP3 can be extrapolated to be about 1.5 dBm.

Mixers are a noisy circuitry in receiver chain, especially for active mixers, which includes more active components (noise sources), compared to passive mixers. For noise measurement of RF circuits, generally there are 3 kinds of methodologies. The first one and the most straight forward one is to use a noise figure meter, which generates a high voltage to drive the noise source (i.e., NC346B from Anritsu). Since the input noise and ENR (excessive noise ratio) is known to the noise figure analyser, the noise figure of the DUT can be displayed directly. However, noise figure meter generally has frequency limitations, moreover, it is no accurate when measuring high noise figures such as mixers with noise figure beyond 10 dB. It is more suitable for low noise amplifier measuring.

The second one for noise measuring is so called gain method, which is based on the noise factor definition:

$$NF = \frac{P_{n,out}}{P_{n,odi}}$$

where $P_{n,out}$ denotes the total output noise power; $P_{n,odi}$ denotes the output noise power due to input source only. In this definition, "noise" is due to two effects, one comes from the noise sources inside the circuitry, which could be from transistors, resistors, etc.;another comes from the input sources and the noise power is amplified by the gain at the output port of the devices, generally this kind of noise is treated as thermal equivalent and has a broadband characteristics, its available noise power is

$$P_{n,ava} = kT \cdot \triangle f$$

NF is defined at temperature of 290 K, so the available noise power density from the input source is :

$$P_{nD,ava} = kT = -174\, dBm/Hz$$

and the noise figure of DUT is

$$\begin{aligned} NF &= \frac{P_{n,out}}{P_{nD,ava} \cdot B} \\ &= P_{n,out} - (-174\, dBm/Hz + 20 \cdot log(BW) + Gain) \\ &= P_{out} + 174\, dBm/Hz - Gain \qquad (4.32) \end{aligned}$$

The problem associated with this method is that the measurement results strongly depend on the power meter's accuracy, which generally requires a well calibrated power meter with very low noise floor. In practice, the gain method is only used to give a rough noise figure estimate of a high gain system, such as receiver with 80 dB gain.

Y factor method is another popular way to measure the noise figure. To use the Y factor method, an ENR noise source is connected at the input of the DUT. The noise source is generally a calibrated diode which provides different levels of shot noise at different biased voltage. The noise figure at specified frequency is set and generally provided in a table with the product. Turning the noise source on and off, a change in the output noise power density, called Y, can be measured by the spectrum analyser, and the ENR is specified as

$$ENR = \frac{T_h - 290K}{290K} \tag{4.33}$$

where T_h is the noise temperature of the noise source when it is turned on; 290K is the standard temperature of noise figure definition. On the other hand, we have

$$Y = \frac{G \cdot (T_h + T_n)}{G \cdot (290K + T_n)} \tag{4.34}$$

where T_n is the noise temperature of the DUT (Device Under Test) and G is the gain of it. Using noise factor definition, $NF = (1 + T_n/290K)$, we have

$$NF = ENR - 10 \cdot log(10^{Y/10} - 1) \tag{4.35}$$

where the recorded power difference Y should be in dB. The Y method for noise figure measuring is more accurate compared to gain method since here the difference of the output noise power is measured instead of the absolute noise power value. And the measuring procedure is simple, but Y method is generally used to measure the devices which have the noise figure in the range of $(ENR \pm 10\ dB)$, since for large NF devices, Y is close to 0 dB, which means no clear difference of output noise power can be observed; for small NF devices, Y is on the order of several tens dB, causing measuring difficulties as well.

For RF1 mixer, the Y method were chosen to characterize its noise performance. MS2721A was used to record the noise power difference Y; noise source is Anritsu NC346B, which has a typical ENR of 15 dB at the frequency of 5 GHz. During the measurement, the LO input DC is set as 2.5 V and the power is set to be 0 dBm at 5 GHz.

It is worth noting here that the RBW and VBW of MS2721A should be set correctly during the measurement, otherwise no clear power difference can be observed. For this measurement, RBW=1MHz, VBW=3KHz, and spectrum span was 10 MHz. A typical value of 2.9 dB for Y was observed in the measurement, which gives a NF of 15.7 dB based on equation (4.35).

From simulation results the implemented Gilbert cell mixer gives a noise figure of 12.6 dB at LO input power of -5 dBm. Correspondingly, a maximum conversion gain of 17.8 dB is achieved at the same LO input power. But in the experiment, an optimized LO input power for maximum conversion gain is around 0 dBm, which differs from that of simulated. The measured Noise Figure of 15.7 dB differs from simulated 12.6 dB as well, the reason for this difference can be caused by LO port matching, the amplitude and phase imbalance at the output of LO driver and the process variation.

Chapter 5

Optical Front end

5.1 Review of CMOS Optical Receiver

Optical receivers with photodetectors integrated in a standard CMOS technology are attractive due to the advantages of low cost, low power consumption, high reliability and potential for large scale integration. However, it is not an easy task to implement high speed photodetector in CMOS process due to the slow photo-generated diffusive carriers caused by relatively long optical absorption length of silicon (about 14 μm at incident light wavelength of 850 nm) [42], as mentioned in chapter 2. For high speed CMOS optical receivers, high speed photodiodes, transimpedance amplifiers (TIA), Limiting Amplifiers (LA) or broad band mixers must be monolithically integrated together, which brings much more challenges than fabricating a fast CMOS photodiode or TIA alone. Under the driving of huge market from long haul, short range optical communications and optical sensing systems (such as in this thesis), numerous designs and solutions have been proposed in literature. An overview of some important recently published results are given in table 5.1.

Table 5.1: Comparison of published fully integrated optical receivers. (* refers to analog bandwidth instead of data rate. **Ge-SOI PD bonded on 0.13-μm CMOS)

Author	Technology	B.W. * (GHz)	sensitivity (dBm)	PD. area. (um^2)	λ (nm)
2001, Woodward [42]	0.35-μm CMOS	< 1	-6.3	274	–
2005, Swoboda [54]	0.6-μm BiCMOS	2.4	-20.5	2070	660
2005, Radovanovic [100]	0.18-μm CMOS	1.5	-19	2500	850
2006, Schow [101]	SOI & CMOS**	6.6	-7.4	100	850
2006, Swoboda [102]	0.5-μm BiCMOS	7.7	-10.8	2500	850
2006, Hermans [103]	0.18-μm CMOS	0.5	-8	6400	850
2005, Jutzi [104]	0.18-μm CMOS	1.1	-8	–	850
This work	0.35-μm CMOS	**4.7**	–	2500	850

In [61], polysilicon is used to implement the photodiode due to its carrier's shorter life time and larger diffusion bandwidth. However, the polysilicon photodiode suffers from the poor responsivity, i.e., a typical value of 1.2 mA/W was reported in [61]. [61] described a new method of high speed photodiode implementation in CMOS, but it was only about the photodiode and not the whole receiver. In [100], the frequency response of n-well p^- substrate photodiode in CMOS process is compensated by an analog equalizer which is placed after the TIA, and a bit rate of 3 Gb/s (-3 dB bandwidth of 1.5 GHz) was achieved for the optical front-end which integrated photodiode, preamplifier, equalizer and output buffer in a 0.18-μm CMOS. The optical receiver demonstrates a sensitivity of -19 dBm with a bit error rate (BER) < 10^{-11}. The power consumption of the system is 50 mW. However, the problem with this optical front-end is that the equalizer only works for small input signals which limits its applications [61].

In [103] and [104], a different way to effectively cancel the effect of slowly diffusive carriers in fully standard CMOS is presented[1]. It has a fully-differential interdigitated structure with n-well p^- substrate photodiodes alternatively shielded

---

[1] section 2.4 gives more details about the so named Spatial Modulated Light (SML) CMOS photodetector.

and exposed. By subtracting the response of the shielded diodes and exposed ones, it eliminates the slow carriers and achieves a high speed rate. Photodetector, transimpedance amplifier, post-amplifier and output buffer were integrated in a 0.18-μm CMOS in [103]. Although a BER of $5 \cdot 10^{-12}$ at the Bit rate of 3.5 Gbit/s with an input V_{pp} of 10mV was measured for the post-amplifier with output buffer, the whole system only demonstrates a BER of $3 \cdot 10^{-10}$ at the Bit rate of 500 Mbit/s with an input V_{pp} of 10mV in [103]. The reported power consumption is 143 mW. In [104], a responsivity of 0.02 A/W and a -3 dB bandwidth of 1.1 GHz was measured for the spatial modulated light photodiode. The photodiode together with transimpedance amplifier, limiting amplifier and output buffer gives an input sensitivity of -8 dBm at the Bit rate of 2Gb/s and BER of 10^{-9}. The reported optical receiver, which was fabricated in an unmodified 0.18-μm CMOS process, has a total power consumption of 50 mW from a power supply of 3.3 V. It was specially noted in the paper that the 3.3 V was only for photodiode reverse bias and input output (IO) MOSFETs. The core MOSFETs in the circuits were not operated from 3.3 V in order to prevent breakdown.

T. K. Woodward in [42] reviewed the monolithically integrated photodetectors, and reported his own work of a photodiode in CMOS technology without any process modification. N-well was used as a screen terminal to block the slow diffusive carriers generated in the substrate. In [42], the p^+ diffusion, which formed the anode of the photodiode, were in parallel in the centre area of n-well. The n^+ diffusion inside the n-well, which formed the cathode of the photodiode, on the other hand, were only made along the boundary of n-well for well contact[2]. The reported photoreceiver integrated the p^+ diffusion to n-well photodiode and a simple three-stage inverter like TIA. The entire circuit was fabricated in Lucent 0.35-μm CMOS. Measurement revealed an input sensitivity of -6.3 dBm at the bit rate of 1 Gb/s with a BER of 10^{-9}. The measurement on the p^+ to n-well photodiode alone showed a responsivity of 0.04 A/W under a reverse bias voltage of 10 V. The reported power consumption for the receiver is 6 mW from 3.3 V power supply.

The p^+ to n-well CMOS photodiode in [42] is an attractive way to implement high

---

[2] the photodiode reported in [42] has a square shape with a size of $16.54\mu m \times 16.54\mu m$.

speed photodiode for its simplicity, but there are still lot of design considerations. One of the potential risks is, the photogenerated carriers inside the n-well still demonstrate diffusive process because of the far aparted electrodes. This problem was solved in this thesis by interdigitated the p^+, n^+ diffusion with a separation width of s as shown in Fig. 2.10. The interdigitation of p^+ and n^+ diffusion is important since it reduces the collection time of the diffusive carriers from inside n-well to the electrodes of the photodiode. Moreover, it minimizes the photodiode's series resistance which could be problematic in the presence of parasitic capacitances of the photodiode.

Silicon-on-insulator (SOI) and BiCMOS technology have advantages in implementation of optical receivers due to their intrinsic insulator layer or buried layers which can be readily used to stop slow diffusive carriers. -3 dB bandwidth of 6.6 GHz and 7.7 GHz optical receivers were addressed in [101] and [102].

In [101], a Ge-on-SOI photodiode was reported, which uses the 350 nm thick, single-crystal Ge as the absorbing layer, sitting on a thin Si layer, with isolation layer underneath. The p^+ and n^+ diffusion are formed and interdigitated on Ge. Anti reflection coating are passivated on the top of the diode. The reported photodiode has an active area of only $10\mu m \times 10\mu m$. The transimpedance amplifier, limiting amplifier and output buffer were fabricated in a standard eight-metal-layer $0.13\mu m$ CMOS process. The Ge-on-SOI photodiode was wire bonded to the TIA, LA, output buffer chip. The receiver exhibits an input sensitivity of -7.4 dBm at bit rate of 15 Gb/s with a BER of 10^{-12}. The small transimpedance gain of the receiver is 91 dBΩ and -3 dB bandwidth is 6.6 GHz. The photodiode shows a responsivity of 0.32 A/W at $\lambda = 850$ nm and reverse bias voltage of 2 V. The power consumption of the receiver at 10 Gb/s is 30 mW from the power supply of 1.5 V.

In [102], a modified 0.5-μm BICMOS process with $f_T = 25$ GHz that offers a vertical PIN photodiode was used to implement the optical receiver. The PIN photodiode has a thickness of $10\mu m$ and is formed by P^+ diffusion, $N-$ intrinsic region (process modified) and N^+ buried layer. A minimum optical power of -8.9 dBm was detected at a bit rate of 11 Gb/s for a BER of 10^{-9}. The 3 dB cut-off frequency of the photodiode at a reverse voltage of 17 V and a wavelength of

850 *nm* was measured to be 2.2 GHz. The whole optical receiver has a power consumption of 310 mW from a power supply of 5 V.

However, these stand-alone implementations clearly can not be fully compatible with today's mainstream technology: CMOS. In our efforts of integration of photodiode in standard CMOS process, we developed a high speed optical front-end which integrated improved photodiode based on the approach mentioned previously in [42] of treating n-well as a screening terminal of slow photo-generated diffusive carriers. But before the above mentioned optical front-end, another initiate, novel idea of implementing optical receiver with a single-balanced mixer was first consolidated in CMOS. Section 5.2 described it in more details.

5.2 Mixer based Optical Front-end

5.2.1 System architecture

In an optical receiver, photodiodes convert the input light into electrical current, which is then converted to voltage by a transimpedance amplifier. If the transimpedance amplifier is followed by a mixer, the transconductance stage in the mixer (i. e., M_1 and M_2 in Fig. 4.21) converts the voltage back into current, which is finally commutated by the switching quad and signal frequency is mixed down to IF. Here it comes up with the new idea of how about directly using the photodiode as the transconductance stage of the mixer? In this way the conversions of light into current and then current into voltage can be omitted and circuits can be greatly simplified. An example of this kind of optical receiver based on single balanced mixer is shown in Fig. 5.1. PD is used as the current source of the single balanced mixer and LO are applied to the gates of transistor M_1 and M_2, the DC component in the incident light gives the bias current of the single balanced mixer while the high frequency components in the incident light are downconverted by the LO. The load of the mixer can be PMOS transistors or resistors which convert the current into voltage for post processing. The most attractive point of this optical front-end is its simplicity, which saves silicon area. Another advantage is its ultra low power consumption, which is determined by the DC component of the incident light. Here we assume that a CMOS n-well photodiode was used in the optical receiver with a responsivity of around 0.3 A/W for an input laser power of 600 μW at wavelength of 850 nm. It is further assumed that the coupling efficiency of the laser to the PD is 40% and the incident laser modulation depth is 5%, the DC current of the mixer then is written as

$$I_{dc} = 600\mu W \times 0.3A/W \times 40\% = 70\mu A \tag{5.1}$$

With all these advantages, this optical receiver does have drawbacks. First, it is difficult to achieve high speed. If we assume the mixer is loaded by small transistors or capacitors, the dominant pole of the circuit is located at common source node of the differential pair M_1 and M_2. The large parasitic capacitor of the photodiode shunts the high frequency signal to ground. Ignoring the body effect of MOSFET, the input impedance of M_1 and M_2 at their common source node can be roughly estimated as $1/g_m$, thus the RC time constant at the common source node is written as g_m/C. With 70 μA current, the g_m of the transistor is about 0.1 m, together with 0.6 pF photodiode capacitance, it gives a cut off frequency of lower than 300 MHz. It is noted here that increasing the size of the differential pair does NOT help much since the C_{gs} increases as well, which offsets the benefits of g_m increasing.

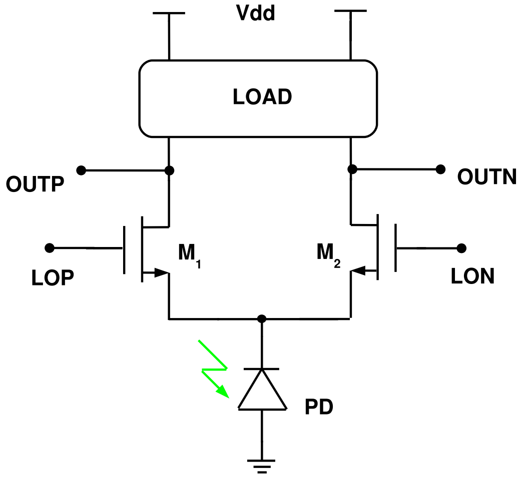


Figure 5.1: Block diagram of the mixer based optical front-end

Noise would be another problem for this kind of optical receiver. It is well known that for a cascade system, the noise factor of the whole system can be written as

$$NF_{total} = NF_1 + (NF_2 - 1)/A_1 + (NF_3 - 1)/A_1 \cdot A_2 + ... \tag{5.2}$$

where NF_i denotes the noise factor of the i_{th} stage, and NF_{total} is the equivalent noise factor of the cascade system. It is clear that the system's noise figure is dominated by the first few stages. To achieve a low noise figure system, the first stage of the system is most crucial, not only a low noise figure is required, it is generally desired to have enough gain for the purpose of noise suppression from following stages. For the optical receiver as in Fig. 5.1, the first stage provides no gain at all, which means that all the noise from the switching effects and the differential pair would be referred to the input directly.

5.2.2 PD and circuit design

The photodiode used for the mixer based optical front end is a n-well CMOS photodiode. This n-well photodiode has been silicon proved by the VLSI group in Electrical and Electronic Engineering Department. It gives a typical responsivity of 0.32 A/W and the size is $50\mu m \times 50\mu m$. The reason why n-well photodiode was chosen for mixer based optical front end was its superior responsivity to the $p+$ n-well photodiode. As calculated in section 5.2.1, with 600 uW light power, we can have 70 μA current for the DC biasing of the differential pair with n-well photodiode; if $p+$ n-well was used instead, only 3.5 μA DC current can be derived from the incident light assuming a responsivity of 0.016 A/W. It would be extremely difficult to operate the circuit at high frequency.

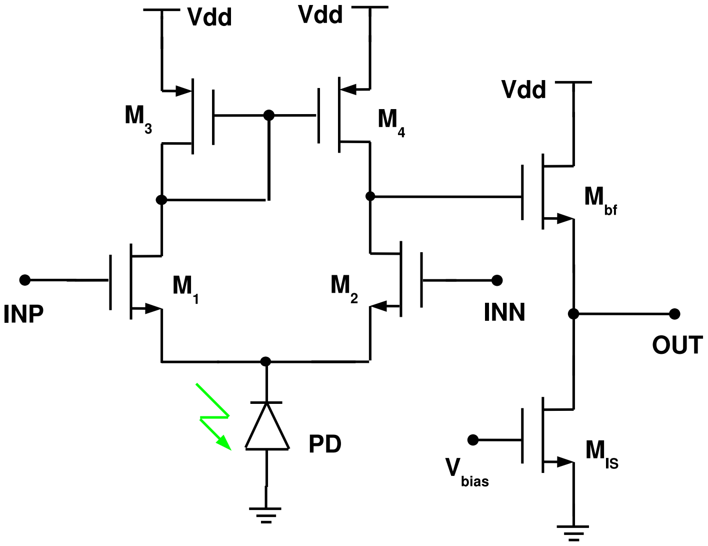


Figure 5.2: Schematic of the mixed based optical front end

The schematic of the single balanced mixer based optical front end is shown in Fig. 5.2. The photodiode converts the modulated light into current, which is then commutated by the differential pair. The modulation depth of the laser decides the AC magnitude of the RF input to the mixer, while the input laser power decides the DC biasing of the circuit. A current mirror was chosen to convert the differential current to a single ended voltage output. For pad driving purpose, a common gate output stage was placed after the single balanced mixer. The transimpedance gain of the optical front end can be written as

$$Z_T = r_{o2} \parallel r_{o4}$$

where r_{o2} is the output resistance of M_2 and r_{o4} is that of M_4. It can be seen that the transimpedance gain can be much higher than that of the transimpedance amplifier mentioned in chapter 3. But it comes with the sacrifices on bandwidth and noise. Since r_{o4} , the output resistance of M_4, not only decides the gain of the mixer, it imposes a pole at the output node of the mixer that can not be ignored in the transfer function of the optical front end. This pole could exceed the RC time constant at the common source node of the switching pair and become the dominate one of the front end if the output buffer has a big size. The output buffer has a self biased circuit on chip to simplify the measurements. Fig. 5.3 gives the simulated frequency response of the optical front end. A -3 dB bandwidth of around 200 MHz was achieved with a fixed IF frequency of 300 KHz. Fig. 5.4 shows a simulated transient response of the optical front end at a DC of 100 nA and modulation depth of 10%. The LO and IF frequency are 5 MHz and 300 KHz, respectively. The design values for all the devices are listed in table 5.2.

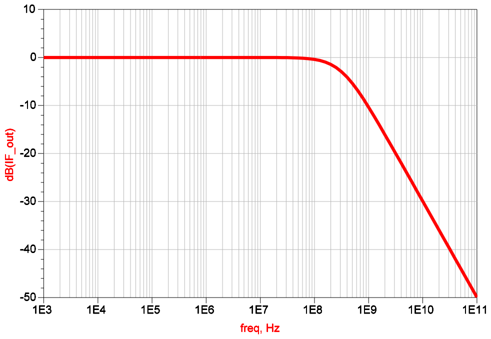


Figure 5.3: Simulated bandwidth of the mixed based optical front end with a DC=100nA.

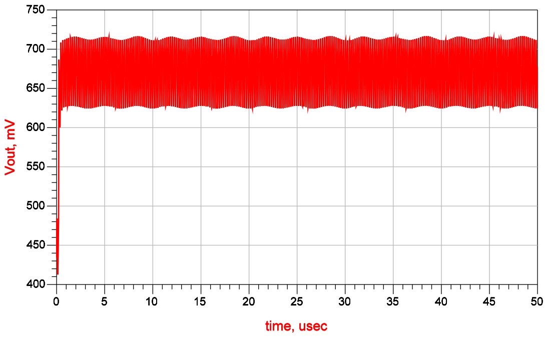


Figure 5.4: Transient simulation on the mixed based optical front end with a DC=100 nA and a modulation depth of 10%. LO= 5 MHz, IF=300 KHz.

devices	design parameter	design value
PD	$50\mu m \times 50\mu m$	
$M_{1,2}$	$5\mu m/1\mu m$	
$M_{3,4}$	$2\mu m/5\mu m$	
M_{IS}	$20\mu m/2\mu m$	mf = 2
M_{bf}	$10\mu m/0.8\mu m$	

Table 5.2: Design values for the devices in the mixer based optical front end

The layout of the optical front end is shown in Fig. 5.5. In practical design, a shield layer of Metal 4 covers all the chip area excluding the photodiode and pads for the purpose of protection on the circuits from laser light interferences. The mixer based optical front end has a very small core area of $50\mu m \times 75\mu m$. Most of the test chip space was occupied by the I/O pads as can be seen in Fig. 5.5.

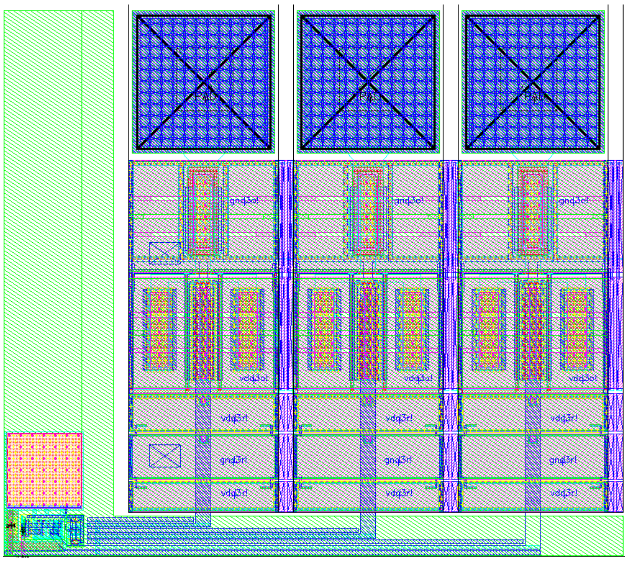


Figure 5.5: Layout of the mixed based optical front end in AMS $0.35\mu m$ CMOS

5.2.3 Measurement results

The mixer based optical front end was investigated in the early stage of the PhD research, which was mainly focused on the functionality verification of the idea of mixer based optical receiver. It was integrated and fabricated on a MPW, the tape out series number is PC3. The chip was packaged and soldered on an evaluation board, as shown in Fig. 5.6. In order to generate differential LO input to the mixer, a transformer (ADT1-6T from minicircuits) was embedded on the PCB, which operates from 0.02~325 MHz with insertion loss lower than -3 dB. The transformer is a 2-port input, 3-port output device, which means that the DC output can be defined by a fixed bias to the middle point of the output winds. For the purpose of saving I/O pads, the power supply of the optical front end was shared by all the circuits inside the chip, which means the power consumption unmeasurable. The optical measurement setup for the mixer based optical front end is basically the same as that of the photodiode measurement set up. The PCB board was sit and fixed on the optical platform while the light from the laser source was coupled to a $50\mu m$ optical fibre and then fed into the photodiode on chip directly. The alignment was achieved by a three dimensional positioner which the optical fibre was assembled.

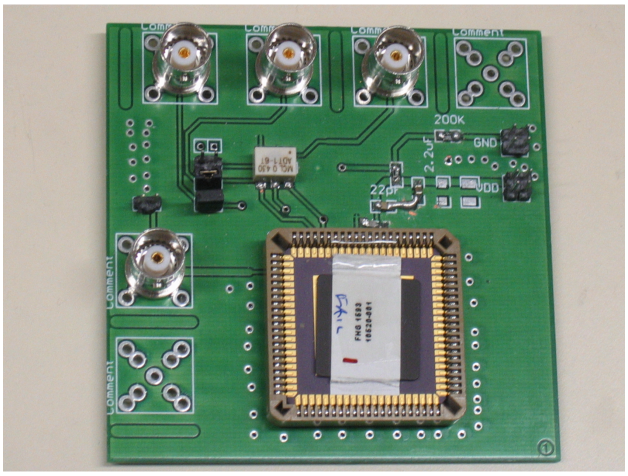


Figure 5.6: Evaluation board of the mixed based optical front end

During the measurement of the frequency response of the optical front end, the LO DC was set to be 2.0 V for a 3.3 V power supply, and 1.6 V for a 2.5 V power supply, respectively. The RF power from VNA 37397D was fixed at 3 dBm. Fig. 5.7 gives the measured output power from the mixer based optical front end versus the frequency sweep. Typically it gives a -3 dB bandwidth of 250 MHz at 3.3 V power supply and 210 MHz at 2.5 V power supply. This results were in good agreement with the hand calculated bandwidth in section 5.2.1, where the dominant pole of the system was assumed to be determined by the parasitic capacitance of the photodiode and the input impedance of the differential pair.

Because the DC current can not be measured due to the power supply sharing with the other circuits, it was not known that at which DC this bandwidth was achieved. However based on the measured responsivity of 0.3 A/W for the n-well photodiode in AMS $0.35\mu m$ CMOS in chapter 2, and the input laser power 600 μW, the DC current through the mixer can be estimated to be around $70\mu A$. What should be noted here is, for this measurement, the laser source used was HFE4080-32X-XBA instead of HFE6391-561, since during the measurement, the later was found to have ripples in the front-end's output at the interested frequencies (DC to several hundred MHz). Furthermore, there is no worry about the bandwidth of HFE4080-32X-XBA (6 GHz of analog bandwidth from the data sheet), which is enough to cover the bandwidth of the mixer based optical front-end.

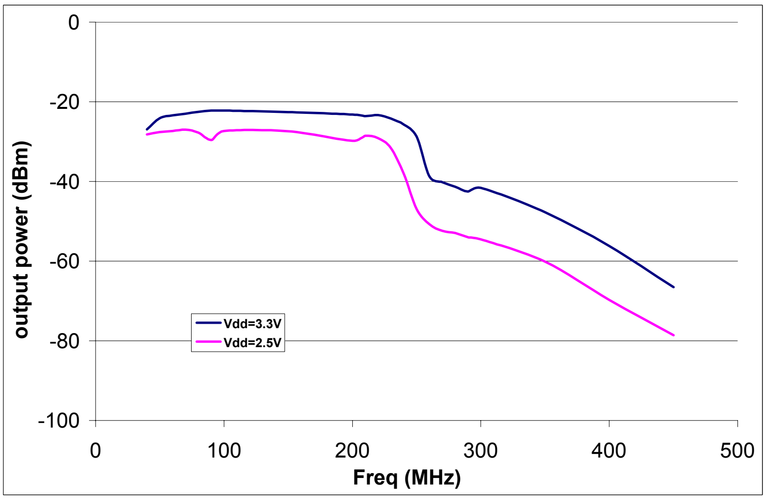


Figure 5.7: Measured frequency response of the mixer based optical front end

5.3 TIA based optical Front end

5.3.1 System architecture

Besides the consideration of speed, photodiodes generally present relatively large parasitic capacitance (from 100 fF to several pF depending on applications), which introduces great difficulties to the design of high speed optical front end. A transimpedance amplifier with low input impedance and a certain gain is necessary for speed consideration and noise consideration. To further alleviate the burden of the photodiode's capacitances on the TIA, ReGulated-Cascade (RGC) input stage can be used as the input stage which reduces the input impedance of the amplifier significantly by applying a local feedback network to a common-gate stage [77]. In this work, an improved RGC transimpedance amplifier was designed and implemented with a current shunt-shunt feedback for gain boosting. The details on the design of the TIA can be found in chapter 3.

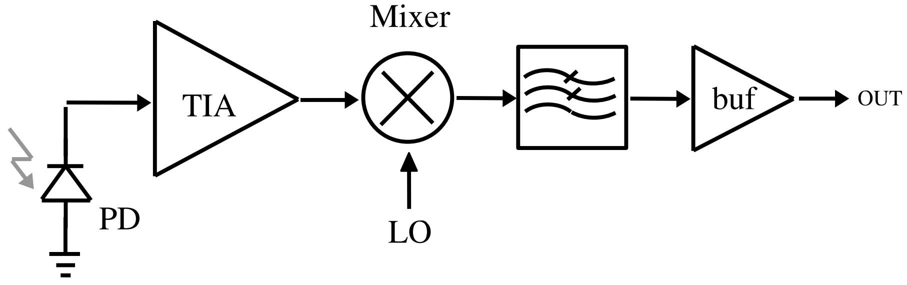


Figure 5.8: Block diagram of the optical front-end

Fig. 5.8 re-plots the block diagram of the proposed CMOS optical receiver. PD converts the light energy into electrical one (in current). The TIA amplifiers the tiny current and gives a voltage output. As mentioned in Chapter 4, a broadband mixer is placed after the TIA for downconversion of the discrete high frequency harmonics since an acoustic wave imaging system has useful signals only located at the mode lock frequency and its harmonics. The IF amplifier and output buffer are for further amplification and match to the measurement equipment.

5.3.2 Circuits and simulation results

The TIA based optical front end consists of three main circuit blocks, namely, photodiode, transimpedance amplifier and mixer. The individual circuits have been described thoroughly in the previous chapters. There is a big change for the D2S circuit in the tape out ATTO1B, compared to the D2S circuit in tape out RF1, as shown in Fig. 5.9 and Fig. 4.19. Resistor load has been changed to PMOS transistor load for this design to save silicon area and improving efficiency. Generally Resistor load is superior in the noise performance compared to PMOS load. But for D2S circuit, the noise is not a problem since D2S is the last stage in the chain of optical receiver. The gain of transimpedance amplifier and mixer greatly suppress its noise contribution to the overall noise performance. Furthermore, the resistor load in Fig. 5.9 directs the signals in both paths of the differential pair to the output by the current mirror, while the resistor load in Fig. 4.19 wastes half of the signal energy since only one signal path of the differential pair is conducted to the output.

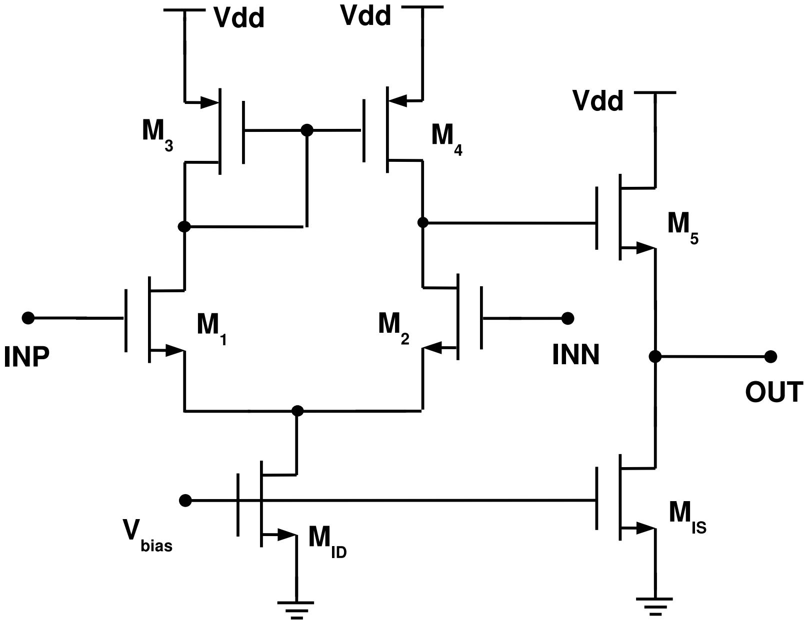


Figure 5.9: Schematic of the Differential to Single circuit in the 5 GHz optical front end

Another update on the D2S circuit is the source follower output buffer. a current source M_{IS} was added to bias the common gate transistor on chip, while in the original design, a DC terminator (50 Ω) is necessary to bias the output buffer. With this modification, the optical front end can be directly measured by the AC input coupled spectrum analyser and no additional 50 Ω DC terminator is needed in the measurement as mentioned in section 4.4. Moreover, the configuration in Fig. 5.9 saves a lot of current compared to the original one, which relieve the burden of large power current through the needle probe, as mentioned again, in section 4.4. in Fig. 5.9, M5 consumes only 3 mA current, while in Fig. 4.19, the common gate transistors (M5 and M6) consumes a current of 23 mA because of the low resistance of the DC terminator.

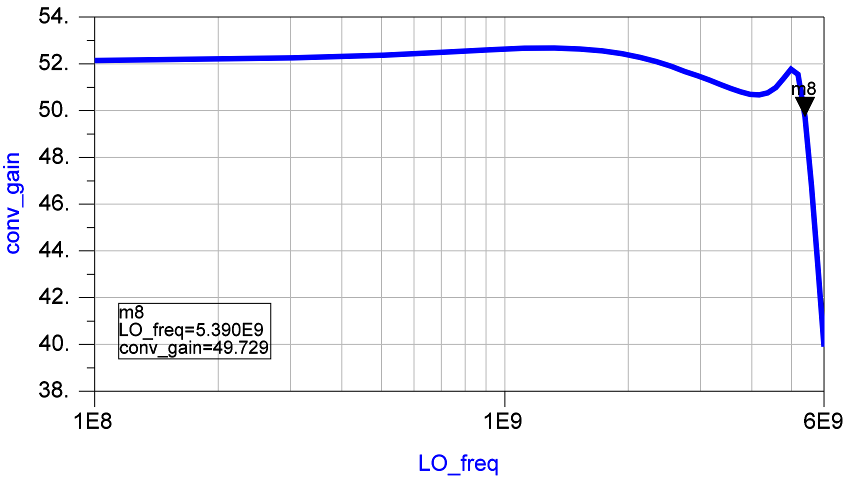


Figure 5.10: Simulated frequency response of the 5GHz optical front end

Fig. 5.10 gives the simulation results of the frequency response of the TIA based optical front end, where the photodiode is modeled as a capacitance of 0.6 pF with a DC current source of 10 μA. It can be seen that the optical front end achieves a conversion gain of 52 dB and a -3 dB bandwidth of 5.39 GHz. Here the conversion gain of 52 dB can be translated into a transimpedance gain of 85 $dB\Omega$ using the equation of $Z_T = Z_0 \cdot S21/(1-S11)$.

However, the simulation shows only the bandwidth of the TIA and mixer, as the frequency of the photodiode has not been taken into account since it is just modeled here as a input capacitance to the TIA. No doubt, the photodiode would dominate the bandwidth of the optical front end if its intrinsic bandwidth is lower than 5.39 GHz, while the techniques to improve the speed of CMOS photodiode have been addressed in details in chapter 2. The layout of the optical front end is shown in Fig. 5.12, with a size of $1037\mu m \times 500\mu m$, again, most of the chip area is consumed by the inductors, pads, and the decoupling capacitors.

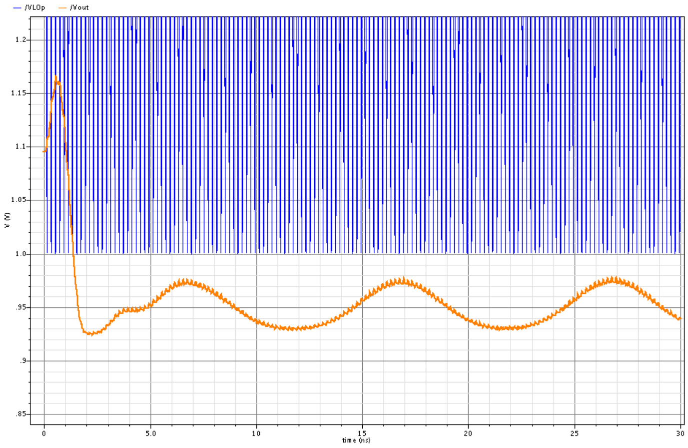


Figure 5.11: Post layout transient simulation results of the 5 GHz optical front-end

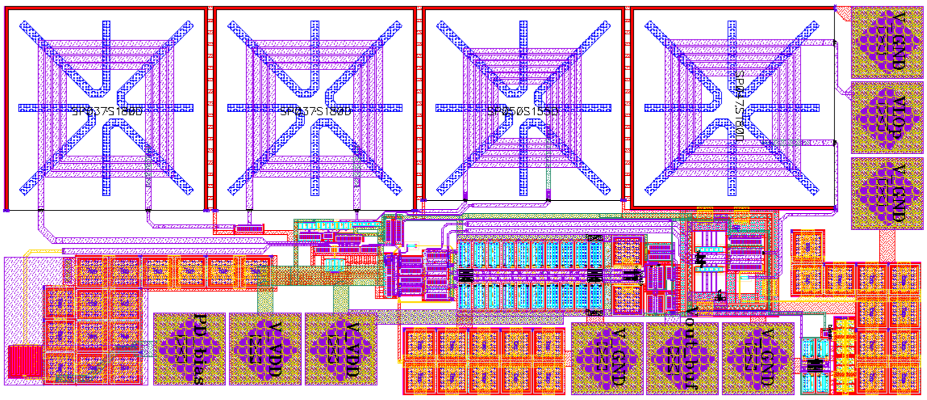


Figure 5.12: The layout of the 5GHz optical front end in AMS C35

5.3.3 Measurement results

For characterization purpose, a sub circuit of TIA and mixer (named TIA_mixer) was also taped out in Atto1b, where the photodiode was excluded. With this option, the bottle neck on the speed performance of the optical front end can be clarified. In the case that the optical front end is not fast enough, we can figure out that which devices or circuits are not within the specification requirements. The measurement on the TIA_mixer was performed first since it does not need an optical input signal and the test setup was relatively simple and can be referred as in Fig. 4.28. During the measurement, the LO DC was set to 2.5 V and LO input power was set to 0 dBm for conversion gain optimization. Fig. 5.13 shows the measured results of the conversion gain of the optical front-end. With RF input power varying from -80 dBm to -60 dBm, the power gain of the optical front end is almost the same, around 50 dB from DC up to 5 GHz. The -3 dB bandwidth is about 5.3 GHz, which is very close to the simulation results. The 50 dB conversion gain is equal to a 83 $dB\Omega$ transimpedance gain, as mentioned earlier. Again, it is in a good agreement to the simulation result. The frequency responses of the optical front end with different IF frequency were also done, as shown in Fig. 5.14. Although there was a peak for the conversion gains at the frequency of about 4.8 GHz, the frequency response at IF=50 MHz and 100 MHz demonstrates a relative flat conversion gain at the frequency range from DC to 5

GHz. On the other hand, when IF is 200 MHz, the optical front end suffers from a gain drop of 7-8 dB at the frequency range from 2 GHz to 4 GHz. This is believed to be caused by the limited bandwidth of the D2S circuit, as stated in 4.3.3. The measurement results on the mixer gave similar results when IF increased to 200 MHz, as mentioned in section 4.4.

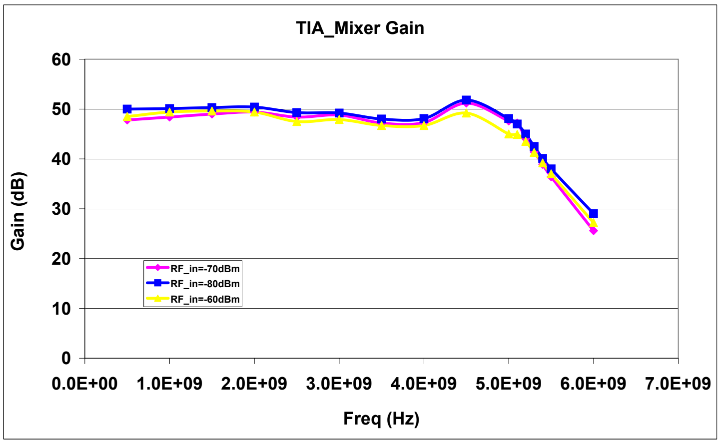


Figure 5.13: The measured frequency response of the TIA based optical front end with different RF input power

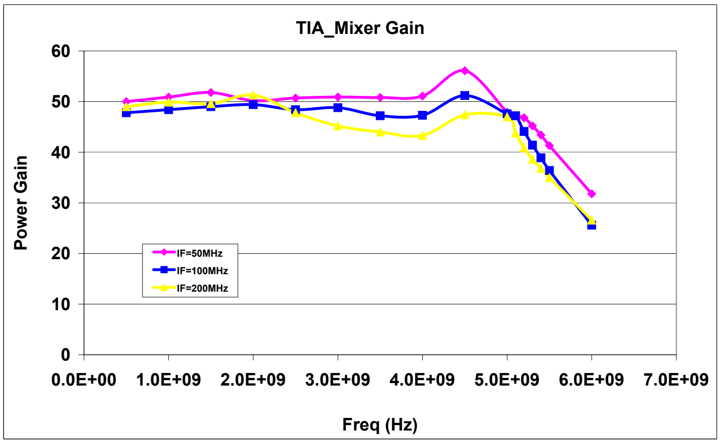


Figure 5.14: The measured frequency response of the TIA based optical front end with IF frequency of 50 MHz, 100 MHz and 200 MHz

The optical measurement setup of the front end is similar to that of the mixer, except that the the electrical RF input signal to the mixer was replaced by a high frequency optical input signal to the photodiode, as shown in Fig. 5.15. On the other hand, it reused the test setup of the photodiode, where the laser light was coupled to photodiode on chip through a 50 μm optical fibre. No additional anti-reflective coating on the top of the photodiode is applied to enhance responsivity. The alignment was achieved by the 3 dimensional positioner under the microscope. Note that the optical fibre should be well flat cleaved by about several *cm*s depending on the fibre's positioning angle to the chip, otherwise it is not easy to have a clear observation of the end of the fibre. The die micrograph of the optical front end is shown in Fig. 5.16.

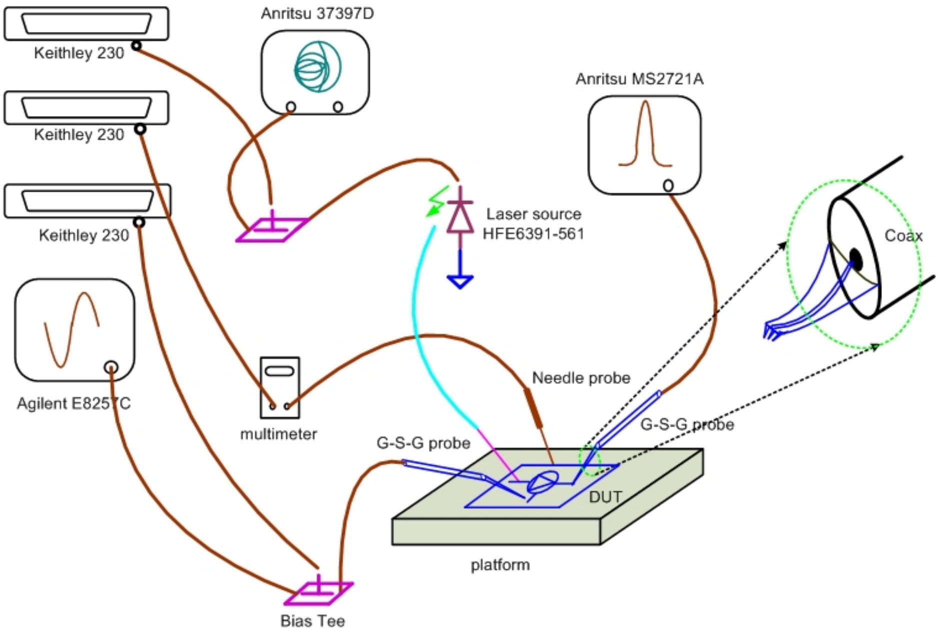


Figure 5.15: The measurement setup of the TIA based optical front end

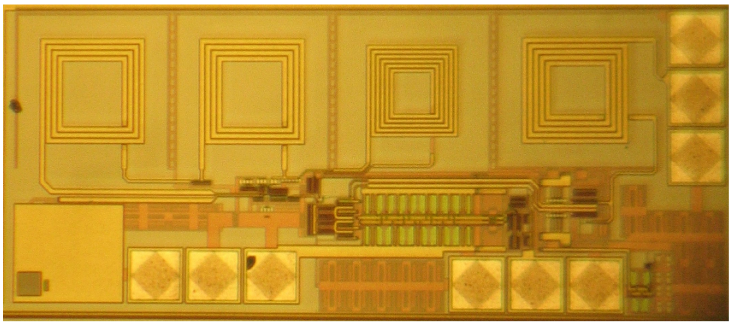


Figure 5.16: The die micrograph of the TIA based optical front end

Anritsu VNA 37397D was used as the 5 GHz signal generator to drive the laser source HFE6391-561 while a 40 GHz signal generator Agilent E8257C was employed to provide the LO signal to the optical front end. During the measurement, the RF power to the laser source is set to be 3 dBm, this value was optimized to have a reasonable output power dynamic range from the optical front end, typically it varies from - 70 dBm to around -20 dBm depending on the input frequency and the reverse biased voltage of the photodiode. For instance, a typical output power increment of 10 dB was observed during the measurement when the reverse biased voltage of the photodiode changed from 7 V to 9 V with an input frequency ranging from DC to 5 GHz. The LO DC was set to 2.5 V and LO input power was set to 0 dBm for conversion gain optimization. The optical front end consumes a total current of 42 mA (including that from the bias circuits and LO driver, Differential to Single circuit and output buffer) from a power supply of 3.3 V.

One of the points that is worth noting here is the set up of the evaluation board of HFE6391-561, which is designed to drive both the differential photodiode and single ended photodiode. Since in our case HFE6391-561 is an anode driven one, only the DATA+ port on the broad was used. However it is better not to let DATA- port float, instead it can be terminated with 50 Ω to give the best high frequency performance.

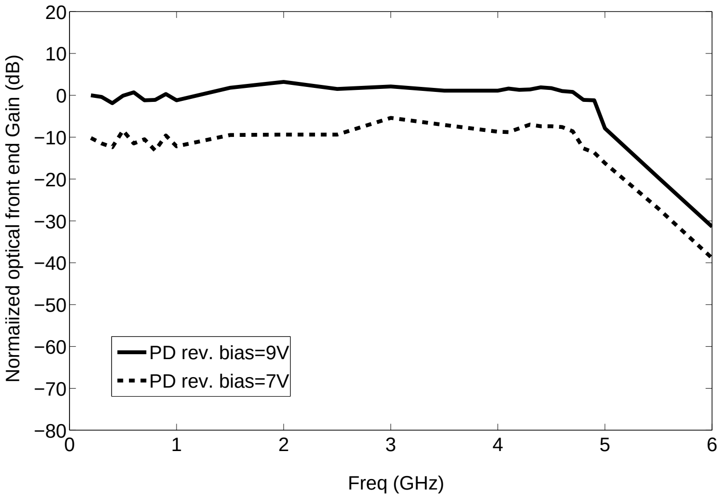


Figure 5.17: The measured frequency response of the TIA based optical front end

Fig. 5.17 gives the measured frequency response of the optical front end. The output power from the optical front end was measured by an Anritsu spectrum analyser with a fixed IF frequency of 100 MHz while RF and LO frequency were swept from DC to 6 GHz. Again, similar to the photodiode's frequency response, at the frequency range from DC to around several hundred MHz, the front end seems to suffer from some ringing or instability issues, a gain ripple of about

±3 dB was observed for both the reverse biased voltage of 7 V and 9 V. The optical front end demonstrates a - 3 dB bandwidth of 4.7 GHz and 4.9 GHz at the photodiode's reverse biased voltage of 7 V and 9 V, respectively. Measurement shows that the optical front end has a slight larger bandwidth than that of the photodiode (4.5 GHz for PD1, which was used in the optical front end). In fact, it is caused by the gain peaking effects of the TIA_mixer circuit at the frequency range of 4.5 GHz to 5 GHz, which was clearly illustrated in Fig. 5.13 and Fig. 5.14. For this gain peaking effects, the TIA_mixer circuit after the photodiode is a "equalizer" which compensates the excessive signal loss from the photodiode at the frequency range of 4.5 GHz to 5 GHz. Radovanovic reported a high speed optical receiver in $0.18\mu m$ CMOS process which operates up to 3 Gb/s[100], despite the n-well photodiode having only a cut off frequency of only several hundred MHz! The secret behind this is, the equalizer employed after the TIA which behaves like a high pass filter, balances the frequency responses of the photodiode and TIA.

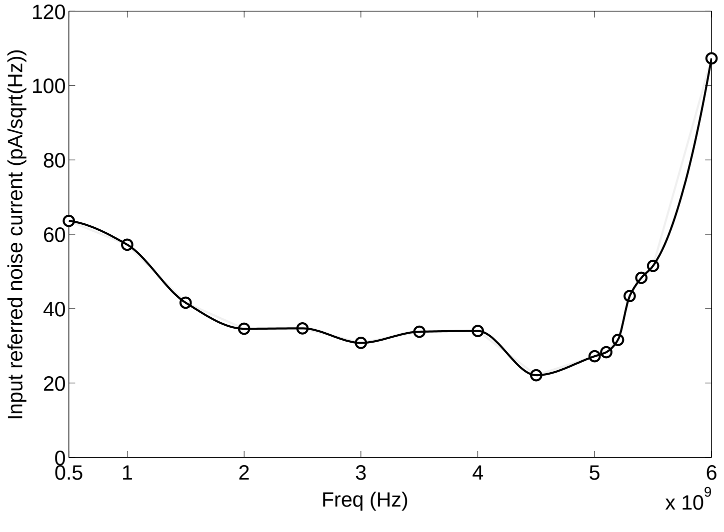


Figure 5.18: The measured input referred noise current of the TIA based optical front end

Sensitivity is an another important performance parameter for optical receivers in communication systems which denotes the smallest detectable optical input power at a specified bit error rate. For our application, sensitivity does not make sense because of the frequency downconversion by the mixer. Instead, the input referred noise current is used to characterize the optical front-end noise performance. For the noise measurement, first the output noise density of the optical front-end was measured with Anritsu spectrum analyser MS2721A which features a noise floor of lower than -150 dBm/Hz. The input referred noise current was then computed by dividing the measured output noise density by the measured transimpedance gain (Z_T) of the whole system. A typical value of 35 $pA/\sqrt{Hz}$ for the input referred noise current was achieved at 5GHz, contributed mostly from the transimpedance amplifier. Fig. 5.18 shows the measured input referred noise current of the optical front-end. Table 5.3 gives a summary of the front end's performance.

performance parameters	measured value
process	0.35 μm CMOS
chip area	$1037\mu m \times 500\mu m$
transimpedance gain	83 $dB\Omega$
-3 dB bandwidth	4.9 GHz
input referred noise current	35 $pA/\sqrt{Hz}$@5GHz
power consumption	42 mA @ 3.3 V

Table 5.3: The performance summary of the optical front end

Since the optical front end was not packaged, it brings difficulties to the on wafer measurement. One of the issues during the measurement was that the output power of the optical front end is relatively sensitive to the accuracy of the optical alignment, specifically, the alignment of the optical fibre to the photodiode on chip. To minimize the effect of the possible drifting of the optical fibre during the measurement, the measurement of the frequency sweep (the frequency of RF and LO were swept simultaneously with a fixed IF frequency) should not last very long time. Or, alternatively, for each frequency sweep, record the maximally achieved output power by slightly tuning the alignment between the optical fibre and the photodiode.

Chapter 6

Conclusion

The work presented in this thesis has concerned the realization of a CMOS optical front end for an optical sensing system. The silicon implementation of the chip as a whole, and its sub block circuits including photodiode, transimpedance amplifier, mixer, has been designed and demonstrated. Although optical receivers for high speed application exist, in the main they are fabricated in expensive III, V compound process, i.e., *GaAs* (Gallium arsenide), or *InP* (Indium Phosphorus). Monolithic, high speed optical receiver is difficult to be implemented in CMOS because of silicon's relatively long optical absorption length (above 10 μm at light wavelength of 850 nm) and relatively slow carrier mobility. On the other hand, signal processing after the front end is generally done in a CMOS chip, which makes high speed CMOS optical front end an attractive solution. The technical bottle neck about high speed CMOS photodiode was detailed in chapter 2. In brief, it is the slow photogenerated carriers in the bulk of silicon that limits the speed of photodiode in CMOS, which presents a "long tail" in the photodiode's transient response at high frequency. Our solutions to overcome the problems caused by the slow diffusion carriers problems were proposed after the review and evaluation of the ideas appearing in publications to improve CMOS photodiode. By treating the N-well as a terminator to screen the slow photogenerated bulk carriers, two kinds of high speed CMOS photodiodes were developed and fabricated using AMS 0.35μm CMOS process. Detailed design considerations were addressed in later part of chapter 2. Measurement results give a -3 dB bandwidth of 4.9 GHz for the

CMOS photodiode and a responsivity of 0.016 A/W.

6 GHz transimpedance amplifier was demonstrated in chapter 3. A significant work was done to implement the high speed, broadband amplifier in $0.35\mu m$ CMOS process. For the first gain stage in the optical front end, not only the bandwidth of the transimpedance amplifier is of great concern, but also the transimpedance gain, noise are taken into account carefully when doing the design optimization of the block circuit. By applying a local feedback on a common gate stage, a so called Regulated Cascode (RGC) circuit was designed to be the input stage of the amplifier. It features a very small input impedance, which is important to alleviate the impact of the parasitic photodiode capacitance on the amplifier's bandwidth. The second stage of the transimpedance amplifier is a shunt shunt feedback common source stage, which behaves as the main gain stage of the TIA. For further bandwith enhancement, multiple inductive series peaking was applied to the shunt shunt feedback stage by isolating the inter stage parasitic capacitances. A ban width enhancement ratio of 2.1 has been observed from simulation results. Throughout chapter 3, the principles and the design optimizations for the noise, gain and bandwidth were described by small signal equivalent circuit analysis on the amplifier. The test set up and on wafer measurement was summarized in the last section of chapter 3. The TIA has a typical power consumption of 53 mW from 3.3 V power supply, while the power supply can be as low as 2.5 V (with the same transimpedance gain while the -3 dB bandwidth and power consumption drop to 3.4 GHz and 29 mW, respectively). It presents a transimpedance gain of 51 $dB\Omega$, with a -3 dB bandwidth of 6.02 GHz. The input referred noise current is about 25 $pA/\sqrt{Hz}$ at 5 GHz.

Different from the transitional optical receiver for high speed communication systems where signals are broadband and non-modulated, an acoustic wave imaging system has useful signals only located at the mode lock frequency and its harmonics. Therefore a broadband mixer is placed after the TIA for downconversion of the discrete high frequency harmonics. The design of a 5 GHz mixer in 0.35 μm CMOS was given in chapter 4. The different categories and topologies of mixer were reviewed first in this chapter, and Gilbert cell was then analyzed in detail. For the mixer design, the key challenge comes from the optimizations and trade offs between gain, power consumption, noise and port matching. For instance, the current of the mixer core, the LO amplitude, the size of the switching transistors, have different and sometimes opposite effects on the mixer's noise, gain and power consumption performances. Optimizations thus need to be done to balance all the specifications. Although the linearity of the mixer plays a important role in wireless receiver communications, it is not a problem in this work since the optical front end only deals with very low power signal. The auxiliary circuits such as LO driver and Differential to Single circuit were also included in the prototype mixer. In fact, the tape out of the high frequency mixer was before the tape out of the transimpedance amplifier and the whole system, the optical front end. Thus some of the design flaws in the D2S circuit were identified and corrected at the subsequent chip fabrication of the whole system, which was mentioned both in chapter 4 and 5. The measurement results show that the mixer achieves a power gain of 16 dB including the IF amplifier at 5 GHz LO frequency. The noise figure was measured to be 17.3 dB. Two tone measurement gave an IIP_3 of 1.5 dBm for the Gilbert cell mixer. The mixer core consumes a current of 10 mA.

The recent publications on CMOS optical receiver were reviewed at the beginning of chapter 5. The system architecture has several choices, namely, single balanced mixer based optical front end and TIA based optical front end. Mixer based optical front end has the advantages of simple topology, low power consumption and small silicon area while the problems associated are the noise and speed. However since there is no gain stage, the noise from the switching pair is directly referred to the input port which has a great impact on the sensitivity of the optical front end. Moreover, the photodiode parasitic capacitance dominates the high frequency response of the mixer based optical front end at the common source node of the differential pair. The proposed optical front end consists of a photodiode, a transimpedance amplifier, and a broadband mixer, auxiliary circuits such as biasing circuits, LO driver, differential to single stage, and output buffer are integrated in the prototype chip as well. High speed, directly modulated VCSEL from Honeywell was used as high speed optical source for the optical front end on wafer measurements. The optical front-end was measured to have a -3 dB bandwidth of 4.9 GHz, with a transimpedance gain of 83 $dB\Omega$. The whole system consumes a current of 40 mA from 3.3 V power supply, while occupying an area of 1037 $\mu m \times$

500μm in AMS 0.35 μm CMOS process. The results were published in proceedings, optoelectronics, SPIE 2007[106].

Of the future work that could be done with the optical front end, passive mixer is perhaps the most exciting. The passive mixer is now popularly employed in wireless receiver for its simple architecture, low power consumption and high linearity. The concern is the power loss (instead of power gain for active mixers) because of passive property. However this can be partially compensated by increasing the gain of the transimpedance amplifier. Currently in the implemented optical front end, mixer occupies a significant silicon area and the current is 10 mA for the core circuit. If passive mixer could be used instead without significant degradation of the front end's noise performance, 40% of the silicon area and 25% of the power consumption would be saved. Besides from the circuit point of view, it is also desirable to implement the optical front end in advanced CMOS process, i.e., using SiGe BiCMOS from AMS would be an optimum choice between cost and system performances. Although the price would go up from 810 $euro/mm^2$ for 0.35 μm CMOS to 1220 $euro/mm^2$ 0.35 μm SiGe BiCMOS[1], but this would be offset by the shrinkage of the chip area since the inductors in the optical front end could be removed with transistor's f_t jumping from 25 GHz to 60 GHz. And another advantage that BICMOS offers is the buried layer in the technology that can be significantly beneficial for the implementation of high speed photodiode in the optical receiver.

Finally, it is worth summarizing the contributions of the thesis in main and the publications on this subject.

- Design, analysis and implementation of a 6 GHz, 51-dBΩ transimpedance amplifier in 0.35μm CMOS process. This is the fastest TIA ever reported in 0.35 μm CMOS technology according to author's knowledge, details can be referred to section 3.4 and reference [91].
- Design and implementation of a 5 GHz, monolithically integrated optical front end in 0.35μm CMOS process. This is the first reported fully integrated

---

[1] refer to MPW shuttle pricelist 2007, Austria Microsystems.

CMOS optical front end that demonstrates a -3 dB bandwidth over 4 GHz, referring to [106].

- high speed photodiode designed and implemented in 0.35μm CMOS process.
- M. Li, B. R. Hayes-Gill and I. Harrison, *6 GHz transimpedance amplifier for optical sensing system in low-cost 0.35-μm CMOS*, Electronics Letters, Vol.42, No.22, pp. 1278 - 1279, Oct 2006.
- M. Li, B. R. Hayes-Gill, M. Clark, M. C. Pitter, M. G. Somekh and I. Harrison, *5GHz front-end for active pixel applications in standard 0.35µm CMOS,* Proceedings of SPIE, Vol. 6477, San Jose, USA, Jan, 2007.

Bibliography

[1] A. G. Bell, *American Journal of Science*, vol. 20, pp. 305, 1880.

[2] A. G. Bell, *Upon the production of sound by radiant energy,* Philos. Mag., vol. 11, pp. 510-528, 1881.

[3] Allan Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy*, John Wiley & Sons 1980.

[4] L. B. Kreuzer and C. K. N. Patel, *Nitric oxide air pollution: detection by optoacoustic spectroscopy,* Science, vol. 173, pp. 45-47, 1971.

[5] L. B. Kreuzer, N. D. Kenyon, and C. K. N. Patel, *Air pollution: sensitive detection of ten pollutant gases by carbon monoxide and carbon dioxide lasers,* Science, vol. 177, pp. 347-349, 1972.

[6] L. B. Kreuzer, *Laser optoacoustic spectroscopy for GC detection,* Anal. Chem., vol. 50, pp. 597A-606A, 1978.

[7] C. F. Dewey, R. D. Kamm, and C. E. Hackett, *Acoustic amplifier for detection of atmospheric pollutants,* Appl. Phys. Lett., vol. 23, pp. 633-635, 1973.

[8] S.L. Firebaugh, *"Miniaturization and Integration of Photoacoustic Detection"* PhD thesis, Massachusetts Institute of Technology, May 2001

[9] S.Sharples, *"0 All-Optical Scanning Acoustic Microscope"* PhD thesis, University of Nottingham, May 2003

[10] H. E. Davis, G. E. Troxell and C. T. Wiskocil, *The Testing and Inspection of Engineering Materials.* McGraw-Hill, 1964.

[11] D. E. Bray and R. K. Stanley, Nondestructive Evaluation, *A Tool in Design, Manufacturing and Service.* CRC Press, 1997.

[12] M. Clark, S. Sharples, and M. Somekh, "*Non-contact acoustic microscopy*" Meas. Sci. Technol,, Published in UK. 11 (2000) 1792-1801. Sep., 2000.

[13] M. G. Somekh, M. Liu, H. P. Ho, and C. W. See, "*An accurate non-contacting laser based system for surface wave velocity measurement,*" Measurement Science and Technology, vol. 6, pp. 1329-1337, 1995.

[14] C. B., Scruby and N. J. Drain, *Laser Ultrasonics*, *Techniques and Applications*, Hilger, Bristol, 1990.

[15] V. Krylov, E. Ponomarev and T. Shtentsel, *Characteristics of thermooptic excitation of sound in metals,* Phys. Bull., Moscow University, 1986.

[16] Y. Sugawara, O. B. Wright, O. Matsuda, M. Takigahira, Y. Tanaka, S. Tamura, and V. E. Gusev, "*Watching ripples on crystals,*" Physical Review Letters, vol. 88, no. 18, pp. 185504-1-185504-4, 2002.

[17] Y. Sugawara, O. B. Wright, O. Matsuda, and V. E. Gusev, "*Spatiotemporal mapping of surface acoustic waves in isotropic and anisotropic materials,*" Ultrasonics, vol. 40, pp. 55-59, 2002.

[18] O. B. Wright, Y. Sugawara, O. Matsuda, M. Takigahira, Y. Tanaka, S. Tamura, and V. E. Gusev, "*Real-time imaging and dispersion of surface phonons in isotropic and anisotropic materials,*" Physical B, vol. 316-317, pp. 29-34, 2002.

[19] Y. Sugawara, O. B. Wright, and O. Matsuda, "*Real-time imaging of surface acoustic waves in thin films and microstructures on opaque substrates,*" Review of Scientific Instruments, vol. 74, no. 1, pp. 519-522, 2003.

[20] O. B. Wright, and K. Kawashima, "*Ultrasonic Detection from Picosecond Surface Vibrations: Application to Interfacial Layer Detection,*" J. Appl. Phys., vol. 32, pp. 2452-2454, 1993.

[21] M. G. Somekh, "*Ultrafast laser ultrasonics-towards practical application,*" Personal communications, 2004

[22] The International Society for Optic Engineering, *Photodetector Materials and Devices,* San Joe, CA, Jan 2001.

[23] R. Sankaralingam High-Speed Photodetectors and Receivers For Long Haul Communication Systems, PhD thesis, University of Nortre Dame, Indiana. Oct 2005

[24] P. Bhattacharya, *Semiconductor Optolectronic Devices*. Prentice Hall (1997).

[25] Behzad Razavi, *Design of Integrated Circuits for Optical Communications,* McGraw-Hill Copmanies, Inc. 2003.

[26] G. Wang, T. Tokumitsu, I. Hanawa, K. Sato, and M. Kobayashi, *Analysis of High Speed PIN Photodiodes S Parameters by a Novel Small Signal Equivalent Circuit Model,* IEEE Microwave and Wireless Components Lett., vol. 12, no. 10, pp. 378-380, October 2002.

[27] J. Harari, G. Jin, J.P. Vilcot, and D. Decoster, *Theoretical Study of PIN Photodetectors, Power Limitations from 2.5 to 60 GHz,* IEEE Trans. on Microwave Theory, vol. 45, no. 8, pp. 1332-1336, August 1997.

[28] V.M. Hietala, G.A. Vawter, T.M. Brennan, and B.E. Hammons, *Travelling wave Photodetectors for High Power, Large Bandwdith Applications,* IEEE Trans. on Microwave Technol., vol. 43, no. 9, pp. 2291-2298, September 1995.

[29] H. Tempkin, T.P. Pearsall, J.C. Bean, R.A. Logan, and S. Luryi, Ge_xSiO_{1-x} *Strained Layer Superlattice Waveguide Photodetectors Operating Near 1.3 um,* Appl. Phys. Lett., vol. 48, no. 15, pp. 963-965, April 1986.

[30] T. Tashiro, T. Tatsumi, M. Sugiyama, T. Hasimoto, and T. Morikawa, *A Selective Epitaxial SiGe/Si Planar Photodetector for Si Based OEIC,* IEEE Trans. on Electron Dev., vol. 44, no. 4, pp. 545-550, April 1997.

[31] J. E. Bowers and C. A. Burrus, *Ultrawide-band long-wavelength p-i-n photodetectors.* IEEE Journal of Lightwave Technology, LT-5 (10): 1339{1350 (Oct. 1987).

[32] R. Stevens, *Modulation Properties of Vertical Cavity Light Emitters,* Ph.D. thesis, Royal Institute of Technology Sweden, November 2001.

[33] J.D. Lambkin, T. Calvert, B. Corbett, J. Woodhead, S.M. Pinches, A. Onischenko, T.E. Sale, J. Hosea, P. Van Deale, K. Vandeputte, A. Van Hove, A. Valster, J.G. McInerney, and P.A. Porta, *Development of a Red VCSEL to Plastic Fibre Module for use in Parallel Optical Data Links,* Proc. of SPIE, vol. 3946, pp. 95-105, 2000.

[34] S.Y. Chou, Y. Liu, W. Khalil, T.Y. Hsiang, and S. Alexandro, *Ultrafast Nanoscale Metal-Semiconductor-Metal Photodetectors on Bulk and Low Temperature Grown GaAs,* Appl. Phys. Lett., vol. 61, no. 7, pp. 819-821, August 1992.

[35] M.Y. Liu, S.Y. Chou, S. Alexandro, C.C. Wang, and T.Y. Hsiang, *110 GHz Si MSM Photodetectors,* IEEE Trans. on Electron Dev., vol. 40, no. 11, pp. 2145 - 2146, November 1993.

[36] P. Fay, Photodetectors. *In Encyclopedia of Materials: Science and Technology*, Elsevier Science Ltd. (2001).

[37] S. M. Sze, *The Physics of Semiconductor Devices.* New York: Wiley, 1981, ch. 10.

[38] J. E. Bowers and Y. G. Wey, High-speed photodetectors. In M. Bass, editor, *Handbook of Optics: Fundamentals, Techniques and Design*, McGraw-Hill Inc., 2nd edition (1995).

[39] L.A. Bergman, W.H. Wu, A.R. Johnson, R. Nixon, S.C. Esener, C.C. Guest, P. Yu, T.J. Drabik, M. Feldman and S.H. Lee, *Holographic optical interconnects for VLSI*, Opt. Eng. 25, 1109-1118 (1986).

[40] W.H. Wu and A.R. Johnson, *CMOS detector cells for holographic optical interconnects in microcircuits,* SPIE vol. 752 on Digital Optical Computing, pp 192-199 (1987).

[41] E. Sackinger, *Broadband Circuits for Optical Fiber Communication*, John Wiley & Sons, Inc., Hoboken, New Jersey.

[42] T. K. Woodward and Ashok V. Krishnamoorthy, *1-Gb/s Integrated Optical Detectors and Receivers in Commercial CMOS Technologies,* IEEE Journal of Selected Topics in Quantum Electronics, vol. 5, no. 2, March/April 1999.

[43] Semiconductor Industry Association, *Semiconductor Industry Association Technology Roadmap,* 1994, available http://notes.sematech.org/ntrs/Rdmpmem.nsf.

[44] Aoife M. Moloney, *A CMOS Monolithically Integrated Photoreceiver Incoporating an Avalanche Photodiode,* PhD thesis, University College Cork, Ireland. Apr 2003.

[45] J. M. Senior, *Optical Fiber Communications: Principles and Practice*. Prentice Hall, second edition (1992).

[46] S. M. Sze, *Physics of Semiconductor Devices*. John Wiley and Sons, second edition (1981).

[47] P. J.-W. Lim, A. Y. C. Tzeng, H. L. Chuang, and S. A. St. Onge, *A 3.3 V monolithic photodetector/CMOS preamplifier for 531 Mb/s optical data link applications,* in Tech. Dig. 1993 Int. Solid State Circuits Conf. (ISSCC), 1993, pp. 96-97.

[48] [18] D. M. Kuchta, H. A. Ainspan, F. J. Canora, and R. P. Schneider, *Performance of fiber optic data links using 670-nm CW VCSEL's and a monolithic Si photodetector and CMOS preamplifier,* IBM J. Res. Develop., vol. 39, nos. 1-2, pp. 63-72, 1995.

[49] S.M. Csutak, J.D. Schaub, W.E. Wu, R. Shimer, and J.C. Campbell, *High-Speed Monolithically Integrated Silicon Photoreceivers Fabricated in 130 nm CMOS Technology,* IEEE J. of Lightwave Technol., vol. 20, no. 9, pp. 1724-1729, September 2002.

[50] S.M. Csutak, J.D. Schaub, W.E. Wu, R. Shimer, and J.C. Campbell, *8-Gb/s CMOS Compatible Monolithically Integrated Silicon Optical Receiver,* in Proc. Optical Fiber Communication Conf (OFC), Mar. 2002, pp.585-586.

[51] Sebastian M. Csutak, Jeremy D. Schaub, Wei E. Wu, Rob Shimer, and Joe C. Campbell, *CMOS-Compatible High-Speed Planar Silicon Photodiodes Fabricated on SOI Substrates,* IEEE J. of quantum Electronics, vol.38, no.2, Feb 2002.

[52] B.Yang, J.D. Schaub, S.M. Csutak, D. L. Rogers, and J.C. Campell, *10-Gb/s All-Silicon Optical Receiver,* IEEE Photonics Technology Letters, vol.15, no.5, May 2003.

[53] R.Swoboda and H. Zimmermann, *A Low-Noise Monolithically Integrated 1.5-Gb/s Optical Receiver in 0.6umBiCMOS Technology,* IEEE J. of Selected Topics in Quantum Electronics, vol.9, no.2, March/April 2003.

[54] R.Swoboda, Johannes Knorr, and H. Zimmermann, *A 5-Gb/s OEIC With Voltage-Up-Converter,* IEEE J. of Solid-State Circuits, vol.40, no.7, July 2005.

[55] Johannes Sturm, Martin Leifhelm, Harald Schatzmayr, Stefan Groi, and Horst Zimmermann, *Optical Receiver IC for CD/DVD/Blue-Laser Application,* IEEE J. of Solid-State Circuits, vol.40, no.7 July 2005.

[56] S. Kimura, K. Maio, T. Doi, T. Shimano, and T. Maeda, *Photodetectors Monolithically Integrated on SOI Substrate for Optical Pickup Using Blue or Near-Infrared Semiconductor Laser,* IEEE Trans on Electron Devices, vol. 49, no.6, June 2002.

[57] C. Rooman, D. Coppee, M. Kuijk, *Asynchronous 250 Mb/s Optical Receivers with Integrated Detector in Standard CMOS Technology for*

Optocoupler Applications, IEEE Journal of Solid- State Circuits, vol. 35, no. 7, July 2000, pp. 953-958.

[58] Natalie Wong, *CMOS Integrated PhotoDiode and Low Voltage Tranimpedance Amplifier in Si-OEIC Application,* university of Toronto, available from http://www.eecg.toronto.edu/~kphang/ece1371/pd_tia.pdf

[59] Kim, D. M., Lee, J. W., Dousluoglu, T., Solanki, R. and Qian. F, *High-speed lateral polysilicon photodiodes,* Semiconductor Sci. Technology, vol. 9, pp. 1276-1278, 1994.

[60] Diaz, D.C., Scho, C.L., Qi, J, Campbell, J.C. *Si/SiO2 resonant cavity photodetector,* Appl. Phys. Letter., vol. 69, pp. 2798-2800, November, 1996.

[61] S. Radovanovic, A.J. Annema and B.Nauta, “High-Speed lateral polysilicon photodiode in stand CMOS technology,” in *Proc. 33rd European Solid-state Device Research Conference*, Estoril, Portugal, Sep. 2003. Available from http://www.imec.be/esscirc/essderc-esscirc-2003/papers/all/331.pdf

[62] A. Aziz, O. Bonnaud, H. Lhermite and F. Raoult, *Lateral polysilicon pn diode: Current-voltage characteristics simulation between 200K and 400K using a numerical approach,* IEEE Trans. on Electr. Devices, vol. 41, pp. 204-211, 1994.

[63] Kamins T, *Polycrystalline silicon for integrated circuits and dis- plays,* Boston : Kluwer Academic Publisherse, 2-nd edition, p. 240, 1998.

[64] Plummer J., Deal M., Grin P, *Silicon VLSI technology; fun- dametals, practice and modelling,* Prentice Hall, p. 560, 2000.

[65] S. Radovanovic, A.J. Annema and B.Nauta, *3-Gb/s Monolithically Integrated Photodiode and Pre-amplifier in Standard 0.18 umCMOS,* IEEE. ISSCC 2004.

[66] AustriaMicroSystems (AMS), *0.35 um CMOS C35 Process Parameters ENG-182,* Revision #2.0, 2003

[67] Chen, Y-J. *A integrated CMOS optical receiver wih clock and data recovery circuit,* MSc. thesis, University of pretoria, South Africa, Aug. 2005.

[68] Sima Dimitrijev, *Understanding Semiconductor Devices,* Oxford university press, 2000.

[69] Jacoboni, C., C. Canali, G. Ottaviani, and A. A. Quaranta, Solid State Electron. 20, 2 (1977) 77-89.

[70] available from *http://www.ioffe.rssi.ru/SVA/NSM/Semicond/Si/electric.html*

[71] P. Gray, P. J.Hurst, S. H. Lewis, R. G.Meyer, *Analysis and Design of Analog Integrated Circuits,* 4th edition, John Wiley&Sons Inc., 2001, chatper 11.

[72] D. Guckenberger, K. Korgegay, *Novel low-voltage, low-power Gb/s tranimpedance amplifier architecture*, available from *aims.ece.cornell.edu/pdffiles/TIA.pdf*

[73] Yong-Hun, Sang-Gug Lee, *An inductance enhancement technique and its application to a shunt-peaked 2.5Gb/s transimpedance amplifier design*, IEEE Trans. on circuits and systems-express briefs, vol.51, no.11, Nov 2004.

[74] B. Analui, A. Hajimiri, *Bandwidth enhancement for tranimpedance amplifiers*, IEEE JSSC, vol.39, no.8, Aug 2004.

[75] Chia-Hsin Wu, C. Lee, W. Chen, S. Liu, *CMOS wideband amplifier using multiple inductive-series peaking technique*, IEEE JSSC, vol.40, no.2, Feb 2005.

[76] S.M. Park, H. Yoo, *1.25-Gb/s regulated cascode CMOS transimpedance amplifier for gigabit ethernet applications*, IEEE JSSC, vol.39, no.1, Jan 2004.

[77] S.M. Park, J. Lee, H. Yoo, *1-Gb/s 80-dBΩ Fully differential CMOS transimpedance amplifier in multichip on oxide technology for optical inerconnects*, IEEE JSSC, vol.39, no.6, Jun 2004.

[78] W. Chen, Y. Cheng, D. Lin, *A 1.8-V 10-Gb/s fully integrated CMOS optical receiver analog front end*, IEEE JSSC, vol.39, no.6, Jun 2004.

[79] B. Razavi, *A 622-Mb/s 4.5-*$pA/\sqrt{Hz}$ *CMOS Transimpedance Amplifier*. IEEE International Solid-State Circuits Conference Digest of Technical Papers, pages 162-163, Feb 2000

[80] B. Razavi, Design of analog CMOS integrated circuits, McGraw-Hill Companies, Inc., 2001.

[81] S. M. Park and C. Toumazou, *A packaged low-noise high-speed regulated cascode transimpedance amplifier using 0.6*μm *N-well CMOS technology,* in Proc. Eur. Solid-State Circuit Conf. (ESSCIRC), 2000, pp. 432-435.

[82] J. Lee, S.-J. Song, S. M. Park, C.-M. Nam, Y.-S. Kwon, and H.-J. Yoo, *A multichip on oxide of 1 Gb/s 80 dB fully-differential CMOS transimpedance amplifier for optical interconnect applications,* in IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers, Feb. 2002, pp. 80-81.

[83] Thomas H. Lee, *The design of CMOS Radio-Frequency Integrated Circuits.* Cambridge University Press, 2^{nd} edition, 2004.

[84] S.S Mohan, M.D. M. Hershenson, S.P. Boyd, and T.H. Lee, *Bandwith extension in CMOS with optimized on-chip inductors,* IEEE JSSC, vol. 35, pp. 346-355, Mar. 2000.

[85] D. A. Frickey, *Conversions Between S, Z, Y, H, ABCD, and T parameters which are valid for complex source and load impedances*, IEEE trans. on microwave and techniques, vol. 42, no.2, Feb 1994.

[86] Anthony Lord, *Advanced Calibration Techniques,* available from web site: http://legwww.epfl.ch/ekv/mos-ak/wroclaw/MOS-AK_AL.pdf .

[87] Chiang, M., Chen, P., Lu, S., and Meng, C., *DC-2.1GHz CMOS multiple feedback transimedance amplifier with high dynamic range and linearity,* Microwave and Optical Technology Letters, 200*3,* **3**6, (1), pp. 60 - 61.

[88] Schrodinger, K., Stimma, J. and Mauthe, M., *A fully Integrated CMOS Receiver Front-End for Optic Gigabit Ehternet,* IEEE J. Solid-State Circuits, 2002*,* **3**7, (7), pp. 874 - 880.

[89] Kuo, C., Hsiao, C., Yang, S., Chan, Y., *2 Gbit/s transimpedance amplifier fabricated by 0.35μm CMOS technologies*, Electron. Lett., 2001*,* **37**, (19), pp. 1158 - 1160.

[90] Terry H. Oxley, *50 Years Development of the Microwave Mixer for Heterodyne Reception,* IEEE trans. on microwave and techniques, vol. 50, no.3, Mar 2002.

[91] M. Li, B. Hayes-Gill and I. Harrison, *6 GHz transimpedance amplifier for optical sensing system in low-cost 0.35-μm CMOS*, Electronics Letters, Vol.42, No.22, pp. 1278 - 1279, Oct 2006.

[92] M. Terrovitis, R. G. Meyer, *Intermodulation Distortion in Current-Commutating CMOS Mixers,* IEEE JSSC, vol. 35, no.10, pp. 1461-1473, Oct. 2000.

[93] G. D. Vendelin, A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques,* John Wiley & Sons, Inc., 1990.

[94] T. Neuyen, N. Oh, V. Le and S. Lee, *A Low-Power CMOS Direct Conversion Receiver With 3-dB NF and 30-kHz Flicker-Noise Corner for 915-MHz Band IEEE 802.15.4 ZigBee Standard,* IEEE trans. on microwave theory and techniques, vol. 54, no. 2, Feb. 2006.

[95] M. T. Terrovitis, R. G. Meyer, *Noise in Current-Commutating CMOS Mixers,* IEEE JSSC, vol. 34, no. 6, pp. 772-783, Jun. 1999.

[96] H. Darabi, A. Abidi, *Noise in RF-CMOS Mixers: A Simple Physical Model,* IEEE Trans. on Solid State Circuits, vol. 35, no.1, pp. 15-25, Jan. 2000.

[97] E. A. M. Klumperink, S. M. Louwsma, G. J. M. Wienk and B. Nauta, *A CMOS Switched Tranconductor Mixer,* IEEE JSSC, vol. 39, no. 8, pp. 1231-1240, Aug. 2004.

[98] Keng Leong Fong, *Design and Optimization Techniques for Monolithic RF Downconversion Mixers,* PhD Thesis, University of California, Berkeley, Fall. 1997.

[99] Agilent Application notes, *Harmonic Balance Simulation.* Advanced Design System Documentation 2005A.

[100] S. Radovanovic, A.-J. Annema, and B. Nauta, *A 3-Gb/s optical detector in standard CMOS for 850-nm optical communication,* IEEE J. Solid-State Circuits, vol. 40, no. 8, pp. 1706 - 1717, Aug. 2005.

[101] C. L. Schow, L. Schares, S. J. Koester, G. Dehlinger, R. John, and F. E. Doany, *A 15-Gb/s 2.4-V optical receiver using a Ge-on-SOI photodiode and a CMOS IC*, IEEE photonics technology letters, vol. 18, no. 19, pp. 1981 - 1983, Oct. 2006.

[102] R. Swoboda and H. Zimmermann, *11 Gb/s monolithically integrated silicon optical receiver for 850 nm wavelength*, ISSCC Dig. Tech. Papers, pp. 904 - 911, Feb. 2006.

[103] C. Hermans, and M. S. J. Steyaert, *A High-speed 850-nm optical receiver front-end in 0.18-μm CMOS,* IEEE J. Solid-State Circuits, vol. 41, no. 7, pp. 1606 - 1614, Jul. 2006.

[104] M. Jutzi, M. Grozing, W. Mazioschek and M. Berroth, *2-Gb/s CMOS optical integrated receiver with a spatially modulated photodetector,* IEEE Photonics Technology Letters, vol. 17, no. 6, pp. 1268 - 1270 , Jun. 2005.

[105] W. Chen, and C. Lu, *Design and Analysis of A 2.5-Gbps Optical Receiver Analog Front-End in a 0.35-μm Digital CMOS technology,* IEEE Trans. Circuits and Systems - Regular Papers, 200*6,* 53, (4), pp. 977 - 983.

[106] M. Li, B. R. Hayes-Gill, M. Clark, M. C. Pitter, M. G. Somekh and I. Harrison, *5GHz front-end for active pixel applications in standard 0.35μm CMOS,* Proceedings of SPIE, Vol. 6477, San Jose, USA, Jan, 2007