UNIVERSITY OF PAVIA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

DOCTORAL THESIS IN MICROELECTRONICS

XXXIV CICLE

Design of Transimpedance Amplifiers for Coherent Optical Communications in 28 nm CMOS Technology

Author:

Laura ASCHEI

Supervisor:

Chiar.mo Prof. Danilo MANSTRETTA

Coordinator:

Chiar.mo Prof. Piero MALCOVATI

Academic year 2020/21

Declaration of Authorship

I, Laura Aschei, declare that this thesis titled, 'Design of Transimpedance Amplifiers for Coherent Optical Communication in 28 nm CMOS Technology' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

ABSTRACT

Facoltà di Ingegneria

Dipartimento di Ingegneria Industriale e dell'Informazione

Doctor of Philosophy

Design of Transimpedance Amplifiers for Coherent Optical Communication in 28 nm CMOS Technology

By Laura ASCHEI

Nowadays the increasing growth of big data applications, high serial data rates communications and cloud computing requires high bandwidth wireline performances. The most common metal connections are the coppers type, but their signal's losing is strongly frequency dependent. The optical interconnections are affected too of the frequency losing but in a minor quantity, for this reason they transport wideband signal over long distance. The optical links are more popular for the high data rate and remote communications. However, the implementation of optical system has many challenges about the sensitivity, the power consumption and the thermal design power constrains of the package. In order to improve more the data rate, in the 2005 the optical communications were associated to coherent modulations that has the high advantage to improve the spectral efficiency of the data [1]. Nowadays the most used technology in optical receiver circuits is the bipolar, due to the high bandwidth and low noise, but in coherent optical receiver the Digital Signal Processing has a very important role to decode the symbol and solve impairments like chromatic and polarization mode dispersion and group delay distortion, and the BiCMOS technology cannot be integrated with the DSP. This research is focused on analyzing the implementation in 28nm CMOS technology of the transimpedance amplifier (TIA) used to convert the photodiode current to voltage. The most critical performances of the TIA usually are the bandwidth and the noise, for that reason at the beginning I analyzed this tradeoff in the most commonly used topology, and I studied the bandwidth extension techniques present in literature. After that two design of TIAs has been realized: the first one uses the complementarity typical of the CMOS to improve the noise performances of the Common Gate topology, while in the second design I proposed a new technique to extend the bandwidth relaxing the tradeoff bandwidth/noise with the cost of higher power consumption. In the thesis is also reported a work realized in collaboration with Lorenzo Gnaccarini for the design of a Closed Loop Variable Gain

ABSTRACT

Amplifier as alternative topology to the traditional Gilbert Cell VGA for enhanced distortion performances.

[1] Kazuro Kikuchi, Fellow, IEEE, Fellow, "OSA Fundamentals of Coherent Optical Fiber Communications" JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 34, NO. 1, JANUARY 1, 2016

INDEX

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

The first studies on optical coherent communications started in the 1980s, motivated by the high sensitivity of the coherent receiver, but is was abandoned due to the progress in wavelength-division multiplexing systems, that allow to have high data-rates without changing the well established and simple modulation and detection methods, i.e. intensity modulation and direct detection (IM DD) [1.1]. In 2005 the interest in optical coherent communications was renewed by the demonstration of digital carrier-phase estimation in the coherent receivers led, that allows to use modulations with very high spectral efficiency such as M-ary Phase Shift Keying and Quadrature Amplitude Modulation (PSK, QAM) [1.1]. Another important advantage of coherent communication systems is the preservation of the phase information, opening up many possibilities to perform equalization of linear transmission impairments such as group-velocity dispersion and polarization-mode dispersion [1.1]. In this chapter we will describe the principle of coherent optical communications, starting from the modulation format to the technique of detection, in particular the homodyne receiver. The chapter will be concluded with a description of the Optical Transimpedance Amplifier and its role in the system.

1.1 Complex Modulations

Complex modulations have long been consolidated in radio applications, where their benefits in term of spectral efficiency are studied and confirmed. We know from Shannon-Hartley theorem that any communication channel has a maximum capacity, for given noise and bandwidth, as provided by (1.1.1) [1.1][1.2][1.3]:

$$C = B \cdot log_2(1 + \frac{S}{N}) \tag{1.1.1}$$

Where B is the bandwidth of the channel and SNR is the signal noise ratio. This means that to increase the capacity of the channel we have to increase the SNR or the bandwidth. The bandwidth is related to the rate of the transmitted symbol, for binary modulations (like NRZ), the transmitted symbol corresponds to a single bit, it means that the bit rate is equal to the symbol rate and it is restricted by the Shannon limit. Moving toward complex modulations, the idea is to transmit more than a bit per symbol, so that, with the same bandwidth, we can transmit higher bit rates [1.2]. This procedure is typically used in high data rate communications, where the symbol rate is called Baud rate and corresponds to (1.1.2) [1.2].

(1.1.2) $$Baud\ Rate = \frac{Bit\ Rate}{\log_2(N_{symbol})}$$

Where Nsymbol is the total number of available symbols per modulation. The most used complex modulation in optical coherent communications is the Quadrature Amplitude Modulation (QAM), which is obtained by the composition of two phase and amplitude modulated signals in quadrature, producing complex symbols, representable with a complex constellation lattice. In particular, for each component (I, in phase, and Q , in quadrature), the symbol can differ in phase by π. Fig.1.1.1 shows 3 possible 2n-QAM schemes, where the symbols are organized in a square lattice [1.2].

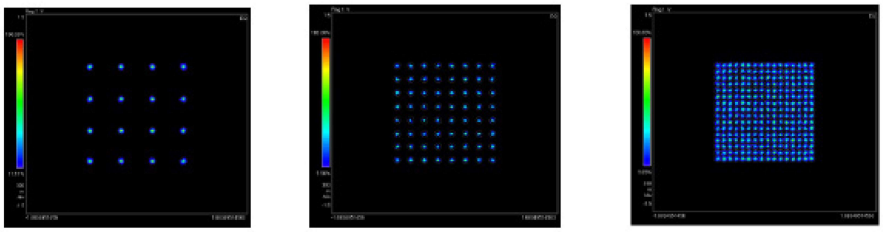
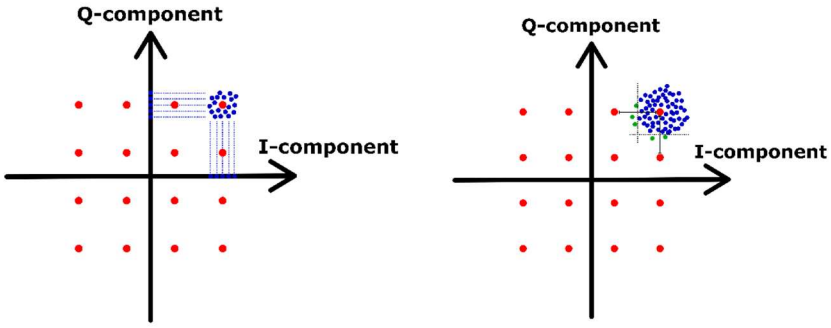


Fig.1.1.1 QAM schemes distributed in a square constellation lattice [1.2]

Usually 2n-QAM schemes with n and odd number are not used, as they can't be distributed in a square lattice and this would impact on the final Bit Error Rate (BER). Another possible complex modulation is the Amplitude and Phase Shift Keying, that organizes the symbols along concentric circles. This modulation was realized for satellite systems, because it is more robust to distortions, but usually it is not adopted in optical coherent communication due to the complexity of the implementation [1.2].

The constellation of transmitted and received signals differ due to non-idealities of the system. If we neglect all the linear non-idealities that the digital signal processing can compensate (they are explained in the next paragraph), we can identify the noise and the distortion as the main contributions to constellation degradation, which can be represented as a displacement of the symbol in the complex plane. If the corrupt symbol crosses the threshold used for the decoding, it will be read incorrectly, resulting in a wrong symbol.



(a) (b)

Fig.1.1.2 Non-linearity and noise effects on the constellation (a) and an excessive non linearity that provide an incorrect decoding of the symbol (green dots)

For this reason, it is important understand how to estimate the quality of the received coherent signal. Usually this is evaluated by two main parameters: the Bit Error Rate (BER) and the Error Vector Magnitude (EVM).

To compute the BER, we can approach using the theory or empirically. In the first case we get the BER by a formula, that is equal to (1.1.3) for a generic square shaped M-QAM [1.2] [1.4] [1.5].

$$BER \approx \frac{2}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3SNR}{2(M-1)}}\right) \tag{1.1.3}$$

Where Q is the complementary error function and SNR is the signal noise ratio. Alternatively, we can express the BER as a function of the EVM if we neglect any contribution of distortion and consider only Gaussian noise [1.2].

$$BER = \frac{2}{\log_2 M}\left(1-\frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3}{2\cdot(M-1)\cdot EVM^2}}\right) \tag{1.1.4}$$

The theoretical estimates considers only white thermal noise as system impairment. We can derive the BER also by a more empirical approach, introducing random modulated bits in the system model, and seeing at the output how many bits are interpreted incorrectly. This empirical approach allows to account for other impairments such as distortion but it can be quite time expensive for very low BER values [1.2][1.5].

The second important parameter that can be used to evaluate the quality of a coherent communication is the Error Vector Magnitude, that can provide information also about small variations of the constellation. It corresponds to the error vector between the ideal symbol and the measured one, as presented in Fig.1.1.3 [1.2].

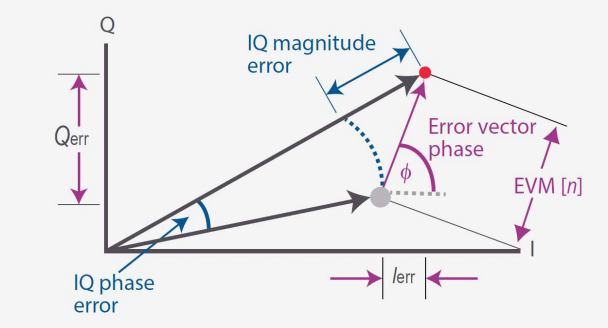


Fig.1.1.3 Vector representation of the Error Vector, where the red dot represents the measured point, and the gray dot is the target symbol [1.2].

According to the vector representation in Fig.1.1.3, the formula of the EVM of the n-th symbol is equal to (1.1.5).

(1.1.5) $$EVM_n = \sqrt{I_{err_n}^2 + Q_{err_n}^3}$$

And the normalized root mean square average EVM results equal to (1.1.6).

(1.1.6) $$\%EVM_{rms} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} EVM_n}{|peak\ reference\ vector|}$$

where the peak reference vector is the vector of the target constellation with the maximum magnitude. Sometimes we can find the EVM normalized not to the maximum target symbol, but to the average target symbol. In that case the formula becomes equal to (1.1.7) [1.2][1.5].

(1.1.7) $$\%EVM_{avg} = \frac{\frac{1}{N}\sum_{n=0}^{N-1} EVM_n}{|average\ reference\ vector|}$$

1.2 Coherent Optical Detection

In the Fig.1.2.1 we can see a standard coherent homodyne receiver, that is the most commonly used, and will be studied in this paragraph. The received signal from the fiber is decomposed in two different polarizations (X and Y POL), that are used to double the bit rate. The received light is then split by a 90° optical hybrid and down-converted, resulting in two orthogonal signal components: In-phase and Quadrature. After that, the light is converted in electrical domain, amplified and converted in digital bits, that are processed in order to rebuild the received information [1.1].

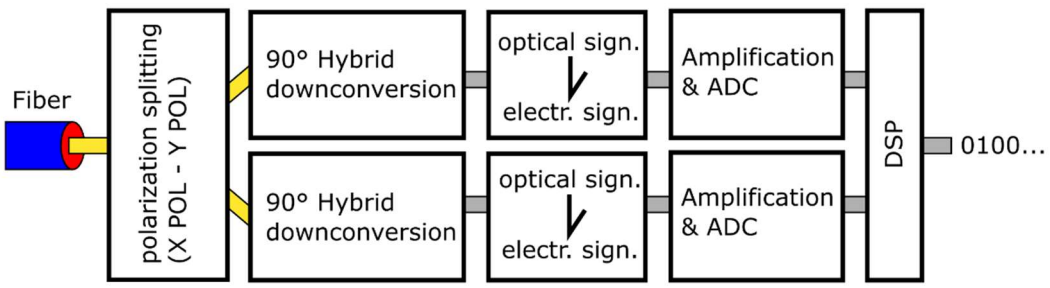


Fig.1.2.1 Generic schematic of an Optical Coherent Receiver

To understand how all this works more in detail it is better to focus only on a single down-converted signal, for example the In-phase signal for the X polarization, under the assumption that the state of polarization of signal and Local Oscillator (LO) are aligned by the polarization controller (PC). Fig.1.2.2 represents how the conversion from optical to electrical domain is performed in a coherent receiver. The electric fields of the signal (ε_s) and of the local oscillator (ε_l) are expressed in (1.2.1) and (1.2.2) [1.1].

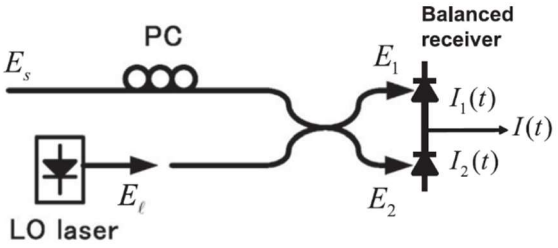


Fig.1.2.2 Configuration of a Polarization Control (PC), 3-dB optical coupler and a couple of photodiodes that are used to down-convert the signal and eliminate the dc component maximizing the beat signal [1.1]

(1.2.1) $$\varepsilon_s(t) = Real\{E_s\} = Real\{A_s(t)e^{i\omega_s t}\}$$

(1.2.2) $$\varepsilon_l(t) = Real\{E_l\} = Real\{A_l(t)e^{i\omega_l t}\}$$

Where ω_s, A_s and ω_l ,A_l are respectively the angular frequency and the complex amplitude of the received signal and the local oscillator, and Es/El are the phasor representation of the received signal and the local oscillator. To suppress the dc component and maximize the receiver signal, a balanced detection is performed by the use of two photodiodes in parallel instead of a single one, as shown in Fig.1.2.2: the first and the second photodiode are excited with two different signals, in particular the LO components are phase shifted by 180° by a 3-dB optical coupler. In this way the incident electric fields result equal to (1.2.3) and (1.2.4) and the resulting output current is equal to (1.2.5) and, for a homodyne system, this corresponds to (1.2.6) [1.1].

(1.2.3) $$E_1 = \frac{1}{\sqrt{2}}(E_s + E_l)$$

(1.2.4) $$E_2 = \frac{1}{\sqrt{2}}(E_s - E_l)$$

(1.2.5) $$I(t) = I_1(t) - I_2(t) = 2R\sqrt{P_s(t)P_l}\cos\{\omega_{IF}t + \theta_s(t) - \theta_l(t)\}$$

(1.2.6) $$I(t) = I_1(t) + I_2(t) = 2R\sqrt{P_s(t)P_l}\cos\{\theta_s(t) - \theta_l(t)\}$$

Where R is the responsivity of the photodiode, θ_s and θ_l are the phases of the signal and the LO respectively, and P_s, P_l are respectively the signal and LO power, computed using (1.2.7), where S_{eff} represent the effective beam area and ζ is the impedance in free space [1.1].

(1.2.7) $$P_i = \frac{S_{eff}}{\zeta} \cdot \frac{|A_i|^2}{2}$$

The expression (1.2.5) can be derived considering the transfer function of the photodiode, reported in (1.2.8) [1.1].

(1.2.8) $$I_i(t) = \frac{S_{eff}}{\zeta} R \left[Real \left\{ \frac{A_s(t) \exp(i\omega_s t) \pm A_l \exp(i\omega_l t)}{\sqrt{2}} \right\} \right]^{ms}$$

Where ms stands for mean square respect to the optical frequencies. It is interesting to notice in [1.6], that the output current is proportional to $\sqrt{P_l}$. This means that the down-conversion operation can also provide amplification of the signal, if the local oscillator power is sufficiently high, thus improving the noise performance. If we analyze (1.2.6), we can observe that it corresponds to the scalar product between the phasors of the signal and LO electric fields. This corresponds to finding the projection of the signal phasor on the LO phasor, as shown in Fig.1.2.3 [1.1].

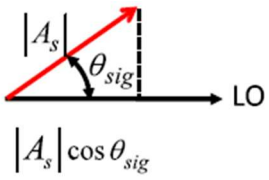


Fig.1.2.3 Phasor diagram of the signal and LO for the homodyne detection.

Hence, by this operation, we obtain the component of the signal that is in phase with the LO. Information can be stored in the amplitude (P_s) or in the phase (θ_s) of the signal, depending on the modulation. We can use the same technique to determine the quadrature components. The device that generates both the In-phase and Quadrature signals is called 90° optical hybrid and its scheme is represented in Fig.1.2.4 [1.1].

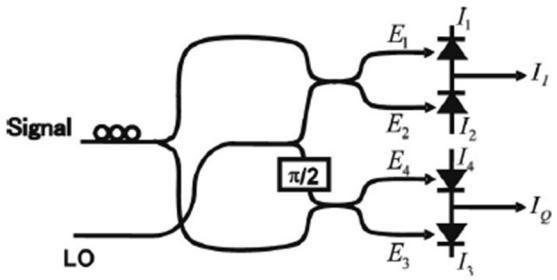


Fig.1.2.4 configuration of the 90° optical hybrid used to perform the phase diversity homodyne reception [1.1]

This type of receiver is called "phase diversity homodyne receiver" and it produces resulting currents I and Q equal to (1.2.9) and (1.2.10), and a complete complex current equal to (1.2.11) [1.1][1.6].

(1.2.9) $$I_I(t) = R\sqrt{P_s(t)P_l}\cos\{\theta_s(t) - \theta_l(t)\}$$

(1.2.10) $$I_Q(t) = R\sqrt{P_s(t)P_l}\sin\{\theta_s(t) - \theta_l(t)\}$$

(1.2.11) $$\boldsymbol{I_{complex}(t)} = I_I(t) - iI_Q(t) = R\sqrt{P_s(t)P_l}\exp\{i[\theta_s(t) + \theta_{sn}(t) - \theta_l(t)]\}$$

Where θ_{sn} is the phase noise. As mentioned above, the signal polarization has to be aligned to the polarization of the local oscillator. Due to random changes in the birefringence of the optical fiber produced by the asymmetries and not uniform solicitations, the received signal presents a deviation of the polarization. Usually this problem is partially solved using polarization diversity. The signal is split in two orthogonal polarization components by a Polarization Beam Splitter (BPS) and the same procedure is applied to the LO. The two signal-LO pairs are demodulated by two different 90° optical hybrids. Fig.1.2.5 shows the schematic of an optical coherent receiver with polarization and phase diversity [1.1][1.7].

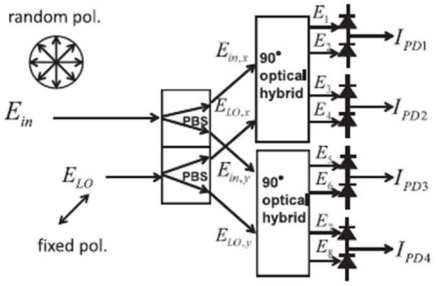


Fig.1.2.5 Configuration of the homodyne receiver where the phase and polarization diversities are explicit [1.1]

After the conversion in electrical domain the current must be amplified by a transimpedance amplifier and converted in digital domain in order to be processed by a Digital Signal Processing (DSP) to recover the information. The DSP has a key role in the de-codification of the signal.

At the output of the homodyne receiver we obtain the two signal components, in-phase and in-quadrature, used to reconstruct the complex signal. In reality the effective signal at this stage is affected by several impairments: carrier frequency and phase offset of the receiver local oscillator and the optical incoming signal, phase noise and amplitude noise due to electrical and optical devices, group delay variation of the fiber, TX and RX and other optical fiber transmission impairments like chromatic dispersion and polarization mode dispersion that gives rise to birefringence effects, mixing up the two polarization contributions. The summary of all these effects produces an essentially unreadable signal. In this scenario the signal is sampled and quantized, in order to move to the digital domain, where most of these impairments can be compensated for. The scheme of the DSP for decoding the symbol is represented

in Fig.1.2.6, where we can identify a first step relative to the Analog to Digital conversion, a consequent Frequency Domain Equalizer (FDE) and Adaptive Equalizer and a final step for the carrier phase recovery. In parallel to this linear impairments compensation, a nonlinearity compensation is realized.

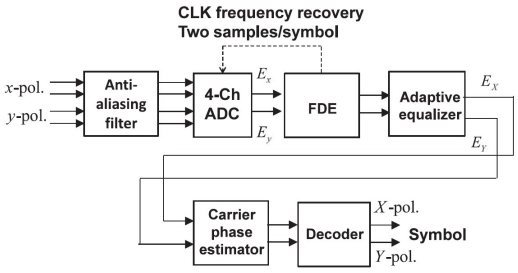


Fig.1.2.6 Common structure of a DSP used to process the symbol [1.1]

Following the signal flow step by step, it starts with the sampling and hold operation, preceded by the antialiasing filter, limiting the signal bandwidth to half of the sampling frequency, corresponding to the symbol rate (or BaudRate). For instance, if we have a dual polarization quadrature phase shift king modulation (DP-QPSK), and a bit rate equal to 100Gb/s, it corresponds to a symbol rate equal to 25 GBaud Rate (the factor of 4 between data rate and the baud rate is due a factor of 2 for the modulation and a factor of 2 for the polarization). In this example the sampling rate is 50 GS/s such that the ADC works in two times-oversampling condition in that way it can avoid aliasing effect but increases the computational complexity. After the sampling operation we find the FDE block. It is realized by an FIR filter and is used to compensate group delay variations (GDV) and chromatic dispersion of the optical fiber. After this compensation the clock frequency is extracted and used to correct the clock sampling frequency and phase to optimize the sampling instant of the symbol stream. Usually, the GDV is constant in time, allowing to have an FIR filter with constant coefficients. On the other hand, the polarization-mode- dispersion (PMD) is a linear time-dependent process, which demands an adaptive equalizer, typically implemented with an FIR butterfly structure, as reported in Fig.1.2.7. Notice that each $\overrightarrow{h_{ij}}$ block (i,j = x,y) is implemented by an FIR filter [1.1][1.8].

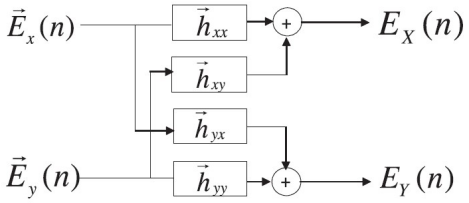


Fig.1.2.7 2x2 butterfly structure FIR filters where the tap coefficients are generated to compensate polarization mode dispersion [1.1]

The matrix $\overline{\overline{\boldsymbol{H}}}$, represented by (1.2.12), is obtained as the inverse matrix of $\overline{\overline{\boldsymbol{H}}}_{\boldsymbol{o}}$, which is the transfer function of the link. In particular, if $\vec{\boldsymbol{E}}_{\boldsymbol{in}}$ is the transmitted electric field and $\overline{\overline{\boldsymbol{H}}}_{\boldsymbol{o}}$ is the link matrix, the received electric field is given by (1.2.13) [1.1].

$$\bar{\bar{\boldsymbol{H}}} = \begin{bmatrix} h_{xx} & h_{xy} \\ h_{yx} & h_{yy} \end{bmatrix} = \bar{\bar{\boldsymbol{H}}}_{\boldsymbol{o}}^{-\mathbf{1}} \tag{1.2.12}$$

$$\vec{\boldsymbol{E}} = \begin{bmatrix} E_x(\omega) \\ E_y(\omega) \end{bmatrix} = \bar{\bar{\boldsymbol{H}}}_{\boldsymbol{o}}(\omega) \begin{bmatrix} E_{inx}(\omega) \\ E_{iny}(\omega) \end{bmatrix} \tag{1.2.11}$$

These FIR filters are in general used to compensate linear impairments, such as the transfer characteristics of optical filters in the links, and electrical circuits in the transmitter and receiver, that generate ISI [1.1].

After the FIR filters, we can identify in Fig.1.2.6 the Carrier Phase Estimator. This block is used to determine the phase of the local oscillator. If we look at (1.2.11), we can observe that the phase of the LO is still present in the expression, and we need to know it in order to subtract $\theta_l(t)$ from the phase information of the signal. In order to perform carrier recovery, the Viterbi-Viterbi 4th power algorithm is commonly used. It starts in taking the modulated signals, e.g. QPSK modulated symbols, that in general is equal to $e^{\frac{j2\pi}{M}} e^{j\theta_l}$, with m that can assume values from 0 to $M-1$, it corresponds to the product of the ideal symbol $e^{\frac{j2\pi m}{M}}$ with the LO component $e^{j\theta_l}$. Raising the modulated signal to the power of 4 (4 is the order of the modulation, i.e. M in general) the components that represents the ideal symbol becomes equal to 1 and you can get simply the component relative to the LO phase. This process usually is repeated on N sample to take their average (1.2.14) [1.1][1.5][1.8].

$$< \theta_l > = \frac{1}{jM} \cdot \arg\left(\sum_{n=0}^{N-1} \left(e^{\frac{j2\pi m}{M}} e^{j\theta_l} \right)_n^M \right) = \frac{1}{jM} \cdot \arg\left(\sum_{n=0}^{N-1} \left(e^{jM\theta_l} \right)_n \right) \tag{1.2.14}$$

Where $e^{\frac{j2\pi m}{M}}$ is the generic received symbol with m that can assume values from 0 to $M-1$. In case of QAM modulation, only the symbol along the diagonal, are used for the carrier recovery as QPSK with different amplitude level.

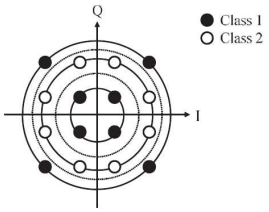


Fig.1.2.8 Constellation of a 16-QAM where the black symbol (class1) are used for phase estimation [1.1]

The last impairments that can be compensated for by the DSP are the deterministic non linearities. To do that, the received signal from the coherent detection is processed by a virtual transmitter, fiber and receiver amplifiers, emulated in digital domain, to reverse the nonlinear characteristic of the optical link [1.1]. To provide a visualization of the DSP role, Fig.1.2.9 shows a qualitative evolution of the received constellation along the digital signal processing. [1.8]

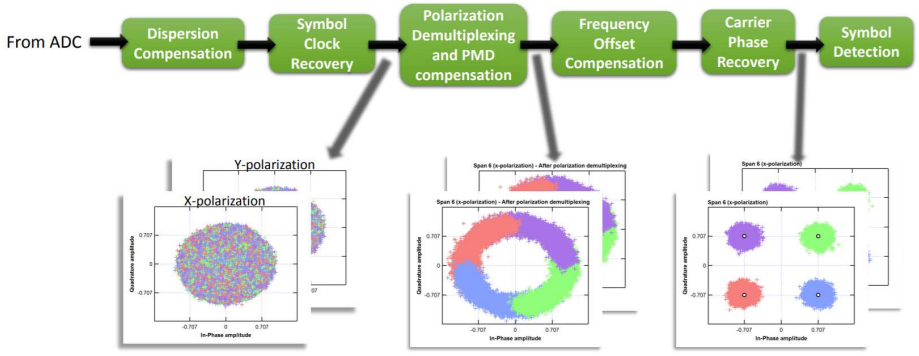


Fig.1.2.9 Evolution of the constellation recovery along the Digital Signal Processing [1.8]

From the above discussion, it is evident the paramount role of the digital processing in determining the final performances of the coherent optical link, and the complexity of such system. For that reason, we decided to evaluate the performances of any proposed design considering only the impairments introduced by the Transimpedance Amplifier, under the hypothesis that all the other impairments relative to the optical link can be compensated for.

1.3 Building Blocks of an Optical Transimpedance Amplifier

In the previous paragraph we have briefly described the receiver of an optical coherent system. This Ph.D. research has focused on the design of an Optical Transimpedance Amplifier (oTIA or TIA) that has the purpose of collecting the current coming out from the photodiodes and to provide a constant output voltage to the following ADC while adding as little noise and distortion as possible. Referring to the Fig.1.3.1, in any coherent receiver we have 4 oTIA: two couple of 2 oTIA for the two polarizations (X and Y), the 2 TIAs are used to collect respectively differential signals for the In-phase and Quadrature components.

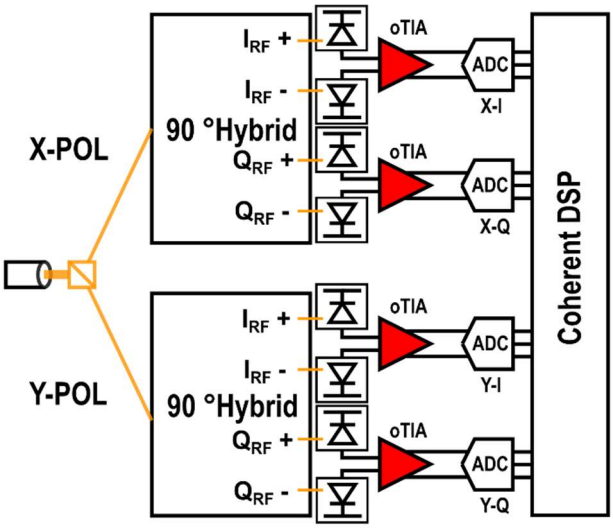


Fig.1.3.1 Generic schematic of a coherent receiver where the oTIA are highlighted in color red

The oTIA has the aim to amplify the current signal and convert into a voltage that will drive the ADCs. The oTIA has to provide a variability of the gain to avoid the saturation of the converter, a typical maximum output voltage swing is of the order of 500mVdpp. The oTIA generally presents the stages reported in Fig.1.3.2.

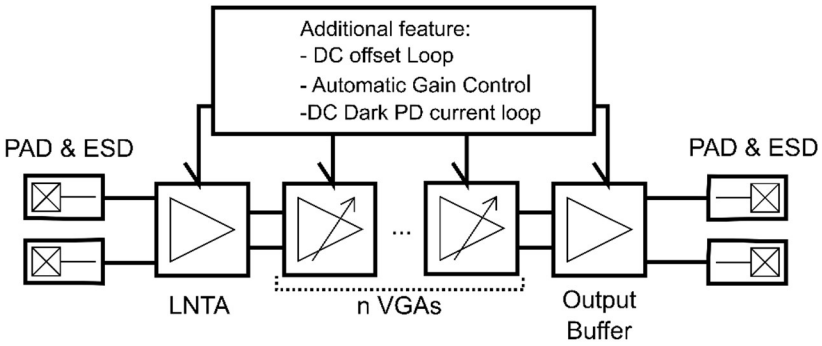


Fig.1.3.2 Schematic of an opto-TIA

We can identify the Low Noise Transimpedance Amplifier (LNTA), that has to perform the conversion of the signal from current to voltage; a certain number (n) of Variable Gain Amplifiers (VGAs), that are used to increase the transimpedance gain and provide variability of the Gain and, sometimes, are used also for equalizing the limited bandwidth of the LNTA; an Output Buffer used for the output impedance matching. In addition, we can have features like the DC offset Control Loop and the Automatic Gain Control. We have already talked about the importance of the bandwidth, noise and distortion in this application. We can now identify which stage is usually more involved in determining each of these characteristics.

The most critical stage of an oTIA is the LNTA. It is at the input of the amplification chain, where the signal is weaker, and therefore more subjected to noise. For this reason, the LNTA has to be a low noise source and has to guarantee enough high gain in order to minimize the noise contributions of the

following stages. This introduces an important tradeoff between noise and bandwidth. In fact, the LNTA's bandwidth is limited by the capacitance of the photodiode that loads its input node. To get over this limitation, the first stage has to show a low impedance at the input that usually is achieved with high gm and wide transistor. Moreover, the final bandwidth is limited by the minimum gain that the LNTA needs to reduce the other noise contributions. For that reason, usually the LNTA is the stage with the most critical design for the bandwidth and noise performances. We will see that in literature we can find different topologies, in open loop and in closed loop, that respectively are focused on improving the bandwidth or the noise.

The Variable Gain Amplifier have the key role of changing the gain of the transimpedance amplifier, so they are fundamental for the noise and distortion performances for the intermediate gain steps. As we already introduced, they are often used to provide some equalization to extend the bandwidth (Fig.1.3.3). Moreover Säckinger demonstrates that realize the transimpedance gain with not only the first stage, but with additional amplification, can help with the gain-bandwidth tradeoff of the TIA [1.9].

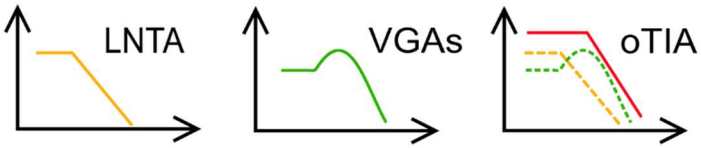


Fig.1.3.3 principle of equalization of the limited bandwidth of the LNTA

Ideally the transfer function of the TIA should be flat in frequency and have a constant group delay. In reality the acceptable ripple in the gain and in the group delay depend on the DSP that follows. The output swing that influences the distortion, depend on the Analog to Digital Converter that receives the signal from the TIA. In this work the output swing results equal to 500 mVppd. The noise and the distortion specifications depend on the modulation that we want to do. For example, a distortion around 1% of THD and a noise around a value of 10 to 20 pA/√Hz can be used for 16 QAM modulation [1.10]. To analyze the achievable modulation QAM with the proposed design, we realized a code in Matlab in order to test the TIA with a signal providing the S_{21}, the THD and the SNR. The algorithm of the program is reported in Fig.1.3.4.

The algorithm is divided in a first section where we define the parameters of the amplifier and the modulation properties. The parameters asked to the users are the transfer function (modulus and phase) of the amplifier to analyze, the THD and the SNR.

The parameters relative to the amplifiers are saved in the *struct amp* while all the data about the communication are saved in the *struct sys* .

The *struct amp* contains the file of the transfer function al all the information useful to read it, the THD, the gain at low frequency, the SNR, the output and input swing and the coefficient used to compute the distortion derived by the THD. In particular, we know from [1.11] that the characteristic of a non-linear system can be approximated by a Taylor expansion in the range of interest

(1.3.1) $$V_{OUT} = \alpha_1 x(t) + \alpha_2 x^2(t) + \alpha_3 x^3(t) \dots$$

For a differential amplifier the component with power 2 can be neglected and assuming that the input signal is equal to $x(t) = Acos(\omega_0 t)$, the expression (1.3.1) becomes equal to

(1.3.2) $$V_{OUT} = \alpha_1 Acos(\omega_0 t) + \alpha_3 A^3 cos(\omega_0 t)^3$$

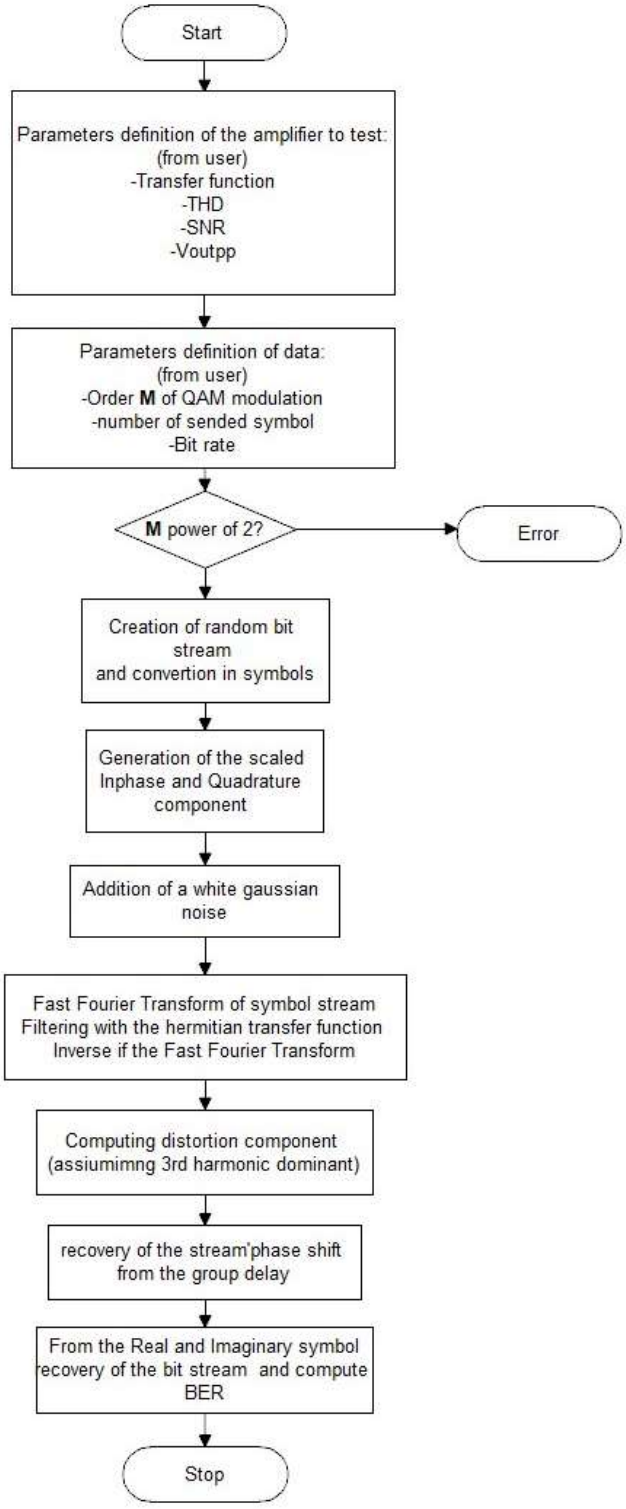


Fig.1.3.4 Flowchart of the algorithm used to test the BER for a given M-QAM modulation of TIA

Razavi demonstrates that this expression can be written in function of the harmonics, in particular it is equal to

$$V_{OUT} = \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right) cos(\omega_0 t) + \left(\alpha_3 \frac{A^3}{4}\right) cos(3\omega_0 t) + \frac{\alpha_2 A^2}{2}(1 + cos(2\omega_0 t)) \quad (1.3.3)$$

If we assume that the DC and second harmonic components are negligible, like in most of the case realized in this work, we can assume that the THD expression contains only the third harmonic component in particular it is equal to

$$THD = \sqrt{\left(\alpha_3 \frac{A^3}{4}\right)^2 \Big/ \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right)^2} = \left(\alpha_3 \frac{A^3}{4}\right) \Big/ \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right) \quad (1.3.4)$$

Under this assumption, (1.3.3) becomes equal to

$$V_{OUT} = \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right) cos(\omega_0 t) + THD \cdot \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4}\right) cos(3\omega_0 t) \quad (1.3.5)$$

In the Matlab code we have the output and input swing from the user at the first harmonic, and it corresponds to the coefficient $\alpha_1 A + \frac{3\alpha_3 A^3}{4}$, and we have also the THD. In that way we can obtain the expression of α_1 and α_3 that we use to consider the distortion in the model. In particular, we have that

$$\alpha_1 = \frac{V_{out}}{V_{in}} - \frac{3\alpha_3 V_{in}^2}{4} \quad (1.3.6)$$

$$\alpha_3 = \frac{4 V_{out} THD}{V_{in}^3} \quad (1.3.7)$$

If we can approximate that $\frac{V_{out}}{V_{in}}$ is equal to the gain of the transfer function at ω_0, the expressions (1.3.6) and (1.3.7) can be simplified in

$$\alpha_1 = G_{\omega_0} - \frac{3\alpha_3 V_{in}^2}{4} \quad (1.3.8)$$

$$\alpha_3 = \frac{4 G_{\omega_0} THD}{V_{in}^2} \quad (1.3.9)$$

α_1 and α_3 are saved in the *struct amp*.

The *struct sys* contains the number of bit that we want to test to evaluate the BER, the order of modulation M-QAM that we want to test, the BitRate and BaudRate and the respective periods.

After some check about the feasibility of the order of the modulation, there is the creation of the random bit stream and in the derivation of the real and imaginary streams relative to the chosen modulation. To perform the filtering with the transfer function we need also to create a signal in time, for that purpose is created another *struct param* to save all the useful parameters used to create the time signal and its fast Fourier frequency transformation, as the uppersampling of the time signal.

After the creation of the signal and its transformation by the FFT (Fast Fourier Transform), we can start to insert the white Gaussian noise to the real and imaginary components of the symbol strings by the function *wgn* already provided by Matlab, filter it with the transfer function and include distortion effect by the previously computed coefficients.
After the filtering, the signal has been delayed by the transfer function, to recover the correct phase of the signal to sample it, we used the low frequency group delay that we derive from the transfer function. Now we can sample the real and the imaginary, demodulate the symbols and compare the original bit stream with the new one to get BER.

The complete Matlab code is reported in the Appendix.

Chapter references

[1.1] K. Kikuchi, "Fundamentals of Coherent Optical Fiber Communications," in Journal of Lightwave Technology, vol. 34, no. 1, pp. 157-179, 1 Jan.1, 2016, doi: 10.1109/JLT.2015.2463719.

[1.2] Keysight Tehcnologies, "Everything You Need to Know About Complex Optical Modulation" published in USE, March 25, 2019, 5992-2888EN

[1.3] C. E. Shannon, "A mathematical theory of communication," Bell Sys.Tech. J., vol. 27, no. 3, pp. 379–423, Jul. 1948 and vol. 27, no. 4, pp.623–656, Oct. 1948

[1.4] Vahid Meghdadi, "BER calculation, reference : Wireless Communications by Andrea Goldsmith" January 2008

[1.5] Harald Rohde Nokia, Germany and Howard Wang, Nokia, USA, "Modeling and Simulation of Optical Transmitter and Receiver Components for Coherent Communications" The Optical Networking and Communication Conference & Exhibition, June 2019, short course SC39

[1.6] A. Davis, M. Pettitt, J. King and S. Wright, "Phase diversity techniques for coherent optical receivers," in Journal of Lightwave Technology, vol. 5, no. 4, pp. 561-572, April 1987, doi: 10.1109/JLT.1987.1075539.

[1.7] K. Kikuchi, "Optical Homodyne Receiver Comprising Phase and Polarization Diversities with Digital Signal Processing," 2007 Digest of the IEEE/LEOS Summer Topical Meetings, 2007, pp. 55-56, doi: 10.1109/LEOSST.2007.4288328.

[1.8] Οπτικά Δίκτυα Επικοινωνιών [Optical Communication Networks], "Coherent Optical Systems" Photonics Communications Research Laboratory Lecture

[1.9] E. Säckinger, "The Transimpedance Limit," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 57, no. 8, pp. 1848-1856, Aug. 2010, doi: 10.1109/TCSI.2009.2037847.

[1.10] M. G. Ahmed, T. N. Huynh, C. Williams, Y. Wang, P. K. Hanumolu and A. Rylyakov, "34-GBd Linear Transimpedance Amplifier for 200-Gb/s DP-16-QAM Optical Coherent Receivers," in IEEE Journal of Solid-State Circuits, vol. 54, no. 3, pp. 834-844, March 2019, doi: 10.1109/JSSC.2018.2882265.

[1.11] Behzad Rezavi "Design of Analog CMOS Integrated Circuits" McGraw-Hill Series in electrical and Computer Engineering, inc. 1221, publication 2001.

CHAPTER 2

In the previous chapter an overview of the coherent optical system is provided in order to describe the application and its criticisms. Moreover, it describes the TIA and its peculiarities. In the following chapter will be described the implementation of the TIA that we can find in literature, and the most used bandwidth extension techniques.

2.1 Low Noise Transimpedance Amplifier (LNTA)

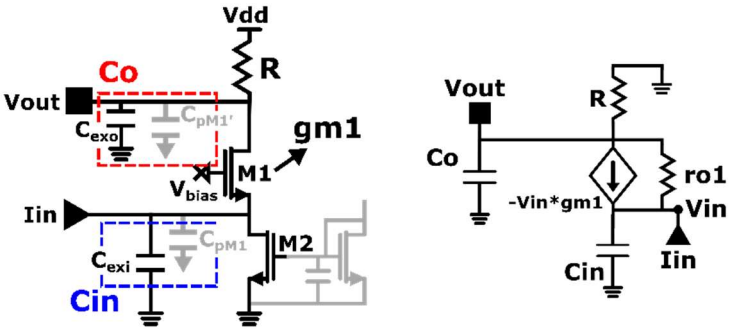
As chapter 1 introduces, the stage that limit more the performances is the first one due to the big input photodiode capacitance that limit the bandwidth, and the gain that has to be high enough to reduce the noise contribution from the next stages. In the literature three main topologies are presented:

- Common Gate (CG)
- Shunt Feedback (SF)
- Regulated Cascode (RGC)

Every topology has its advantages and disadvantages that will be presented in this paragraph.

2.1.1 Common Gate

The principle behind any TIA is to provide a high gain but with a low input impedance in order to compensate the photodiode capacitance. The Common Gate configuration can satisfy this requirement as ideally it shows an input impedance equal to the inverse of the transistor transconductance. The single ended schematic and the small equivalent circuit of a Common Gate is presented in Fig. 2.1.1, where we can identify a current generator (M2), a common gate transistor (M1) and a resistive load (R) [2.1] [2.2].



(a) (b)

Fig. 2.1.1 (a) Single ended schematic of a common gate, (b) small equivalent circuit of a common gate where the output resistance of the current mirror has been neglected.

Where Cin includes the external capacitance as the photodiode, the PAD and the ESD, and the intrinsic input capacitances of the stage, while Co represent the load capacitance and the intrinsic output capacitances of the common gate. The resulting transimpedance gain of the common gate is equal to

$$\frac{Vout}{Iin} = R * \frac{1}{1 + s\left(RCo + \frac{(R + ro_1)Cin}{1 + gm_1 ro_1}\right) + s^2 \frac{CinCoRro_1}{1 + gm_1 ro_1}} \tag{2.1.1}$$

If in first approximation we neglect ro, the gain becomes equal to [2.1]

$$\frac{Vout}{Iin} = R * \frac{1}{(1 + sRCo)(1 + \frac{sCin}{gm_1})} \tag{2.1.2}$$

The gain is directly related to the value of R, while the bandwidth is determined by two real poles, the first at the frequency $f1 = \frac{1}{2\pi}$, and the second in $f2 = \frac{gm_1}{2\pi Cin}$. Considering that C_{in} and C_o are in part dependent to the dimension of M1, we can express the two poles as (2.1.3) and (2.1.4):

$$f1 = \frac{1}{2\pi R C_o} = \frac{1}{2\pi R(Cex_o + \frac{\alpha gm_1}{\omega_T})} \tag{2.1.3}$$

$$f2 = \frac{gm_1}{2\pi C_{in}} = \frac{gm_1}{2\pi(Cex_i + \frac{gm_1}{\omega_T})} \tag{2.1.4}$$

Where respectively Cex_i and Cex_o are the fixed external input and output capacitances, ω_T is the intrinsic cutoff pulse of the MOSFET and α is a technological parameter that represent the ratio of the input and output intrinsic capacitances of the transistor, in the 28nm High Performance Computing Plus (HPC+) technology it is about 0.5. For gm that tends to infinite, $f1$ continues to decrease due to the intrinsic output capacitances of M1 that dominates to the external fixed one, while $f2$ will tend to the intrinsic cutoff frequency of the MOSFET. For the opposite behavior of the two poles we can find an optimum bandwidth that corresponds to the gm_1 that force the two poles to be equal [2.3].

Sackinger decided to determine the relationship between the gain and the bandwidth of a transimpedance amplifier in [2.3]. He observed that this relationship in the common gate results equal to (2.1.5).

$$Gain_{LF} = R_T \leq \frac{\sqrt{\zeta^4 + 6\zeta^2 + 1} - \zeta^2 - 1}{2\zeta} \cdot \frac{A_0 f_A}{2\pi C_{in} BW_{-3dB}^2} \tag{2.1.5}$$

Where $A_0 f_A$ is the gain-bandwidth product of the common gate, equivalent to $(gmR) \cdot \left(\frac{1}{2\pi RC_o}\right)$ where the first term represent A_0 and the second f_A; ζ is theratio between the two poles f_2/f_1. He observed that the common gate solution can achieve high gain with a given bandwidthif ζ is low, it corresponds to a lower frequency input pole respect the output one, but it leads to the sensitivity of the circuit to the photodiode capacitance [2.3].

After the analysis of the bandwidth, it is important also analyze the noise performances, as the first stage influences the noise of the entire TIA. To do that we have to consider also the output resistance (ro_2) of the current mirror, that has been neglected until now, meanwhile, for simplicity, we neglect the output resistance of M1. The total equivalent input current noise is equal to (2.1.6).

$$S_{in} = \frac{In_{in}^2}{Hz} = 4KT\gamma gm_2 + \frac{4KT}{R}\left|\frac{1 + gm_1 ro_2}{gm_1 ro_2}(1 + s\frac{C_{in} ro_2}{1 + ro_2 gm_1})\right|^2 + 4KT\gamma g_m \left|\frac{1 + sro_2 C_{in}}{gm_1 ro_2}\right|^2 \tag{2.1.6}$$

Where gm_2 is the transconductance gain of the current mirror. Usually the frequency $\frac{1+ro_2 gm_1}{2\pi C_{in} r0_2}$ is outside the bandwidth, it means that inside we can identify two white noise contributions dependent to the load resistance and the current mirror, and a color contribution provided by gm that becomes more and more important when the noise of M1 stops to recirculate in the transistor and starts to go inside the input capacitance, in other word it becomes important when the input pole starts to be important [2.1][2.2]. If we take into account that gm_{CS} is dependent to the same current that flows in M1, for a given technology and constant gmOverId we can define β where $gm_2 = \beta gm_1$. Assuming that $ro_2 >> 1$, the equation in of the noise in bandwidth can be approximated to

$$S_{in} \approx 4KT\gamma\beta gm_1 + \frac{4KT}{R} + \frac{4KT\gamma}{gm_1} \cdot \omega^2 C_{in}^2 \tag{2.1.7}$$

This expression suggests that the higher is gm_1, higher is the noise contribution of the current mirror. As [2.1] observes, this technique leads to headroom problem typical of this topology. In fact, the main limit of this stage is the tradeoff between bandwidth, noise and voltage supply. To maximize the input pole, we would like to increase the current, as the gm, but it leads to an higher voltage drop over the resistor R and an higher current mirror noise, at the same time if we want to reduce the resistor to accommodate a smaller supply voltage we degrade also the noise of R. For that reason, this topology usually is considered a possible solution where a large bandwidth but poor noise performances are required [2.1][2.2].

For example, in 28nm HPC+ CMOS technology we can perform a design to optimize the bandwidth of a common gate with a gain of 200Ohm, an input external capacitance of 80fF and output external

capacitance of 40fF and sweeping gm_1 of the common gate MOSFET. The resulting bandwidth and average input noise versus the transconductance gain of M1 are reported in Fig. 2.1.2.

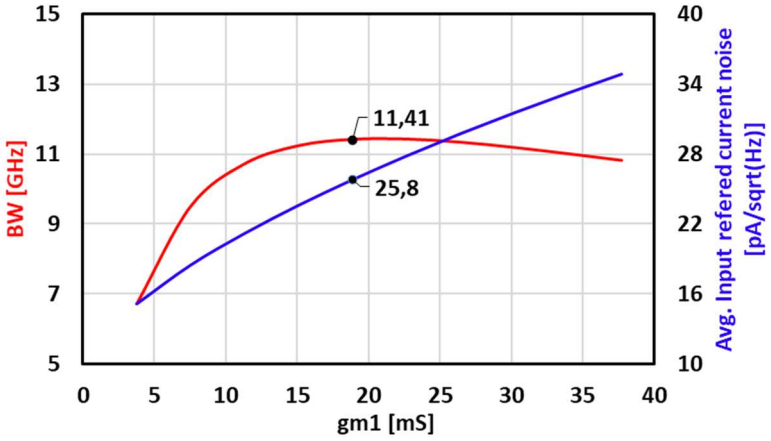


Fig. 2.1.2 Bandwidth and average input current noise versus the transconductance gain of the common gate.

As we expect the bandwidth presents an optimum, around 11GHz, and an equivalent input current noise equal to 25.8 pA/sqrt(Hz). In that condition the power consumption is equal to 4 mW for a supply voltage of 1.55 V. We can analyze the noise contributions in Fig. 2.1.3.

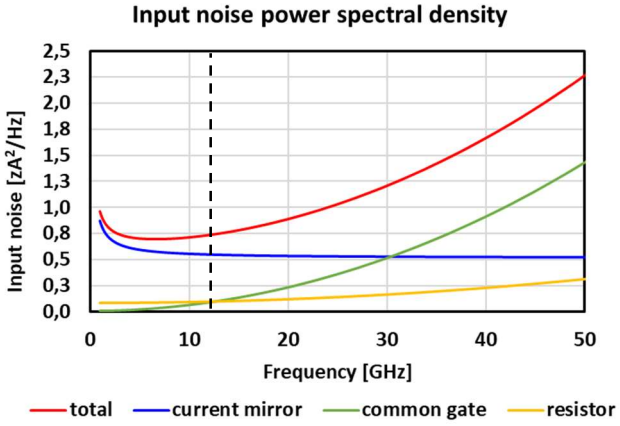


Fig. 2.1.3 Equivalent input noise contributions of the common gate at the bandwidth optimum design.

We can observe that in the bandwidth (highlighted by the black line), as we expect, the dominant noise contribution is provided by the current mirror (blue line). We will see that often in the TIA the first stage has a limited bandwidth due to the photodiode capacitance, but the overall bandwidth is amplified by the next stages by an equalization realized with inductive peaking. In this case the integration of the noise is realized above the bandwidth of the TIA, that is higher respect the LNTA's bandwidth. In this case also the colored M1 noise contribution can become important in the design.

An example of Common Gate in literature is proposed by [2.4] where they proposed a single ended to differential TIA in 0.13 μm BiCMOS designed to operate in a 4-channel 25 Gb/s system. It presents a gain of 76 dBΩ with a bandwidth from 22 kHz to 23 GHz and input-referred noise equivalent to 2.4 μA_{rms}. The schematic of the proposed LNTA is shown in Fig. 2.1.4 [2.4].

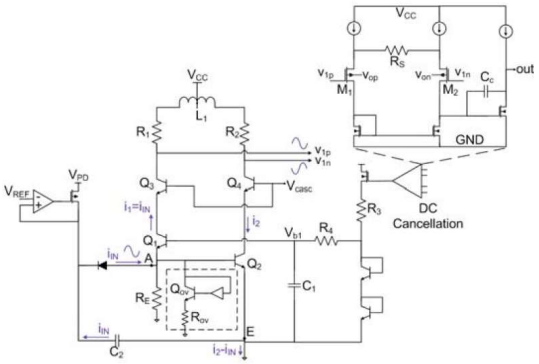
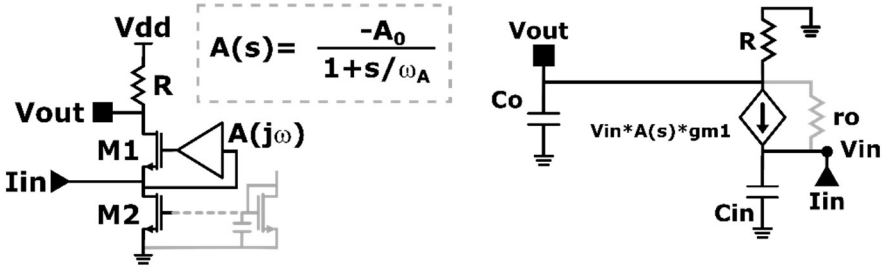


Fig. 2.1.4 First stage circuit and DC cancellation amplifier on the top left [2.4].

In this paper a variant of the standard common gate is proposed where the capacitor C1 is used to bootstrap the transistors Q2 and Q1 in order to mirror any voltage variation on the emitter of Q2 also to the base of Q1, and consequently to its emitter. The impedance from node E to ground has to absorb the difference between the current i_{IN} and i_2. In that way they could to create a differential output signal increasing the power supply rejection respect a traditional common gate [2.4].

2.1.2 Regulated Cascode

The regulated cascode is a closed loop topology that is a derivation of the common gate, in fact it is based on a boosting of the gm by an amplifier. The basic schematic of a regulated cascode and its equivalent small equivalent circuit are represented in Fig. 2.1.5.



(a) (b)

Fig. 2.1.5 (a) Single ended schematic of a regulated cascode, (b) small equivalent circuit of a regulated cascode where the output resistance of the current mirror has been neglected.

If we ignore the output resistance of M1, the new input impedance become equal to $\frac{1}{gm(A_0+1)}$, so in correspondence of the same input impedance of a common gate, M1 can carries less DC current reducing power consumption, headroom limitation and the noise contribution of the current mirror. Analyzing the transfer function of the small equivalent circuit, it is equal to

(2.1.8) $$\frac{V_{out}}{I_{in}} = R \cdot \frac{(1+\frac{s}{\omega_a(A_0+1)})}{(1+sC_oR)(1+s\left(\frac{gm+\omega_aC_{in}}{gm\omega_a(A_0+1)}\right)+s^2(\frac{C_{in}}{gm\omega_a(A_0+1)}))}$$

We can identify that respect the classic common gate we have a zero at $\omega_a(A_0+1)$ that usually is at high frequency [2.3], and two complex poles characterized by $\omega_0 = \sqrt{\frac{gm(A_0+1)}{C_{in}} \cdot \omega_a}$ and $Q = \frac{\sqrt{gm\omega_a(A_0+1)}}{gm+\omega_aC_{in}}$. This results in 3 total poles, where the output pole is pushed at higher frequencies respect than in common gate topology because M1 is smaller, but at the same time we add additional capacitance at the input node by the amplifier. If we want a flat response we need to impose that $Q \leq \frac{1}{\sqrt{2}}$, under this condition the bandwidth becomes smaller than $\sqrt{\frac{gm}{C_{in}}A_0\omega_a}$ [2.5].

Also for the regulated cascode Sackinger performed the transimpedance limit and he determined that is results equal to (2.1.9).

(2.1.9) $$Gain_{LF} = R_T \leq \frac{\sqrt{\zeta^4+6\zeta^2+1}-\zeta^2-1}{2\zeta}\frac{A_0f_A}{2\pi C_{in}BW_{-3dB}^2} \cdot \frac{A_0f_A}{\zeta BW_{-3dB}}$$

This expression is equal to (2.1.5) except the factor $\frac{A_0f_A}{\zeta BW_{-3dB}}$ that changes the relationship between the gain and the BW from the square to the cube, ant it also improves the relation by a factor of A_0f_A/ζ. This is justified by the lower output capacitance of the common gate [2.3].

As we see from (2.1.8) also the finite bandwidth of the amplifier has an important role, as it determines when the input impedance start to increase. This inductive behavior can help to enhance the bandwidth [2.6]. If we look to the noise performances the equivalent input current noise expression becomes equal to (2.1.10).

(2.1.10) $$S_{in} = 4KT\gamma g_{mCS} + \frac{4KT}{R}\left|\frac{1}{1+\frac{s}{\omega_{zero}}}(1+\frac{s}{Q\omega_0}+\left(\frac{s}{\omega_0}\right)^2)\right|^2 + \frac{4KT\gamma}{g_{mA}}\left|\frac{sA_0C_{in}}{1+A_0}(\frac{1}{1+\frac{s}{\omega_{zero}}})\right|^2 + 4KT\gamma g_m\left|\frac{sC_{in}}{gm(1+A_0)}(\frac{1+\frac{s}{\omega_a}}{1+\frac{s}{\omega_{zero}}})\right|^2$$

As we did for the common gate, we can consider the noise only in the bandwidth and assume that $A_0 \gg 1$, the expression becomes equal to

(2.1.11) $$S_{in} \approx 4KT\gamma g_{mCS} + \frac{4KT}{R} + \frac{4KT\gamma}{g_{mA}} \cdot \omega^2 Cin^2 + \frac{4KT\gamma}{A_0g_m} \cdot \frac{\omega^2C_{in}^2}{A_0}(1+\frac{\omega^2}{\omega_a^2})$$

As we introduced before, the transconductance gain of the current mirror can be reduced respect the common gate solution, thanks to the lower DC current, so the white noise can be improved respect the previous topology, but at the same time we add a colored noise of the amplifier and also the noise of M1 increase with ω^4 instead of ω^2 after the amplifier pole. We can achieve a noise improvement respect to the common gate solution if the white noise is reduced more than the increment of the colored noise.

In literature we can find some example that use this topology, as for example the work presented by Sung Min Park and Hoi-Jun Yoo in 2004 at IEEE Journal of Solid State Circuit [2.7]. They presented an

transimpedance input stage realized by 0.6μm digital CMOS for Gigabit Ethernet applications, it is based on a regulated cascode in order to boost the input transimpedance and compensate the capacitance of the photodiode equal to 500 fF. The measured gain is equal to 58 dBΩ with a bandwidth of 950 MHz. The average noise current spectral density is equal to 6.3 pA/√Hz, while the total dissipated power is 85 mW for 5 V of voltage supply. Fig. 2.1.6 represents the schematic of the TIA where we can identify the photodiode that pushes the current in a common gate (M1-R1), whose gm is boosted by an N-only amplifier (MB-RB). As the authors explain, the crucial point of this topology is the stability [2.7].

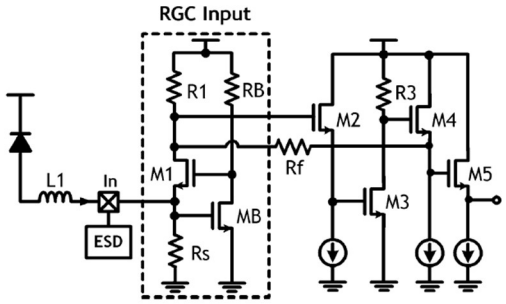


Fig. 2.1.6 Circuit scheme of the entire TIA [2.7]

Another example of regulated cascode is provided by S. Bashiri Amid, C. Plett and P. Schvan in 2010 at Bipolar/BiCMOS Circuit and Technology Meeting [2.6]. They realized a regulated cascode in 0.13 μm BiCMOS technology for a photodiode with an equivalent capacitance equal to 300 fF. The resulting bandwidth is equal to 28GHz for a differential gain of 53.6 dBΩ and an integrated input referred noise of 6.11 μArms. The total chip consumes 110 mW with 3 V of voltage supply. The schematic of the first stage is shown in Fig. 2.1.7. where an active load (M1-M2) is used instead of the resistance to reduce headroom issues and noise contribution [2.6].

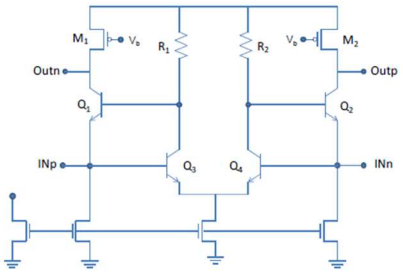


Fig. 2.1.7 Schematic of the fully differential regulated cascode with active load [2.6]

2.1.3 Shunt Feedback

Most of the LNTA design in literature are realized with the shunt feedback topology, to avoid the current mirror noise contribution and overpass the headroom limit. To understand we have to analyze its schematic that is presented in Fig. 2.1.8.

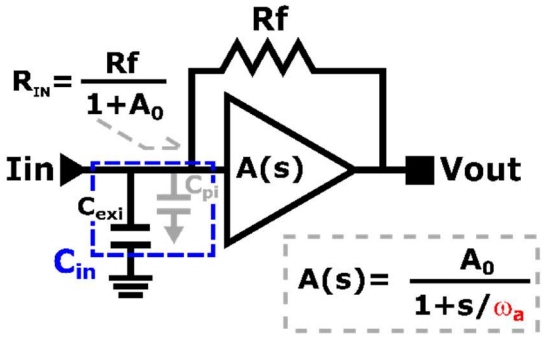


Fig. 2.1.8 Generic block diagram of a shunt feedback amplifier

The shunt feedback topology has the purpose to provide a gain equal to the feedback resistance, but due to the loop gain the input impedance is reduced to a value equal to $\frac{R_f}{A_0+1}$. If we consider the limited bandwidth of $A(j\omega)$, the transfer function becomes equal to

$$\frac{V_{out}}{I_{in}} = -\frac{R_f A_0}{A_0+1} * \frac{1}{1+\frac{s}{A_0+1}\left(\frac{1}{\omega_a}+R_f C_{in}\right)+\frac{R_f C_{in}}{\omega_a(A_0+1)}s^2} \tag{2.1.12}$$

Where C_{in} is the sum of the photodiode capacitance (C_{PD}) and the parasitic input capacitance of the amplifier (C_{Pin}). The gain can be approximated to R_f, if A_0 is higher than 1, and the transfer function is characterized by two complex poles at the angular frequency equal to $\omega_o = \sqrt{\frac{\omega_a(A_0+1)}{R_f C_{in}}}$ and with $Q = \frac{\sqrt{R_f C_{in}\omega_a(A_0+1)}}{\omega_a R_f C_{in}+1}$ [2.1][2.3]. If we want to avoid any peaking, we have to set the quality factor equal to $1/\sqrt{2}$, for ω_a higher than $\frac{2A_0}{R_f C_{in}}$, the resulting cutoff frequency becomes (2.1.13) [2.1].

$$f_{-3dB} = \frac{1}{2\pi} \cdot \frac{\sqrt{2A_0(A_0+1)}}{R_f C_{in}} \tag{2.1.13}$$

Computing the equivalent input noise in the bandwidth, it results [2.1]

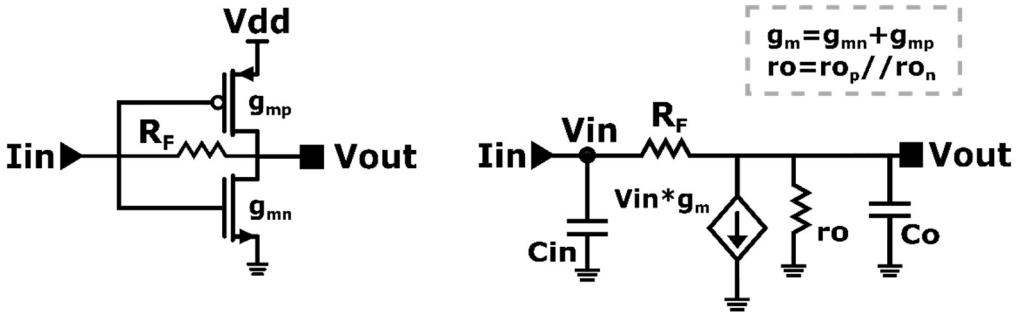
$$S_{in} = \frac{4KT}{R_f} + \frac{V_{InNA}^2}{Hz}\left|\frac{1+sCinR_f}{R_f}\right|^2 \tag{2.1.14}$$

Where also in this case we can identify a white noise produced by the feedback resistor, and a colored noise produced by the amplifier, in particular V_{InNA}^2 is the equivalent noise of the amplifier at its input. The main advantage of this topology is the feedback resistance that can curry no dc current, so it can be increased without headroom issues, of course it determines a tradeoff with the bandwidth. Another advantage is produced by the feedback: even if the white noise corresponds to a resistance R_f, the input impedance corresponds to $R_f/(A_0+1)$ [2.1][2.2].

If it is not possible to push the pole of the amplifier (ω_a) at high frequency, the expression of f_{-3dB} becomes more complicate and the bandwidth is reduced, in fact usually the quality factor is selected in order to have a little boost of bandwidth using the peaking [2.1].
Also for the shunt-feedback Sackinger analyzed the transimpedance limit that results equal to (2.1.15).

(2.1.15) $$Gain_{LF} = R_T \leq \frac{A_0 f_A}{2\pi C_{in} BW_{-3dB}^2}$$

Where $A_0 f_A$ is the gain-bandwidth product of the amplifier A.
Comparing (2.1.5) and (2.1.15) he observed that for most of the topologies the relation between the gain and the bandwidth is not linear, but quadratic, and that the limit is always dependent to $A_0 f_A / C_{in}$. It depends mainly to the photodiode capacitance and to $A_0 f_A$ that is in some way proportional to the technology cut-off frequency [2.3]. If we compare the limits of the common gate with the shunt feedback, he highlighted that their ratio depends only to ζ (ratio between the input and output pole of the common gate). If we plot the ratio between the two transimpedance limits we can identify an optimum for the common gate transimpedance expression that exactly corresponds when the two real poles are equal as we already explained before and in this optimum the shunt feedback achieves a transimpedance gain higher of a factor 2.44 respect the common gate solution.
Another possible realization of the shunt feedback LNTA is presented in Fig. 2.1.9.



(a) (b)

Fig. 2.1.9 Inverter based shunt feedback circuit (a) ant its small signal equivalent circuit (b)

Its study is less intuitive than the one presented previously because the amplifier stage in this case has a high output impedance. We can study this circuit with two methods:

- Closed Loop study
- Open Loop study

The Closed Loop Transfer function results equal to

$$\frac{V_{out}}{I_{in}} = \frac{r_o(R_f g_m - 1)}{1 + g_m r_o} \cdot \frac{1}{1 + s\left[\frac{C_o r_o}{1 + g_m r_o} + \frac{C_{in}(R_f + r_o)}{1 + g_m r_o}\right] + \frac{s^2 C_o C_{in} r_o R_f}{1 + g_m r_o}} \tag{2.1.16}$$

If $R_f g_m \gg 1$ and $g_m r_o \gg 1$ the expression becomes equal to

$$\frac{V_{out}}{I_{in}} \approx R_f \cdot \frac{1}{1 + s\left[\frac{C_o}{g_m} + \frac{C_{in}(R_f + r_o)}{g_m r_o}\right] + \frac{s^2 C_o C_{in} R_f}{g_m}} \tag{2.1.17}$$

Where the gain is R_f and we can identify two complex poles at the angular frequency $\omega_o = \frac{g_m}{C_o C_{in} R_f}$ with $Q = \frac{r_o\sqrt{C_o C_{in} R_f g_m}}{r_o C_o + C_{in}(R_f + r_o)}$.

The second method studies the open loop gain of the circuit. To compute it we have to refer not to the Return Ratio, that is achieved inserting a signal at the input of the gate of the MOSFET, but to the Bode Closed Loop Gain (Gloop) that we can get from the Y matrix [2.8][2.9]. To simplify the computation of Gloop we can use an equivalent circuit that has exactly the same transfer function, but that explicits the direct and feedback gains. The equivalent circuit is presented in Fig. 2.1.10.

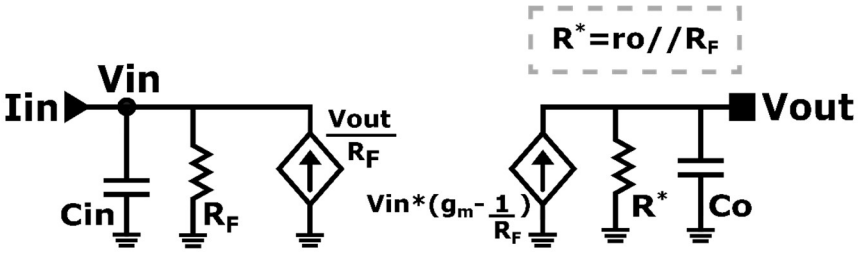


Fig. 2.1.10 Equivalent circuit of the inverter based shunt feedback

As for any closed loop circuit the bandwidth of the closed loop gain is strictly related to the unity gain bandwidth of the Gloop, whose expression is

$$Gloop = (g_m - \frac{1}{R_f})(\frac{R_f r_o}{R_f + r_o}) \cdot \frac{1}{1 + sC_o \cdot \frac{R_f r_o}{R_f + r_o}} \cdot \frac{1}{1 + sC_{in} R_f} \tag{2.1.18}$$

We can identify two poles respectively at $p_1 = \frac{1}{C_{in} R_f}$ and $p_2 = \frac{R_f + r_o}{C_o r_o R_f}$ with a low frequency gain $G_{LF} = (g_m - \frac{1}{R_f})(\frac{R_f r_o}{R_f + r_o})$. We can rewrite the Gloop replacing $r_o = \frac{A_o}{g_m}$; $C_{in} = C_{xi} + \frac{g_m}{\omega_T}$ and $C_o = C_{xo} + \beta\frac{g_m}{\omega_T}$ where $A_o$is the intrinsic gain of the transistors, ω_T is the cut off pulse of the transistors, β is the ratio between the output and the input intrinsic capacitances of the MOSFETs assuming that the PMOS and the NMOS have the same technological parameters in the selected bias conditions, and C_{xi}, C_{xo} are respectively the fixed input and output capacitances of the stage. Under this condition the Gloop becomes equal to

(2.1.19) $$Gloop = (g_m - \frac{1}{R_f})(\frac{R_f A_o}{g_m R_f + A_o}) \cdot \frac{1}{1 + s(C_{xo} + \beta \frac{g_m}{\omega_T}) \cdot \frac{R_f A_o}{g_m R_f + A_o}} \cdot \frac{1}{1 + s(C_{xi} + \frac{g_m}{\omega_T})R_f}$$

The closed loop bandwidth is strictly related to the unity gain frequency of the Gloop, to determine if an optimum for different g_m exists, we have to understand how G_{LF}, p_1 and p_2 behave in function of it. Usually the dominant pole is p_1 and the cut off frequency relative of only this pole can be obtained as the product of the gain and p_1:

(2.1.20) $$\omega_{T1} = (g_m - \frac{1}{R_f})(\frac{R_f A_o}{g_m R_f + A_o}) \cdot \frac{1}{(C_{xi} + \frac{g_m}{\omega_T})R_f}$$

If we compute the derivative $\frac{d\omega_{T1}}{dg_m}$ we obtain a fraction with a second order polynomial expression of g_m at the numerator and a square expression at the denominator, to find the sign of the derivative, we can analyze only the numerator. The expression of the numerator results equal to

(2.1.21) $$N = A_o^2 R_f^2 C_{xi} + \frac{A_o^2 R_f}{\omega_T} + A_o R_f^2 C_{xi} + g_m \left(\frac{2A_o R_f^2}{\omega_T}\right) - g_m^2 (\frac{R_f^3 A_o}{\omega_T})$$

Where the Δ of the second order equation is

(2.1.22) $$\Delta = \frac{A_o^2 R_f^4}{\omega_T}(A_o + 1)(\frac{1}{\omega_T} + R_f C_{xi})$$

All the parameters involved in Δ are positive, it means that exist two values of g_m that delimit the area where $\frac{d\omega_{T1}}{dg_m}$ is positive. If we solve the homogeneous equation $N = 0$, we find that these two values are equal to

(2.1.23) $$g_{m_{1,2}} = \frac{1 \pm \sqrt{(A_o + 1)(1 + \omega_T R_f C_{xi})}}{R_f}$$

Due to the fact that $(A_o + 1)(1 + \omega_T R_f C_{xi})$ is always higher than 1, it means that the solution of the inequality $N > 0$ is $g_m < g_{m_2} = \frac{1+\sqrt{(A_o+1)(1+\omega_T R_f C_{xi})}}{R_f}$. ω_{T1} optimum corresponds to $g_m = g_{m_2}$.
If $p2$ is inside the unity gain frequency of the Gloop, we have to take into account its dependency to g_m. $p2$'s expression results equal to

(2.1.24) $$p_2 = \frac{g_m R_f + A_o}{R_f A_o \left(C_{xo} + \beta \frac{g_m}{\omega_T}\right)}$$

Also in this case we can compute the derivative respect g_m and it results that $p2$ is monotonic, in particular if $R_f C_{xo} > \frac{\beta A_o}{\omega_T}$ it is growing, otherwise it is descending. The value of $p2$ also limits the stability of the feedback.

About the equivalent input noise, if $R_f g_m \gg 1$, it can be approximated to

$$S_{in} = \frac{4KT\gamma}{g_m R_f^2} \cdot \left|1 + sC_{in}R_f\right|^2 + \frac{4KT}{R_f} \cdot \left|1 + \frac{sC_{in}}{g_m}\right|^2 \tag{2.1.25}$$

If $\frac{g_m}{C_{in}}$ is outside or delimit the bandwidth, inside it we can identify a white noise contribution produced by the feedback resistance and a colored noise of the transistor's channel that is reduced by the value of R_f. If we have an additional external load resistance R_L, we should also add its contribution equal to $S_{in_{RL}} = \frac{4KT}{R_L} \cdot \left|\frac{1+sC_{in}R_f}{g_m R_f}\right|^2$.

If the external input capacitances are dominant to the intrinsic ones of the MOSFETs, the noise can only improve for high transimpedance gain, but for very high g_m the input intrinsic capacitances increase as the colored noise, it means that the noise can improve for high g_m only if the white noise reduction is higher than the increment of the colored noise.

As we did for the common gate, also for the shunt feedback we can evaluate the bandwidth optimum for a technology of 28nm HPC+ CMOS. To do it, we fixed a gain equal to 200Ohm, an input external capacitance of 80fF and output external capacitance of 40fF. The optimization of the bandwidth and the average input current noise in function of the transconductance are shown in Fig. 2.1.11.

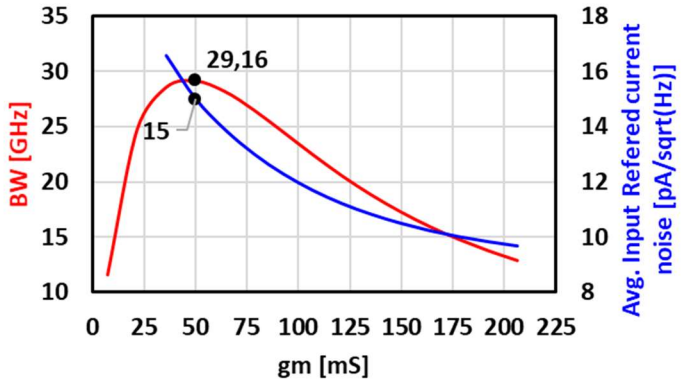


Fig. 2.1.11 Bandwidth and average input referred noise of an inverter based shunt feedback in function of the gm

The plot confirms the optimum bandwidth of 29 GHz in correspondence of a transconductance equal to 50 mS, the equivalent input current noise results equal to 15 pA/sqrt(Hz) with a total power consumption of 5.3 mW and a supply voltage equal to 1.2 V. In this condition the phase margin of the loop results 67.42 ° with a peaking less than 1 dB. Fig. 2.1.11 explicit the noise spectral density for the 3 contributions at the optimum bandwidth.

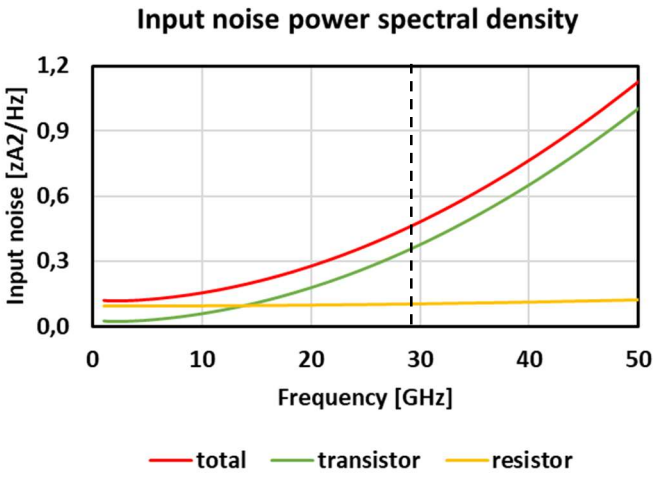


Fig. 2.1.12 Equivalent input noise contributions of the shunt feedback at the bandwidth optimum design.

We can observe in Fig. 2.1.12 that at low frequency the noise is dominated by the resistive contribution, while at higher frequency the channel noise of the MOSFET becomes dominant. As we already said for the common gate, if we get a higher bandwidth with the next stages by equalization, the colored noise of the transistor becomes more and more significant. Also in this case the bandwidth of the LNTA is highlighted by the dashed black line.

Also [2.10] performed a very interesting analysis to understand the performances of the shunt feedback topology. In particular, the authors evaluated the bandwidth and input referred current noise density in function of the power consumption that was varied, keeping constant theratio between the Pmos and the Nmos and changing the number of fins, fingers and multiplicity. They studied three cases: (a) without any load, (b) with the input and output load including the bonding inductance of the photodiode, (c) using the series inductive peaking. Fig. 2.1.13 shows the respectively schematic and plots where we can observe that the case (c) produced a bandwidth enhancement about 70% at 3 mW of power consumption and a noise equal to 11 pA/√Hz.

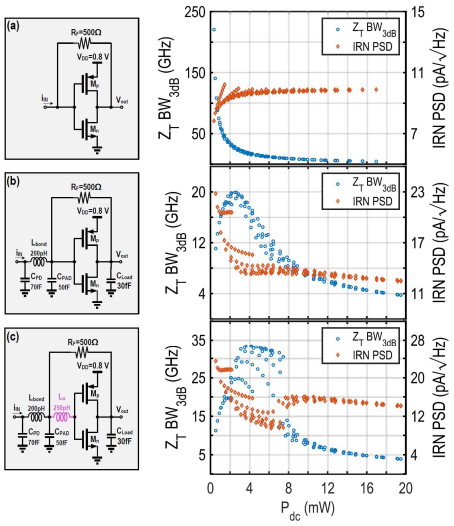


Fig. 2.1.13 Schematic of the Shunt Feedback TIA with its bandwidth and noise performances without any load (a), loaded with the Photodiode and an output capacitance (b) and with series inductive peaking (c) [2.10]

In literature there are many example of shunt feedback, here eight examples are reported.
The first example is proposed by Gertjan Coudyzer, Peter Ossieur, Johan Bauwelinck and Xin Yin in the Journal of Solid State Circuit in 2020 realized in 0.25 μm BiCMOS technology [2.11]. The schematic of the transimpedance amplifier is reported in Fig. 2.1.14 where we can identify in the grey rectangle the photocurrent monitor to absorb the photodiode current.

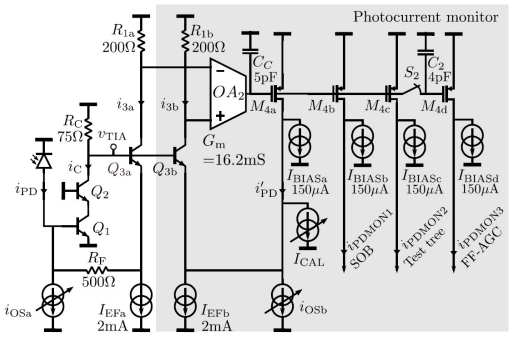


Fig. 2.1.14 Transimpedance amplifier and photocurrent detector [2.11]

In this example the feedback is realized by a resistance of 500 **Ω** and a follower (Q3a) used to decouple the amplifier and the resistive load. To improve the bandwidth of the shunt feedback they studied the closed loop transfer function identifying a secondary pole at the output of the amplifier (Q1, Q2, R_C) that is additionally loaded by the emitter follower. This transistor adds another pole at high frequency that degrades the phase margin of the loop, for that reason they decided to reduce the value of Rc resistance that forced to increase the power consumption. The resulting feedback unity gain product was of 5.5 GHz with a bandwidth closed loop transfer function of 26.4 GHz. Fig. 2.1.15 shows the simulated Gloop and closed loop transfer function of only the first stage of the system [2.11].

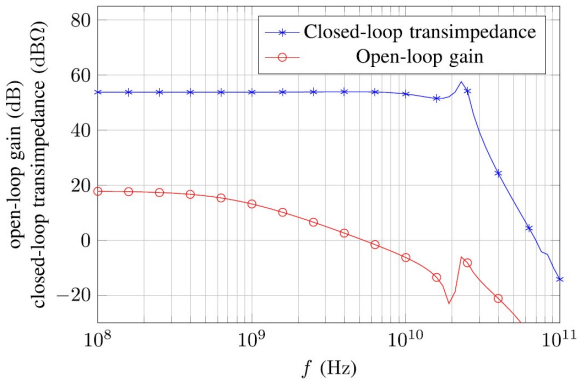


Fig. 2.1.15 Simulated loop gain and closed loop transimpedance gain of the LNTA [2.11]

After the transimpedance amplifier, the TIA is composed by three stages of post amplification and an output buffer realized for 50 **Ω** matching. The overall integrated system is represented in Fig. 2.1.16.

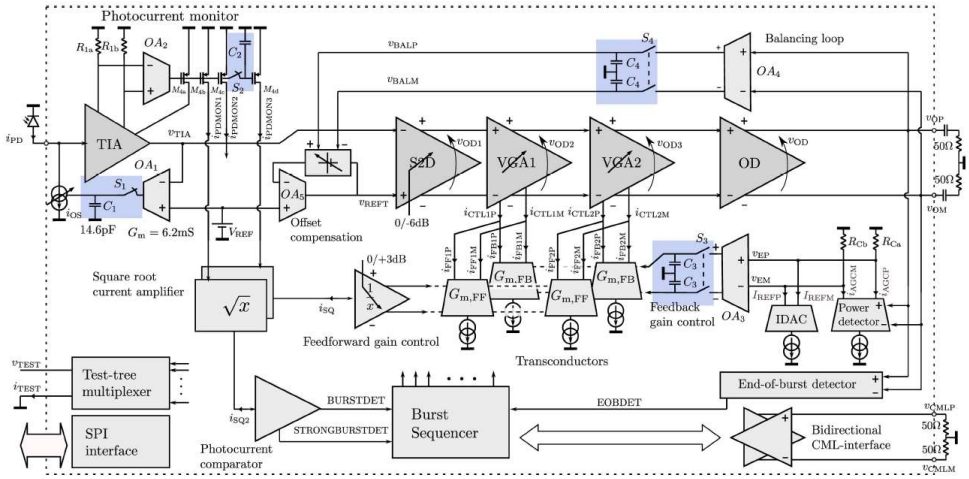


Fig. 2.1.16 Complete architecture of the receiver [2.11]

In the image we can identify a combination of feed forward and feedback for the automatic gain (AGC) and offset control (AOC), they are realized by fast detection of rms- and average- values of the output bursts in order to avoid peaking detector which is difficult to realize at high frequency. The resulting maximum settling time is 82.7 ns. The measured TIA presents an overall bandwidth around 18 GHz with an equivalent input current noise at maximum gain (70 dBΩ) of 1.1 μA_{rms}. The dynamic range is of 21.6 dB for NRZ modulation and 15.8 dB for PAM 4 and the final bit rate results equal respectively to 25 Gbit/s and 50 Gbit/s for NRZ and PAM 4. The overall power consumption is 280 mW for a supply voltage of 2.5 V [2.11].

The second proposed example from literature is a work of Ke Li et al. that was presented on the Optical Express in 2020 [2.12]. The authors propose a differential transimpedance amplifier in 28 nm CMOS co-designed with a balanced photodetector. The optical receiver circuit is shown in Fig. 2.1.17 [2.12].

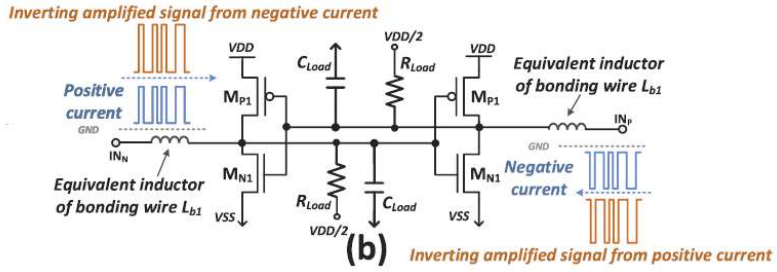


Fig. 2.1.17 Schematic of the LNTA or Cross Coupled Inverter topology [2.12]

The authors realized a new topology based on positive feedback, the idea is that the transistors MN1 and MP1 could be sized with low dimensions saving power because at low frequency the loop absorbs the differential currents from the photodiode producing an inverted and amplified signal on the other side, this allows to reuse the amplifier's transconductance. At high frequency the capacitor Cload creates a resonance peak with the bonding inductance Lb1 and expands the bandwidth. They explain that to avoid latching from the structure, a resistance should be introduced (Rload). In the final implementation they

decided to replace the resistance with a small inverter based shunt feedback amplifier (MP2, MN2, RF) as Fig. 2.1.18 shows. [2.12]

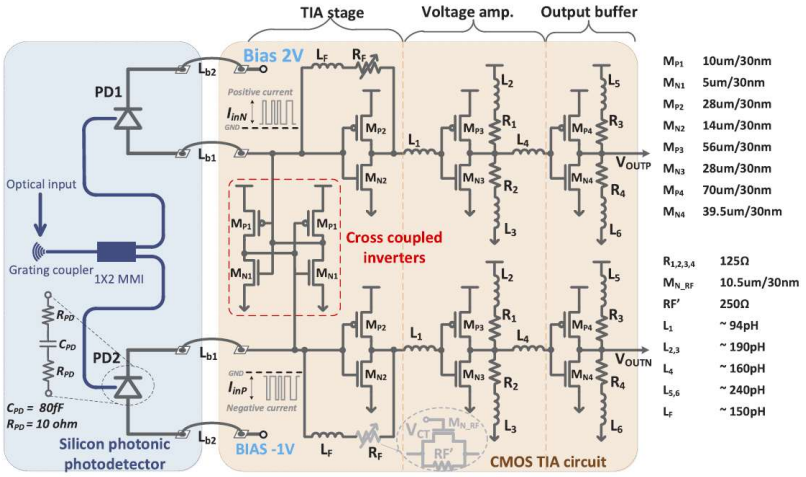


Fig. 2.1.18 Schematic of the receiver [2.12]

The simulation of the transfer function of the TIA is in Fig. 2.1.19 [2.12].

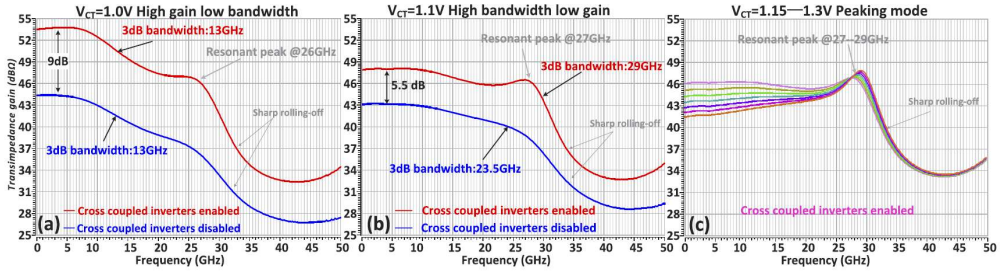


Fig. 2.1.19 Transimpedance amplifier of the TIA with the activated (red curve) and deactivated (blue curve) cross coupled inverters at maximum gain (a) and minimum gain (b); peaking mode (c) [2.12]

As the authors explain that the presence of the cross coupled inverter enhance the gain of 9 dB at maximum gain and of 5 dB at low gain where also a 3-dB bandwidth enhancement is produced from 23.5 GHz to 29 GHz, overpassing the traditional tradeoff between bandwidth and gain. The measured results present a TIA that is able to produce an output signal at 54 Gb/s with a power consumption of 53.1 mW and 0.98 pJ/bit. About the noise performances Fig. 2.1.20 presents the SNR for different data rate. In the article is explained that the bigger limitation in the SNR is the limited bandwidth, in fact there is an improvement in the SNR activating the peaking mode [2.12].

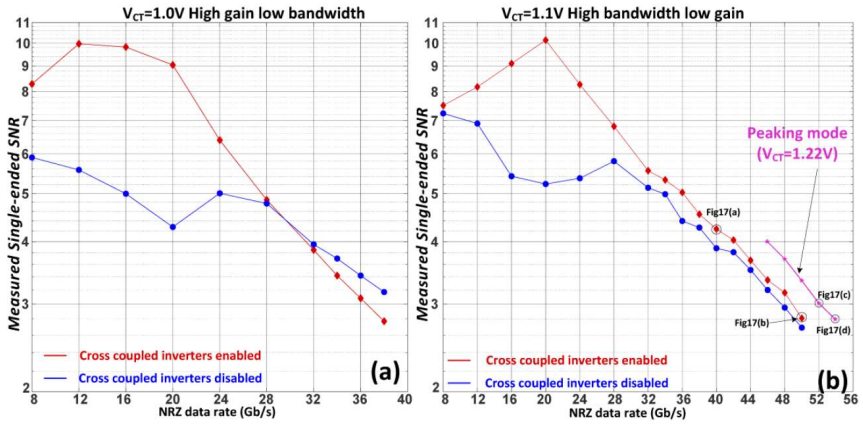


Fig. 2.1.20 Single ended measured SNR with an input OMA of -8.7dBm at high gain mode (a), low gain mode and peaking mode (b) [2.12]

Another interesting work is the one proposed by Kadaba R. Lakshmikumar et al. in 2019 at the Journal of Solid State Circuit realized in 16 nm FinFet [2.13]. They proposed an alternative way to boost the bandwidth of a Shunt Feedback. Usually the most common way is the one presented in Fig. 2.1.21 (a) where the inductance is used before the loop, while they suggest to put the inductance inside the loop in front of the amplifier to shield the transistor capacitance of the inverter (Fig. 2.1.21 (b)) [2.13].

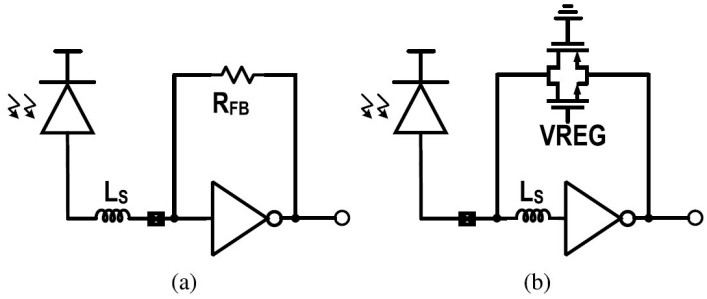


Fig. 2.1.21 (a) Traditional shunt feedback schematic, (b) proposed solution [2.13]

The proposed solution represents the one with the higher bandwidth respect the traditional solution or without any inductor. The scheme of the overall TIA is presented in Fig. 2.1.22where we can identify a stage to convert from single ended to differential signal, three programmable gain amplifier, an output buffer and two DC feedback [2.13].

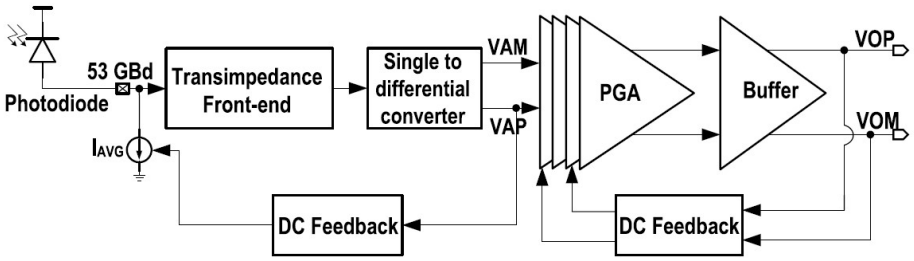


Fig. 2.1.22 Schematic of the complete TIA [2.13]

The measured result presents a bandwidth of 27 GHz, an average input referred noise of 16.7 pA/√Hz and a distortion equal to 1.8 % with an output swing of 600 mV. The transimpedance gain can varies from 78 dBΩ to 57 dBΩ. The design was realized form PAM 4 modulation; indeed, it can support communications with 106.25 Gb/s rate with a power consumption of 60.8 mW [2.13].

The fourth shunt feedback example is a work presented in Journal of Solid State Circuit in 2019 by Mostafa G. Ahmed, Tam N. Huynh, Christopher Williams, Yong Wang, Pavan Kumar Hanumolu and Alexander Rylyakov [2.14]. They designed a linear TIA in 0.13 μm SiGe BiCMOS with a block diagram equal to Fig. 2.1.23 where an Input DC Current Compensation (IDDC), an Analog Gain Control (AGC) and DC Offset Compensation (DOC) are implemented [2.14].

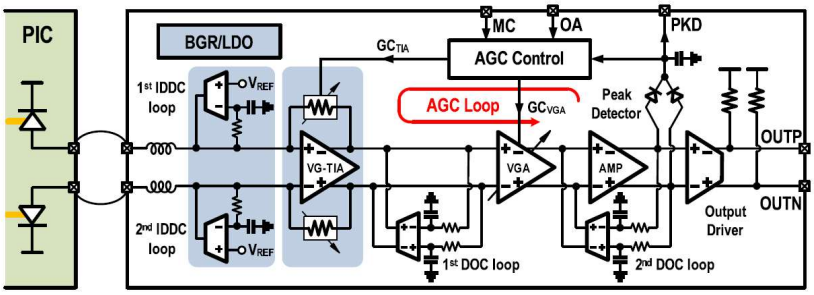


Fig. 2.1.23 Scheme of the TIA [2.14]

The schematic of the proposed Variable Gain Transimpedance Amplifier is shown in Fig. 2.1.24.

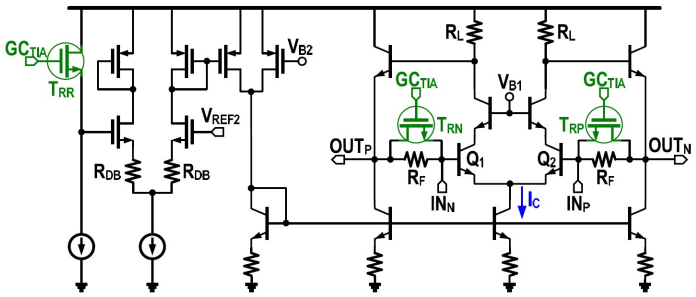


Fig. 2.1.24 Circuit of the LNTA [2.14]

They realized a fully differential LNTA based on a shunt feedback solution with an amplifier Q1-Q2 followed by an emitter follower where there is the output, with this technique the input capacitance of the next stage is isolated to the loop allowing a higher bandwidth. The variability of the gain was realized with a NFET in parallel to the feedback resistance where the maximum value is selected for a given bandwidth, while the minimum is limited by the distortion [2.14]. To prevent TIA's phase margin degradation, with the feedback resistance also the feed forward amplifier gain has to be changed in order to keep constant their ratio. To do that the authors decided to vary Ic together with the variable feedback resistance. The measured transimpedance gain of the entire TIA can varies from 73 dBΩ to 30 dBΩ with a 3-dB bandwidth higher than 27 GHz. The average input noise results equal to 20 pA/√Hz and the distortion is equal to 2.2 % at minimum gain with an output swing of 500 mVppd. The total power

consumption is of 313 mW and the archived data rate is of 200 Gb/s for a dual polarization 16 QAM [2.14].

I would like to present a work of Hao Li, Ganesh Balamurugan, James Jaussi and Bryan Casper that was presented at the European Solid State conference in 2018 [2.15]. They designed a TIA for PAM 4 in 28 nm bulk CMOS technology composed by a single ended to differential TIA, a multiple stage Variable Gain Amplifier and a Continuous Time Linear Equalizer driver. The first stage, that has the purpose to realize the transimpedance amplification, is inverter based in order to accommodate a low supply voltage of 0.9 V provided by an integrated regulator. In the stage a passive peaking network was realized in order to enhance the bandwidth: a shunt inductance (Lshunt1) that increase the transimpedance gain at high frequency in order to compensate the roll off; an additional Lshunt2 in series to R_{L1} in order to increase the open loop gain and two series inductances (Lseries1 and Lseries2) to distribute the capacitances. The schematic of the described LNTA is presented in Fig. 2.1.25 [2.15].

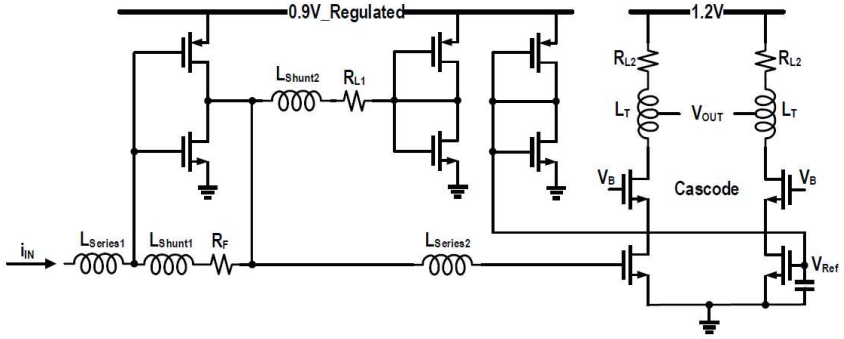


Fig. 2.1.25 Schematic of the shunt feedback transimpedance amplifier [2.15]

The measured Zt transfer function with an equivalent photodiode capacitance of 70 fF and bonding inductance of 200 pH presents a bandwidth of 60 GHz with a gain of 65 dBΩ. The equivalent input noise results equal to 19.3 pA/√Hz with a power consumption of 107 mW. The authors realized also a simulation of 112 Gb/s PAM4 that presented a sensitivity of -5.1 dBm with a BER of 2x10-4 resulting in an energy per bit equal to 0.96 pJ/bit [2.15].

A fully differential TIA with average input referred noise less than 10 pA/√Hz was proposed by Iria García López, Ahmed Awny, Pedro Rito, Minsu Ko, Ahmet Cagri Ulusoy and Dietmar Kissinger in 2018 always on the Journal of Solid State Circuit [2.16]. They realized a TIA in 130 nm SiGe BiCMOS technology for 100 Gb/s for PAM 4 modulation. The measured TIA presents a bandwidth of 66 GHz with gain of 65 dBΩ and a noise of 7.6 pA/√Hz. The schematic of the TIA is presented in Fig. 2.1.26 [2.16].

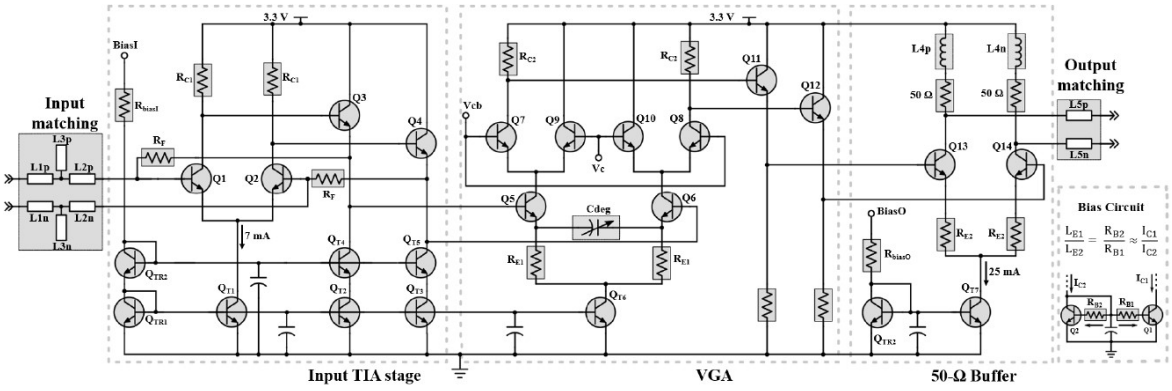
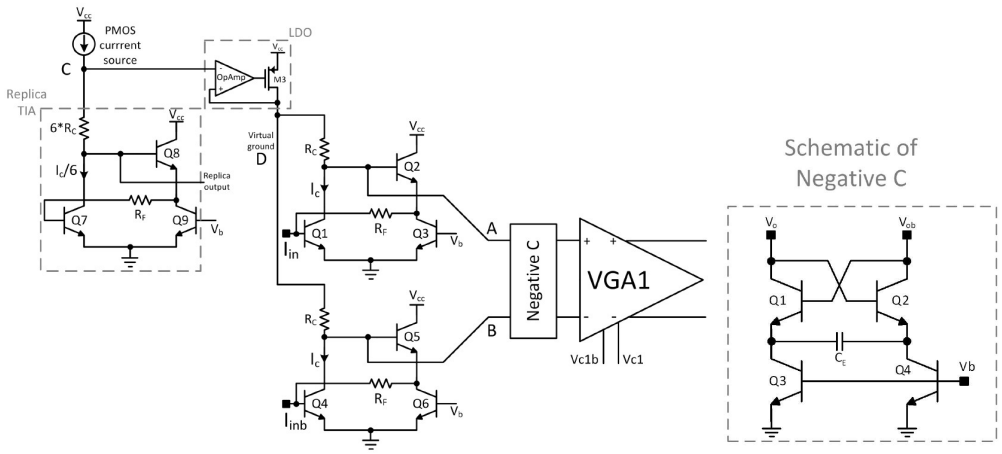


Fig. 2.1.26 Schematic of the complete TIA [2.16]

The realized shunt feedback for the input stage is obtained with an amplifier followed also in this case by an emitter follower that is connected to the feedback resistance. After the first stage we find the variable gain amplifier that is a degenerated common emitter in order to improve linearity. The resistance of degeneration is in parallel to a variable capacitance that introduces a zero in order to control the peaking of the transfer function at middle frequencies, while the variability of the gain is realized with a current steering. The enhancement of the bandwidth is provided by the 50-**Ω** buffer with a shunt inductance (L4p-n). The power consumption results equal to 150 mW with a supply voltage of 3.3 V [2.16].

The last example that I would like to present is a work presented at the International Solid State Circuit Conference in 2016 by Ahmed Awny et al. [2.17]. They realized a TIA in 0.13 μm BiCMOS for coherent optical modulation where the schematic of the shunt feedback is presented in Fig. 2.1.27 (a) [2.17].



(a) (b)

Fig. 2.1.27 Schematic of the transimpedance amplifier (a) and of the negative capacitance (b) [2.17]

The shunt feedback topology was selected for better noise performances respect the common base solution. The innovation of the presented work was the use of a replica of the transimpedance amplifier to determine the bias current of the LNTA acting to the supply voltage by a LDO. The input of the next stage is taken from the base of the emitter follower to accommodate headroom issues of the VGA. In this example the bandwidth extension is realized

by inductive peaking and a negative capacitance (Fig. 2.1.26 (b)). The other stages are two variable gain amplifier realized by degenerated Gilbert Cell, and an emitter degenerated output buffer design to drive 100 Ω differential resistance with 900 mVppd. The measured TIA presents a 3-dB bandwidth of 53 GHz and maximum gain of 80 dBΩ. The linearity is measured with a tone at 1 GHz and corresponds to a THD of 5 % for 900mVppd and 3% for 600mVppd for an input current of 3mApp. The average input referred noise results equal to 24.86 pA/√Hz. Finally they tested the TIA with an optically modulated QPSK demonstrating a possible transmission data up to 128 Gb/s. The power consumption results equal to 277 mW for 3.3 V supply voltage [2.17].

2.2 Variability of the Gain

After the conversion of the signal into voltage by the LNTA, the VGAs are used to amplify it more and provide some variation of the gain in order to avoid the saturation of the Analog Digital Converters that follow the TIA. Often not only the Amplifiers can provide variability of the gain, but also the LNTA can in order to have smaller output swing and improve distortion at low gain condition, when the input current can be very high.
In literature there are two main topologies to perform gain variability: one based on the Gilbert cell or Current steering and one based on variable resistance/transconductance.

2.2.1 Gilbert Cell and Current Steering VGA

The basic schematic of the Gilbert Cell is reproduced in Fig. 2.2.1. As we can observe it is based on a common source and a cascode, where in part is straight and in part is crossed. The straight cascode pushes the current to the output, while the crossed cascode carries the currents with opposite sign to the same output in order to subtract the current signal. Controlling the ratio between these two currents by the gate voltages, you can control how much signal current is subtracted and change the gain.

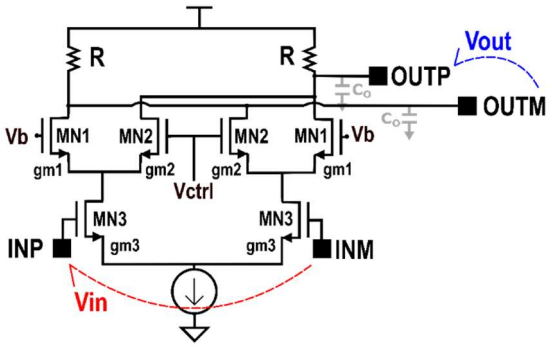


Fig. 2.2.1 Schematic of the first type of Gilbert Cell VGA

The transfer function of this VGA is expressed in (2.2.1), where we can observe that the low frequency gain depend on the ratio between gm_1 and gm_2. About the bandwidth, it is determined to the pole identified by the load resistance and the output capacitance C_O. At high frequency also the pole at the source of the cascode can introduce undesired losses. Moreover the output resistance of the cascode is affected by Miller effect, but for different gain setting its value changes and move the secondary pole of the cascode in different frequencies. This effect can change the shape of the transfer function for different gain settings.

$$\frac{V_{out}}{V_{in}} = -gm_3 R \cdot \frac{1-\frac{gm_2}{gm_1}}{1+\frac{gm_2}{gm_1}} \cdot \frac{1}{1+sRC_O} \tag{2.2.1}$$

Another very cascode based VGA is the Current Steering, reported in Fig. 2.2.2. The principle is the same of the VGA in Fig. 2.2.1, but we reject the "crossed" current instead to subtract it to the straight current. In the previous schematic the gain could change the sign, while in this other topology it is not possible, but at the same time we have less capacitance that load the output.

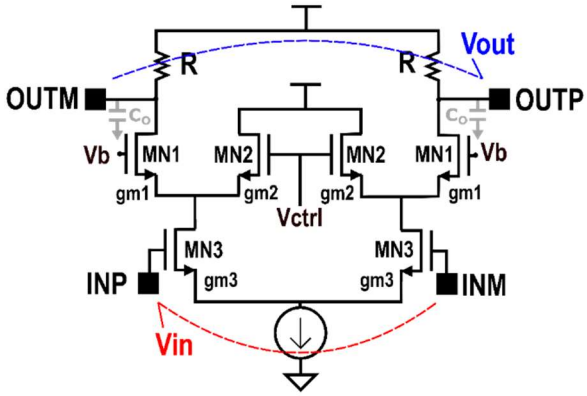


Fig. 2.2.2 Schematic of the Current Steering VGA

The transfer function of this topology is equal to (2.2.2) where we neglect in this case the pole et the source of the cascode. Also in this case this pole can becomes not negligible at high frequency, but its output resistance is no more affected by Miller effect. On the other hand, the variation of the gain, produce a variation of the bias current through the load resistance and consequently a variation of the output DC voltage.

$$\frac{V_{out}}{V_{in}} = -gm_3 R \cdot \frac{1}{1+\frac{gm_2}{gm_1}} \cdot \frac{1}{1+sRC_O} \tag{2.2.2}$$

In literature we can find many example of this type of gain variation, for example [2.14] [2.11] [2.16] used the Gilbert Cell as described above, while [2.15] decided to avoid the cascode, but control the current provided to the Gilbert Cell to change the ratio between the two transconductances.

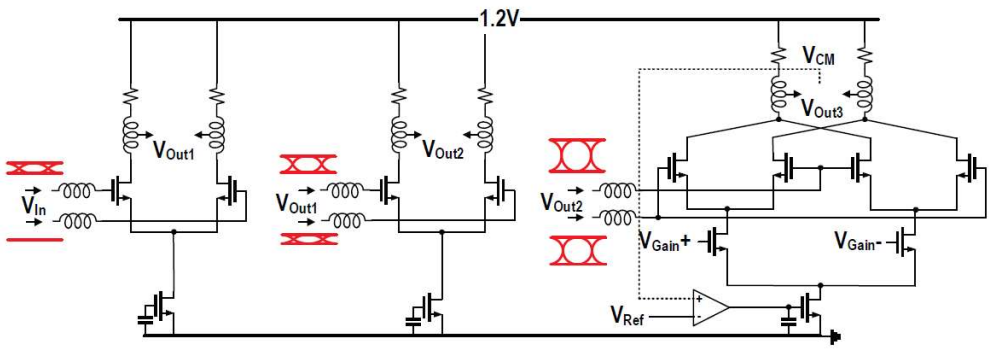
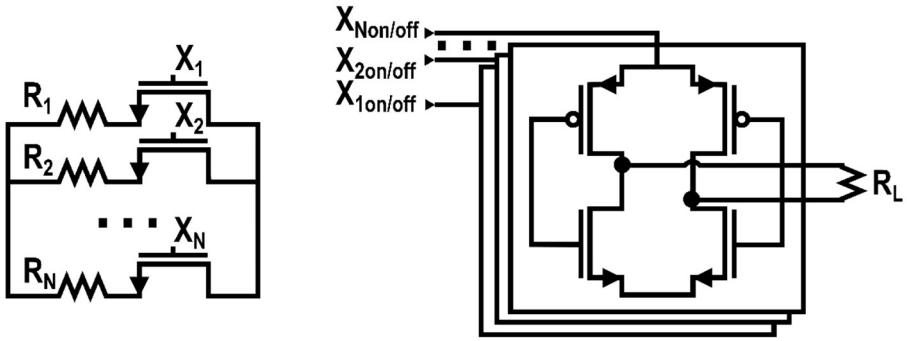


Fig. 2.2.3 Variable gain Amplifier proposed by [2.15]

2.2.2 Variable Resistance/Transconductance

The second technique used to vary the gain is based on variability of the load resistance or of the transconductance of the transistor. This is the most conventional technique and can be realized realizing a discrete or continuous gain variation. A discrete variable resistance is realized with series resistances that can be turned on or off digitally with switches, while a discrete variable transconductance can be created with N parallel MOSFET that can be turned on or off. Fig. 2.2.4 shows some example of these techniques.



(a) (b)

Fig. 2.2.4 Examples of discrete variable (a) resistance and (b) transconductance

Usually in coherent modulation the granularity of the gain variation is so small that a continuous gain variation is necessary and often it is realized with a transistor in triode region that works as a resistance, it can be used as a load or to create a variable resistive degeneration of the transistor. Often in literature this technique is used to vary the gain of the LNTA if it is realized with a shunt feedback. For example, [2.14] realized the feedback resistance with a discrete passive resistance in parallel to a MOSFET in triode

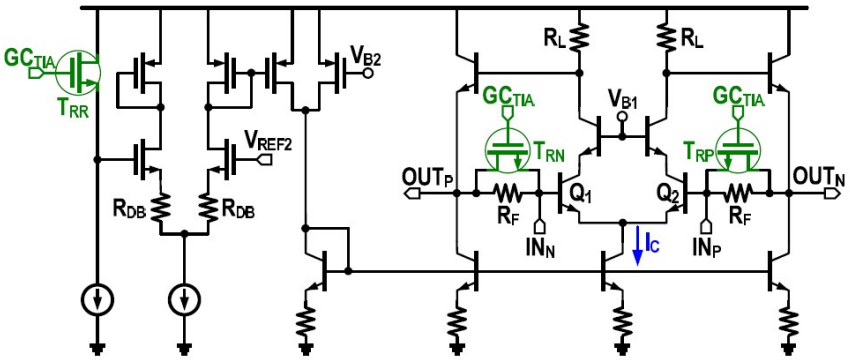


Fig. 2.2.5 Variable Gain Transimpedance Amplifier schematic [2.14]

Another example of VGA based on variable resistance is proposed by Stefano Facchin et. al. in [2.18]. They realized the Variable Gain Amplifier with the Cherry Hopper topology, that will be described in the section 2.3 as it is used for high bandwidth performances. As we can observe in Fig. 2.2.6 this topology is characterized by a transistor that pushes the signal current in a low impedance created by a resistance in feedback. The low frequency gain is mainly determined by this resistance that can be tuned with a CMOS in triode region.

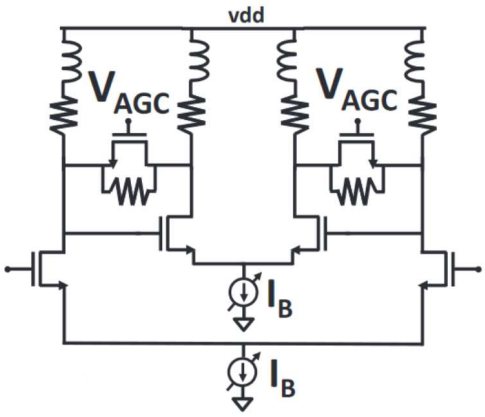


Fig. 2.2.6 Cherry-Hooper and differential variable gain amplifier schematic [2.18]

2.3 Enhancement Bandwidth Techniques

The issue relative to bandwidth limitations becomes more and more important as we grow up un frequency, usually to compensate the losses a combination of equalization and bandwidth extension techniques is fundamental to reach the desired data rate. In this section the most common techniques will be presented.

2.3.1 Inductive Peaking

The simplest techniques used to enhance the bandwidth is the inductive peaking. It involves the use of inductances in order to produce resonances and increase the impedance at high frequency. Limit of this techniques are the not negligible parasitics of the integrated inductances, in particular the quality factor, that usually is not so fundamental for these type of application, and the parasitic capacitance that limits the maximum frequency within wich the impedance behaves like an inductance. The fundamental techniques that involve the inductance peaking are:

- Shunt Peaking
- Bridged Shunt Peaking
- Series Peaking
- Triple Resonance Network (or Shunt Series Peaking)
- Bridged Shunt Series Peaking
- Reverse Triple Resonance Network
- T-coil Peaking
- Active Inductive Peaking

The last technique in the list in reality doesn't involve inductances, but an active circuit that behaves as an inductance.

2.3.1.1 Shunt Peaking

The schematic of a generic shunt peaking is presented in Fig. 2.3.1 where we can identify that the resistive load is replaced by a resistance in series to an inductance. Increasing the impedance load we increase the gain and we push more signal current to the output improving the circuit's speed [2.19].

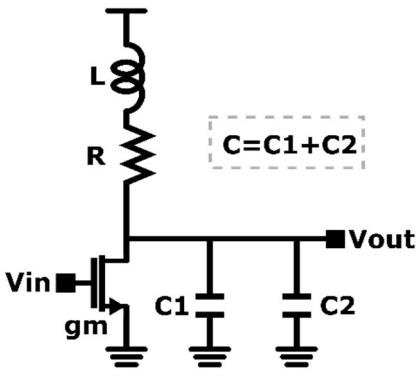


Fig. 2.3.1 Schematic of the shunt inductance peaking

To understand the bandwidth advantages of this technique we have to evaluate the transfer function determined by the added network [2.1].

(2.3.1) $$\frac{V_{out}}{V_{in}} = -g_m R \cdot \frac{1+\frac{sL}{R}}{1+sCR+s^2CL} = -\frac{g_m R\omega_0}{2\zeta} \cdot \frac{s+2\zeta\omega_0}{s^2+2\zeta\omega_0 s+\omega_0^2}$$

The transfer function is characterized by a zero that is determined by the resistance and the inductance, and two complex poles. If we impose a flat response that corresponds to $\zeta = {\sqrt{2}}/{2}$, we can compute the -3 dB cut off frequency that results equal to (2.3.2) [2.1].

(2.3.2) $$\omega_{-3dB} = \left[\frac{1}{4\zeta^2}+1-2\zeta^2+\sqrt{\left(\frac{1}{4\zeta^2}+1-2\zeta^2\right)+1}\right] \cdot \frac{4\zeta^2}{(RC)^2}$$

The complexity of the formula is determined by the zero, in fact if the neglect it in first approximation, the resulting bandwidth is (2.3.3) [2.1].

(2.3.3) $$\omega_{-3dB} \approx \frac{\sqrt{2}}{RC}$$

Without any inductance the bandwidth corresponds to the single pole $\omega_{-3dB} \approx \frac{1}{RC}$, it means that the shunt inductive peaking can provide at least a bandwidth improvement equal to $\sqrt{2}$. [2.1] studied the effect of the shunt peaking and the resulting overshooting considering or not the zero.

Overshoot	5%	7.5%	10%
ζ (with zero)	0.73	0.69	0.65
ζ (without zero)	0.69	0.64	0.59
BW enhancement (with zero)	78%	82%	84%

Table 2.3.1 The BW and the dumping factor with and without zero in correspondence of overshooting's percentage

In function of the desired overshooting we can have different bandwidth enhancements, that can reach also 84%. The overshooting can be desired and used to perform equalization of the other stages [2.1].
In literature we can fine many examples where the series indicative peaking is used, any person tries to find the best way to take the advantage to split capacitances with inductances for the considered design. An example is [2.13], how it was already introduced in the previous section. Other two examples are presented by [2.16] and [2.15]. In particular, very interesting, in the work of Balamurugan et al., is the comparison with and without the inductive peaking reported also in Fig. 2.3.2 where it is evident how the bandwidth is enhanced passing from 32.3 GHz to 72.3 GHz.

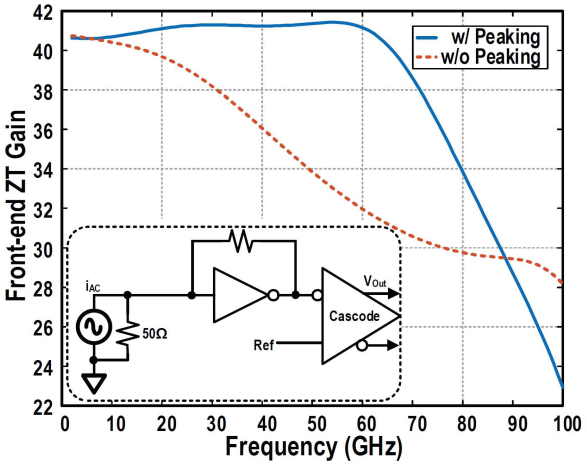


Fig. 2.3.2 Transfer function with and without the inductive peaking [2.15]

2.3.1.2 Bridged Shunt Peaking

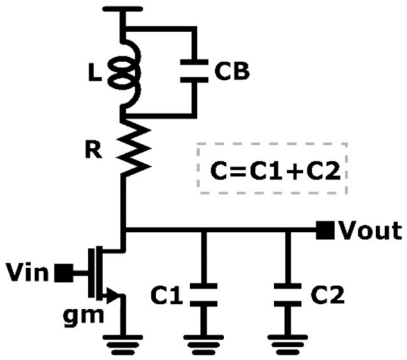


Fig. 2.3.3 Schematic of the bridged shunt inductance peaking

In this techniques the shunt inductor is in parallel to a capacitor. In reality any inductance has its own parasitic capacitor, it means that C_B can be not necessarily an explicit capacitance, but simply the parasitic capacitor. The bridged shunt peaking is a variant of the shunt peaking; it derives from the necessity to control the peaking of the shunt technique. If we look to the shunt peaking transfer function, it can be rewritten as (2.3.4) [2.19].

$$\frac{V_{out}}{V_{in}} = -g_m R \cdot \frac{1 + \frac{s}{m\omega_x}}{1 + \frac{s}{\omega_x} + \frac{s^2}{m\omega_x}} \tag{2.3.4}$$

Where $\omega_x = 1/RC$ and $m = (R^2\,C)/L$. The maximum bandwidth extension corresponds to a factor of 1.84 at $m = \sqrt{2}$, but it produces also a peaking of 1.5 dB, while for a flat response the maximum bandwidth extension is of 1.72. The shunt capacitor should be small enough to not degrade the bandwidth

extension, but also high enough to reduce the peaking [2.19]. The resulting transfer function is (2.3.5) [2.19].

$$\frac{V_{out}}{V_{in}} = -g_m R \cdot \frac{1 + \frac{sL}{R} + s^2 LC_B}{1 + sCR + s^2 L(C_B + C) + S^3 LC_B RC} = -g_m R \cdot \frac{1 + \frac{\left(\frac{1}{m}\right)s}{\omega_x} + \frac{\left(\frac{k_B}{m}\right)s^2}{\omega_x^2}}{1 + \frac{s}{\omega_x} + \frac{\left(\frac{k_B + 1}{m}\right)s^2}{\omega_x^2} + \frac{\left(\frac{k_B}{m}\right)s^3}{\omega_x^3}} \quad (2.3.5)$$

Where $k_B = \frac{C_B}{C}$. [2.19] explains that for a $k_B = 0.3$ it is possible enhance the bandwidth of 1.83 with a flat response. Fig. 2.3.4 shows the normalized transfer function for the uncompensated response and for the bridged shunt peaking technique by varying k_B [2.19].

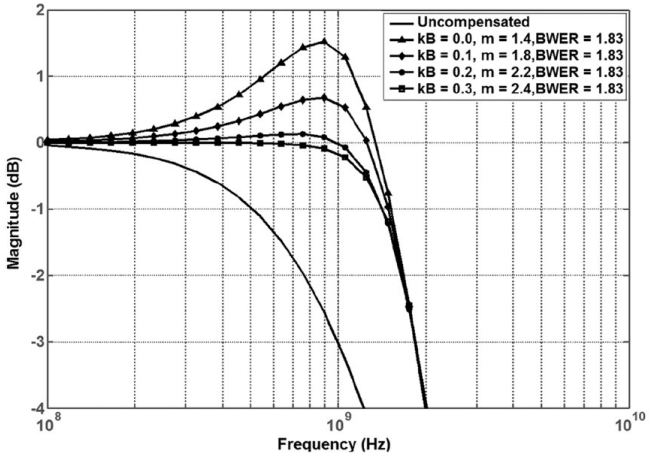


Fig. 2.3.4 Ideal transfer function where bridged shunt peaking is implemented versus k_B [2.19]

2.3.1.3 Series Peaking

The series transfer function schematic is represented in Fig. 2.3.5. The advantage of the capacitance splitting is that without the inductance the current flows from the transistor to the sum of C1 and C2, but with the inductor the currents through L is delayed and the MOSFET has to fill only C1 at the beginning [2.19].

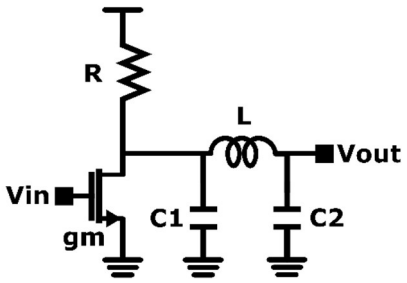


Fig. 2.3.5 Schematic of series peaking

We can observe that exits two different type of series peaking, the type represented in the previous image is the most common to find in literature. Its transfer function is equal to (2.3.6) [2.19].

$$\frac{V_{out}}{V_{in}} = -R \cdot \frac{1}{1 + sR(C_1 + C_2) + s^2LC_2 + s^3LC_1C_2R} = -R \cdot \frac{1}{1 + \frac{s}{\omega_x} + \frac{\left(\frac{1-k_c}{m}\right)s^2}{\omega_x^2} + \frac{\left(\frac{k_c(1-k_c)}{m}\right)s^3}{\omega_x^3}} \tag{2.3.6}$$

Where $k_c = \frac{C_1}{C_1+C_2}$, $\omega_x = \frac{1}{R(C_1+C_2)}$ and $m = \frac{R^2(C_1+C_2)}{L}$. In (2.3.6) we can identify two additional poles respect the RC starting solution, [2.19] demonstrates that the bandwidth can be improved by a factor of 2.52 for a $k_c = 0.3$ without peaking, but the bandwidth can be increased more if the resulting peaking is acceptable, or for example used for equalization purpose [2.7].

[2.7] proposed also a variation of the series peaking in case that $k_c < 0.1$. The other series peaking technique is based on the schematic in Fig. 2.3.6.

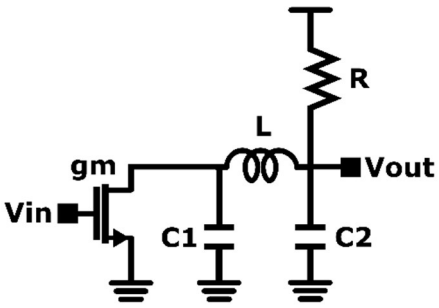


Fig. 2.3.6 Schematic of the second variant of series peaking

The transfer function becomes equal to

$$\frac{V_{out}}{V_{in}} = -R \cdot \frac{1}{1 + sR(C_1 + C_2) + s^2LC_1 + s^3LC_1C_2R} \tag{2.3.7}$$

It is almost equal to (2.3.6) by the except of the second order term at the numerator that instead of C_2, R multiplies C_1, it means that if we approximate $C_1 + C_2 \approx C_2$, (2.3.6) becomes

$$\frac{V_{out}}{V_{in}} \approx -R \cdot \frac{1}{(1 + s^2LC_1)(1 + sRC_2)} \tag{2.3.8}$$

As we can imagine the two complex poles at LC_1 presents a quality factor close to infinite, it means that the peaking produced can be very high, it can be useful if this peaking is desirable, but if it is excessive the only way is to decrease the quality factor of the inductance [2.19].

Fig. 2.3.7 shows exactly this concept representing the normalized transfer function for a give load resistance when $C_1 \ll C_2$ [2.19].

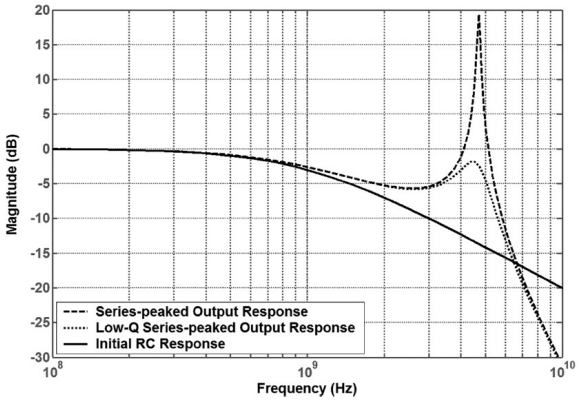
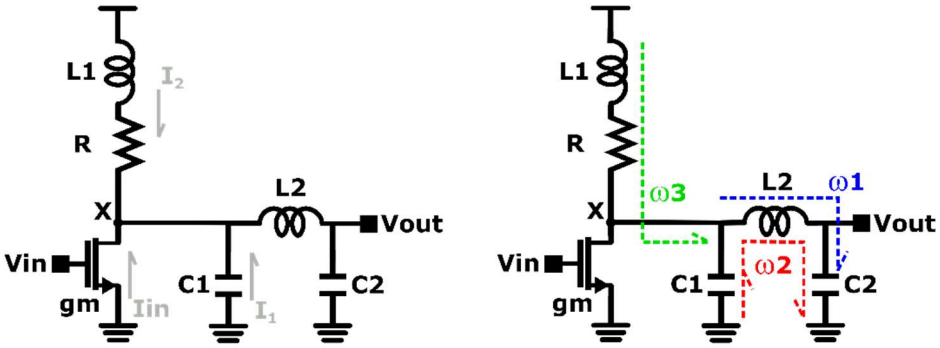


Fig. 2.3.7 Simulated second type of series peaking including an ideal inductance [2.19]

2.3.1.4 Triple Resonance Network (TRN)

The schematic that represents the TRN is represented in Fig. 2.3.8. The aim of this technique is to use resonances in order to push much current as possible to the load in order to increase the bandwidth. In particular, we can identify three resonances, to understand them we start describing the behavior of the network in frequency.



(a) (b)

Fig. 2.3.8 (a) Schematic of the Triple Resonance Network with the signed currents and voltages, (b) and the highlighted resonances

At very low frequency the current provided by the MOSFET is all collected by the resistance as long as the inductance L_1 resonates with the parallel between C_1 and C_2, where a greater portion of I_{in} is pushed in C_2. This behavior is enhanced by the inductance L_2 that is sized in order to resonate with C_2 at the same frequency (ω_1). In that way the signal current is almost all collected by the output capacitance due to the low impedance produced by the series resonance. If, for simplicity we assume that $C_1 = C_2 = \frac{C}{2}$, to set equal the two frequencies $\frac{1}{L_1 C}$ and $\frac{1}{L_2 C_2}$ the relation $L_2 = 2L_1$ must be satisfied. Usually the gain at this frequency is set equal to the low frequency gain producing the relation (2.3.9) [2.20].

(2.3.9) $$R = 2\sqrt{\frac{L_1}{C}}$$

After ω_1 the voltage at node X changes the sign producing an incoming current to C_2 from C_1 (I_1) and L_1, R (I_2). I_2 produces a voltage in V_{out} almost constant in frequency and proportional to theratio of the two capacitances, while I_1 produces a roll-up in V_{out} for the inductive component. The transfer function continues to rise until the resonance of the π network realized by L_2, C_1, C_2. At that frequency due to the high impedance produced by the parallel resonance, the signal current from the MOSFET is most absorbed by the series of the inductance and resistance. Can be demonstrated that the frequency of the second resonance is equal to $\omega_2 = \sqrt{2}\omega_1$. Since the two capacitances are equal and at the resonance carries the same current in modulus, we can assume that the output voltage is equal to the voltage at node X producing a gain of (2.3.10) [2.20].

(2.3.10) $$|V_{out}| = |I_{in}|\sqrt{\frac{3}{2}}R$$

After w2 the π-network becomes capacitive and V_{out} comes back to the midband value of $I_{in}R$ again only when L_1 resonates with the π-network that corresponds to a frequency equal to $\omega_3 = \sqrt[4]{6}\omega_1$. The final approximated bandwidth results equal to (2.3.11) [2.20].

(2.3.11) $$\omega_{-3dB} \approx \frac{2\sqrt{3}}{RC} = \sqrt{3}\omega_1$$

The final transfer function results something like Fig. 2.3.9.

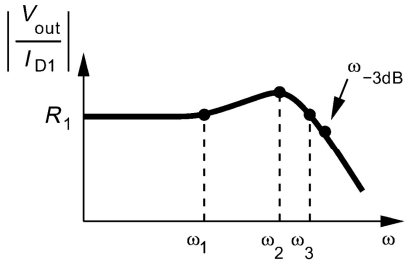


Fig. 2.3.9 Generic transfer function of a TRN [2.20]

2.3.1.5 Bridged Shunt Series Peaking

Starting from the shunt and the series peaking, different passive network can be realized, the first one that we present here is the bridged shunt series peaking. It is a composition of the previous described techniques, in particular its schematic is shown in Fig. 2.3.10.

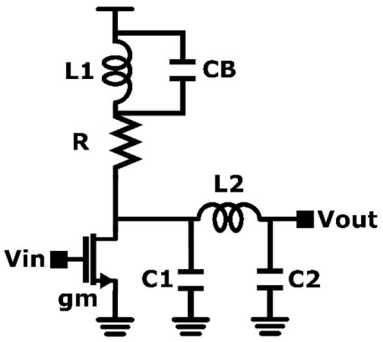


Fig. 2.3.10 Schematic of the bridged shunt series peaking

The expression of the transfer function becomes more complicated; we write it here below [2.19].

$$\frac{V_{out}}{V_{in}} \approx -R \cdot \frac{1+\frac{\left(\frac{1}{m_1}\right)s}{\omega_x}+\frac{\left(\frac{k_B}{m_1}\right)s^2}{\omega_x^2}}{1+\frac{s}{\omega_x}+\frac{\left(\frac{1+k_B}{m_1}+\frac{1-k_C}{m_2}\right)s^2}{\omega_x^2}+\frac{\left(\frac{k_B}{m_1}+\frac{k_C(1-k_C)}{m_2}\right)s^3}{\omega_x^3}+\frac{\left(\frac{(k_C+k_B)(1-k_C)}{m_1 m_2}\right)s^4}{\omega_x^4}+\frac{\left(\frac{k_B k_C(1-k_C)}{m_1 m_2}\right)s^5}{\omega_0^5}} \tag{2.3.12}$$

Where $k_c = \frac{C_1}{C_1+C_2}$, $\omega_x = \frac{1}{R(C_1+C_2)}$, $k_B = \frac{C_B}{C_1+C_2}$, $m_1 = \frac{R^2(C_1+C_2)}{L_1}$ and $m_2 = \frac{R^2(C_1+C_2)}{L_2}$. This technique can be considered ad an evolution of the Triple Resonance Network, or also called shunt series peaking. The advantage is the additional degree of freedom of C_B that can helps in reaching wider bandwidth with a limited peaking [2.19].

2.3.1.6 Reverse Triple Resonance Network (RTRN)

The Reverse Triple Resonance Network is a variant of the TRN but the resistive load is after the series inductance (Fig. 2.3.11).

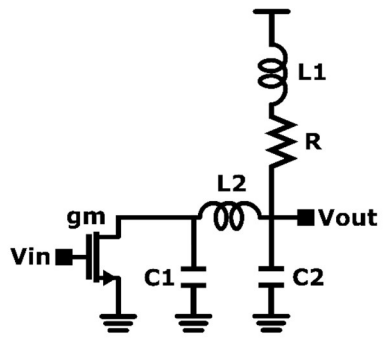


Fig. 2.3.11 Schematic of the Reverse Triple Resonance Network

[2.21] suggests a methodology to study this network in a simplified way. First of all the authors decided to express the two capacitances in function of the parameter α, in particular $C_1 = \frac{1+\alpha}{2}$ and $C_2 = \frac{1-\alpha}{2}$ where usually α is greater than 1 due to Miler effect of the next stage. Then they rewrote the RTRN to an equivalent circuit equal to the parallel of an equivalent capacitance and a series of an equivalent resistance and inductance (Fig. 2.3.12).

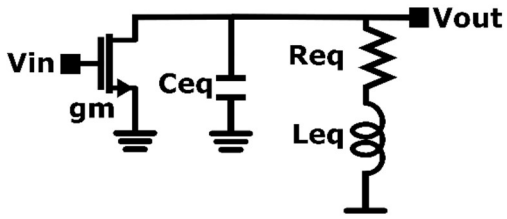


Fig. 2.3.12 Schematic of the equivalent circuit of the RTRN

The total output impedance becomes equal to (2.3.13) [2.21].

(2.3.13) $$Z_{tot} = (R_{eq} + j\omega L_{eq}) \| \frac{1}{j\omega C_{eq}} = \left[\left(\frac{R}{1 - \omega^2 L_2 \frac{C}{2}(1-\alpha)} \right) + j\omega \left(\frac{L_1}{1 - \omega^2 L_2 \frac{C}{2}(1-\alpha)} \right) \right] \| [\frac{1}{j\omega C} \cdot \frac{1}{1 - \omega^2 L_2 \cdot \frac{C}{4}(1-\alpha^2)}]$$

If we observe the previous equation we can identify three resonances: when the series L_{eq} and R_{eq} resonates ($\omega_1 = 1/\sqrt{\frac{L_2 C(1-\alpha)}{2}}$), when C_{eq} resonates ($\omega_2 = \sqrt{\frac{L_2 C(1-\alpha^2)}{4}}$) and when Z_{tot} becomes purely real. Usually the frequency of the first resonance is design in order to be equal to the resonance frequency of a shunt peaking, it means that ω_1 is equal to $\frac{1}{\sqrt{L_1 C}}$. Under that assumption and introducing the constrain that the gain at this resonance must be equal to the low frequency gain, we can get the expression of the inductances in function of C, R and α [2.21].

(2.3.14) $$L_1 = \frac{(1-\alpha)^2 C R^2}{4}$$

(2.3.15) $$L_2 = \frac{(1-\alpha) C R^2}{2}$$

We can use (2.3.14) and (2.3.15) to determine the gain and the frequency of the other resonances that results equal to the following equations [2.21].

(2.3.16) $$\omega_2 = \sqrt{\frac{2}{1+\alpha}}\,\omega_1$$

(2.3.17) $$|Z_{tot}|_{\omega_2} = \frac{1}{1 - \left(\frac{\omega_2}{\omega_1}\right)^2} \sqrt{R^2 + \omega_2^2 L_1^2}$$

(2.3.18) $$\omega_3 = \frac{f(\alpha)}{(1-\alpha)^2} \cdot \frac{2\sqrt[4]{6}}{CR}; \qquad f(\alpha) = \sqrt[4]{1-\frac{\alpha}{3}} \cdot \sqrt{\frac{-\alpha\sqrt{6-2\alpha}+\sqrt{1+4\alpha^2-2\alpha^3+\alpha^4}}{1+\alpha}}$$

(2.3.19) $$|Z_{tot}|_{\omega_3} = \left| \frac{1+\frac{\left(\frac{\omega_3}{\omega_1}\right)^2 (1-\alpha)^2}{4}}{1-\left(\frac{\omega_3}{\omega_1}\right)^2} \right| R$$

In function of α the transfer function can be affected by a significant roll off at low frequency that can be adjusted by tuning the value of L_1 [2.21].

2.3.1.7 T-coil Peaking

The purpose of the T-coil is to use a special network in order to see a constant impedance in frequency in a specific node. Referring to Fig. 2.3.13, ideally at low frequency the network allows to keep Z_{in} equal to the load resistance thanks to the inductances, while at high frequency the capacitance C_B bypass the load capacitor and also in this case Z_{in} ideally results equal to R.

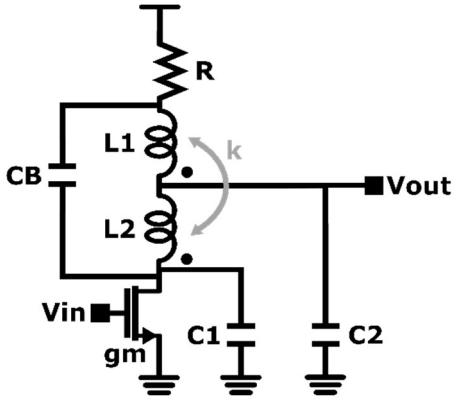


Fig. 2.3.13 Symmetric T-coil schematic

[2.1] explains that in order to have a bandwidth enhancement with a dumping factor of ζ the following relations must be used [2.1].

(2.3.20) $$L_1 = L_2 = \frac{C_2 R^2}{4}\left(1+\frac{1}{4\zeta^2}\right)$$

(2.3.21) $$C_B = \frac{C_2}{16\zeta^2}$$

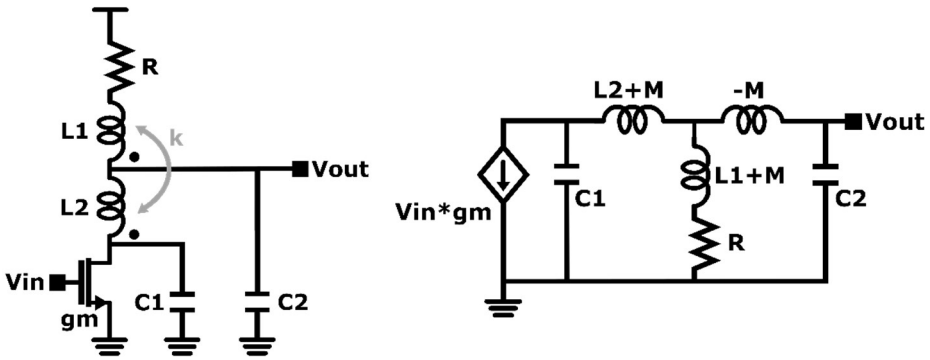
(2.3.22) $$k = \frac{4\zeta^2-1}{4\zeta^2+1}$$

Under these assumptions the transfer function becomes equal to

(2.3.23)
$$\frac{V_{out}}{Vin} = R \cdot \frac{1}{\frac{1}{4}\left(\frac{1-k}{1+k}\right)R^2C_2^2s^2 + \frac{1}{2}RC_2s + 1}$$

If k is 0.5 the bandwidth can be enhanced by a factor of 2.72 [2.1].

We can find in literature also the use of the T-coil in an asymmetric way [2.19]. When the load capacitance of the next stage becomes dominant to the intrinsic output capacitance of the stage ($C_2 > 0.7(C_1 + C_2)$) usually the bridged shunt series peaking doesn't provide enough bandwidth enhancement because the splitting action of L_2 becomes less effective. The mutual inductance is designed to be negative producing an equivalent circuit with the transformer model equal to Fig. 2.3.14 [2.19].



(a) (b)

Fig. 2.3.14 Schematic of the asymmetric T-coil (a) and equivalent schematic of the T-coil with the T model of the transformer, where $M=k\sqrt{L_1L_2}$ and k is the coupling factor of the transformer (b)

At the beginning due to the inductance L_2 the current flows only in C_1. When the currents start to go through L_2, a option of it goes also to C_2, but due to the equivalent negative inductance $-M$ connected in series, there is a boost in the current flow to the load capacitance improving the raise time. [2.19] explains that the symmetric T-coil uses a zero pair to cancel a pole couple, reaching a enhancement of the bandwidth of a factor equal to 2.83, while the asymmetric T-coil if realized with negative mutual inductance can achieve an enhancement of the bandwidth of the order of 5 for a peaking less than 2dB [2.19].

2.3.1.8 Active Inductive Peaking

The advantages in term of bandwidth extension by the use of inductances are clear, but the use of a passive inductance take a lot of area, so sometimes can be useful to have an inductive behavior without the use of a real inductance. An active inductance is a circuit that provides a low impedance at low frequency and a high impedance at high frequency. It can be realized in different ways, the most famous load that behave just like that are represented in Fig. 2.3.15.

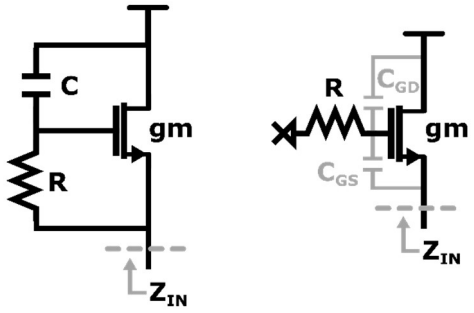


Fig. 2.3.15 Two possible implementations of an active inductance

In both the presented cases the purpose is to show an impedance equal to $\frac{1}{gm}$ at low frequency, and R at high frequency, in the first case thanks to the short of the explicit capacitance C, in the second case by the short of the parasitic capacitance C_{GS}. Of course if we take into account the parasitic elements of the transistor, we have to consider also the output resistance ro and all the others parasitic capacitances [2.1]. If we take for example the second implementation, also called Hara active inductive peaking [2.30], the impedance of the load results equal to

$$Z_{IN} = \frac{1 + sR(C_{GS} + C_{GD})}{gm(1 + sRC_{GD})\left(1 + \frac{sC_{GS}}{gm}\right)} \tag{2.3.24}$$

As we can see the equivalent impedance corresponds to a resistance of $\frac{1}{gm}$ up to the zero at $\omega_z = \frac{1}{R(C_{GD}+C_{GS})}$. After this frequency the impedance behave like an inductance up to the first pole at $\omega_{p1} = \frac{1}{RC_{GD}}$ and after $\omega_{p2} = \frac{gm}{C_{GS}}$ Z_{IN} is equal to a capacitance of value $\frac{C_{GD}C_{GS}}{C_{GD}+C_{GS}}$. It means that the bandwidth of interest of the active inductive peaking is from ω_z to ω_{p1}.

An interesting example of active inductive peaking is [2.30] where they implemented the Hara topology as Fig. 2.3.16 represents.

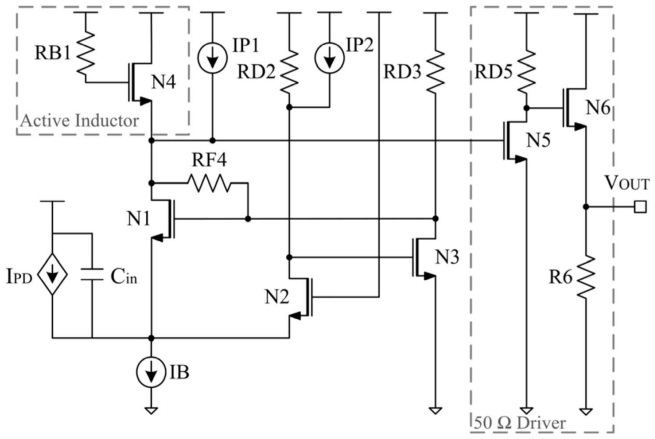


Fig. 2.3.16 Example of implementation of the active inductive peaking [2.22]

[2.22] discovered a bandwidth enhancement of 15% under the condition that $Rgm \gg 1$.

2.3.2 Negative Capacitance

The negative capacitance is a very interesting technique that uses the Miller effect to create a negative capacitance to subtract to the fixed capacitances of a node. In reality it doesn't create only a negative capacitance but it adds also positive capacitances due to parasitics, so this technique is more effective when the added positive capacitances are less than the negative ones. The first way to realize a negative capacitance is also called neutralization, it is a technique that is easy suitable for fully differential structure. It consists into insert a capacitance between the input and the output of a stage with the same sign (Fig. 2.3.17). In this way, if the stage has a gain of A at the input of the stage a capacitance equal to $C(1-A)$ will be added, while at the output there will be a capacitance equal to $C(1-\frac{1}{A})$. This solution can be useful if the dominant capacitance is at the input of the stage [2.1].

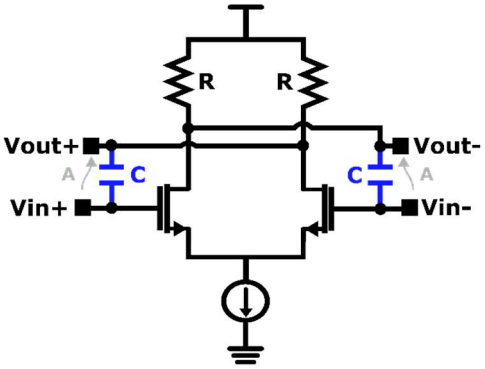


Fig. 2.3.17 Example of Neutralization

The principle presented before can be applied not only to an existing stage, but also adding to the node where we want a negative capacitance a stage with a gain A and the neutralization. In that we are adding a negative and a positive capacitance at that node, if the negative contribution is higher than the positive we can have a bandwidth enhancement.

The negative capacitance can also be created by a positive feedback structure presented in Fig. 2.3.18.

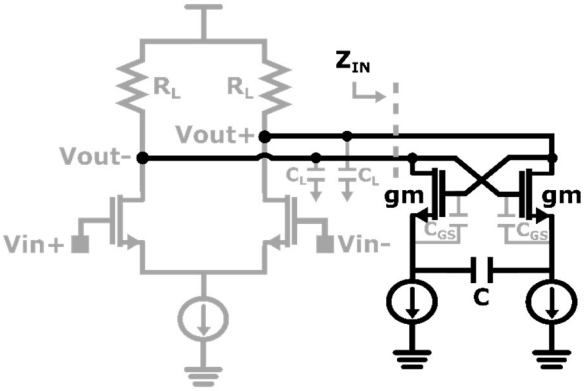


Fig. 2.3.18 Negative capacitance implementation

If the working frequency range is well below the f_T, the impedance Z_L results equal to

(2.3.25) $$Z_{IN} = -\frac{1}{sC} - (\frac{C_{GS}}{C} + 2) \cdot \frac{1}{gm}$$

As we can see in this way we are introducing a series between a negative capacitance equal to C and a negative resistance proportional to $\frac{1}{gm}$, to avoid oscillation the additional negative resistance has to be not dominant respect the resistive load at the node where we are introducing the negative capacitance. In particular, for the example reported in Fig. 2.3.18 the stability condition results equal to (2.3.16) [2.1].

(2.3.26) $$gmR_L \geq \frac{C}{C - \frac{C_L}{2}}$$

In literature we can find the implementation of a negative capacitance in [2.23], whose schematic is shown in Fig. 2.3.19. The authors decided to use a negative capacitance instead of the typical inductor based enhancement bandwidth techniques to minimize the area.

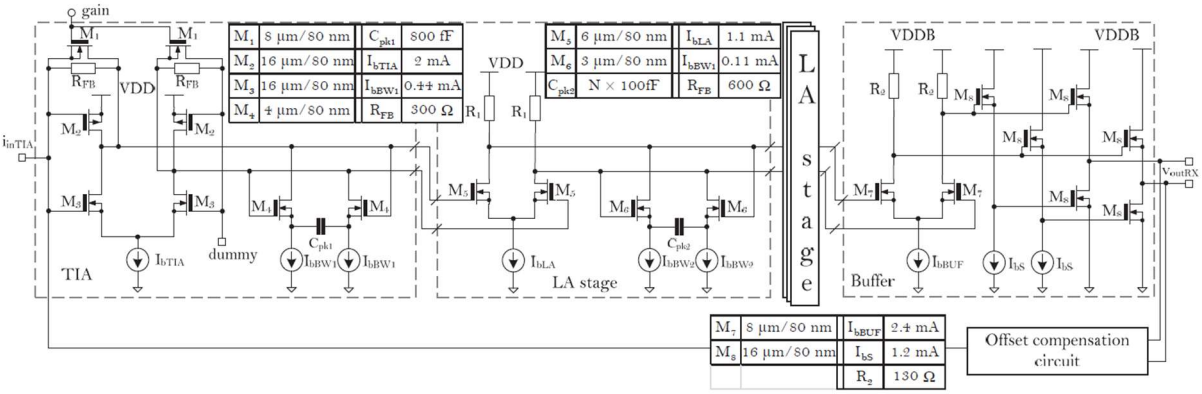


Fig. 2.3.19 Example of implementation of the negative capacitance [2.23]

2.3.3 Capacitive Degeneration

There are some techniques to enhance the bandwidth that doesn't involve a particular load, but that can be applied to particular topologies of the amplifiers. A first example is the capacitive degeneration that can be applied for example to a common source structure. The principle behind this technique is that a common source degenerated shows a transimpedance gain that is related to the impedance of degeneration, for example if the degeneration is a resistance, if the gm of the transistor is high enough, the equivalent transconductance is $\frac{1}{R_{deg}}$. The capacitive degeneration uses this principle to introduce a zero in the transfer function. An example of schematic that uses this technique is presented in Fig. 2.3.20 [2.1].

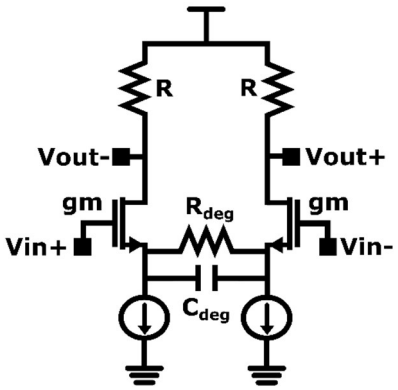


Fig. 2.3.20 Example of differential amplifier with capacitive degeneration (a) and its equivalent half circuit

The degenerated impedance is equal to

(2.3.27) $$Z_{deg} = \frac{1}{2C_{deg}} // \frac{R_{deg}}{2} = \frac{R_{deg}}{2\left(1 + R_{deg}C_{deg}\right)}$$

It means that the effective transconductance of the transistor becomes equal to

(2.3.28) $$gm_{eff} = \frac{gm}{1 + gmZ_{deg}} = \frac{gm}{1 + \frac{gmR_{deg}}{2}} \cdot \frac{1 + sC_{deg}R_{deg}}{1 + \frac{sC_{deg}R_{deg}}{1 + \frac{gmR_{deg}}{2}}} \approx \frac{1}{R_{deg}} \cdot \frac{1 + sC_{deg}R_{deg}}{1 + \frac{sC_{deg}}{gm}}$$

If the additional pole at $\omega_p = \frac{gm}{C_{deg}}$ is at enough high frequency, we can set the zero $\omega_z = \frac{1}{C_{deg}R_{deg}}$ in order to compensate the dominant pole of the transfer function and improve the bandwidth. We have to take into account that the choice of R_{deg} determines not only the zero position, but also the gain, for that reason this technique has the intrinsic tradeoff between the gain and the bandwidth [2.1]. Often this technique is used in the design in BiCMOS technology due to the higher gmOverId respect the CMOS technology. In fact due to the degeneration, as (2.3.18) shows, the effective transconductance is lower than the gm of the transistor. [2.6], [2.16] and [2.17] are some example of the implementation of this technique, that usually is used to add peaking at middle frequency [2.16], while for high frequency the gain is enhanced by other techniques like negative capacitances [2.17] or inductive peaking techniques [2.16].

2.3.4 Cherry Hooper Amplifier

The topology of the Cherry Hooper is very used in literature for its broadband performances. Its basic schematic is shown in Fig. 2.3.21.

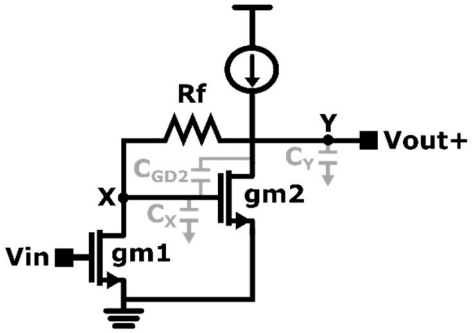


Fig. 2.3.21 First type of Cherry Hooper implementation

Its gain results equal to the product of gm_1 and the feedback resistance R_f, if $gm_2R_f \gg 1$, it means that it can provide the same gain of the traditional common source, but in very first approximation the impedance at the two node results equal to $\frac{1}{gm_2}$ producing poles at high frequency around $\omega_{X,Y} = \frac{gm_2}{C_{X,Y}}$ [2.1]. In reality the expression of $\omega_{X,Y}$ is optimistic, because as soon as C_Y short the output node, the loop start to have less gain and the input pole raise up, the same is valid respectively for the C_X and the output impedance [2.1]. If we take into account the also the capacitance C_{GD2}, we can derive the transfer function that results equal to

$$\frac{V_{out}}{V_{in}} = \frac{gm_1(gm_2R_f - 1)}{gm_2} \cdot \frac{1 + \frac{sC_{GD2}R_f}{R_f gm_2 - 1}}{1 + s\left(\frac{C_Y + gm_2R_fC_{GD2} + C_X}{gm_2}\right) + \frac{s^2R_f(C_XC_Y + C_{GD2}C_Y + C_{GD2}C_X)}{gm_2}} \tag{2.3.29}$$

If the two poles are equals and the term $gm_2R_fC_{GD2}$ is negligible, can be demonstrated that the two poles results equal to $\omega_{p1,2} = \frac{2gm_2}{C_X + C_Y}$ that is higher than the pole without the feedback $(R_fC_{X,Y})$ [2.1].
In literature we can find a variant of the Cherry Hooper presented in Fig. 2.3.22, where also a emitter follower is implemented.

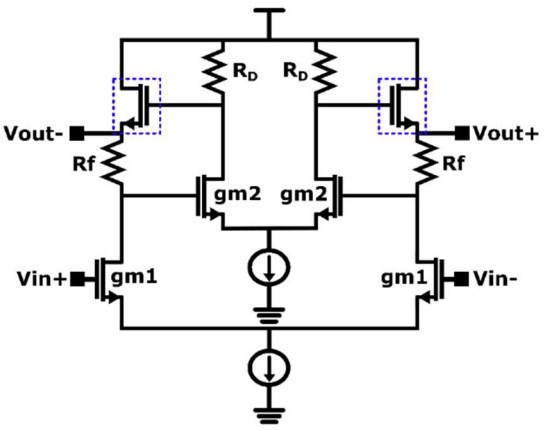


Fig. 2.3.22 Second type of Cherry Hooper implementation

In this way the follower is used to drive the load capacitance, where the follower impedance is also lowered by the feedback. The bigger disadvantage of this solution is the issue relative the headroom limitations [2.1].

2.3.5 f_T Doubler

The bandwidth is determined by the resistances and the capacitances at any node, but higher is the frequency and more the intrinsic capacitances of the MOSFET becomes not negligible, sometimes they can also become the dominant capacitances at the nodes. The f_T doublers has the purpose to find some way to double the intrinsic bandwidth of the MOSFET, or in other way, they want to half their intrinsic capacitances. This operation impacts on the power consumption and add parasitic capacitances on other nodes of the circuit, it means that this solution is good if the system presents a dominant pole on the transfer function that limit the overall bandwidth [2.1].

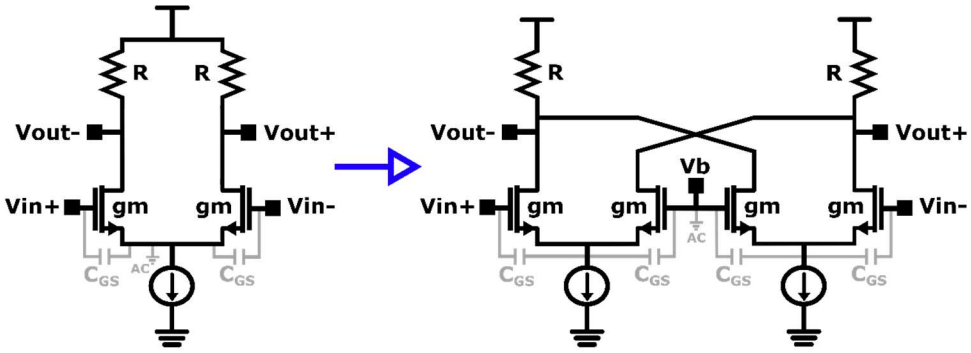


Fig. 2.3.23 Step from the common source to the f_T doubler

In Fig. 2.3.23 we can analyze the step performed to pass from a classic common source to an f_T doubler. In both the two figure the low frequency gain is determined by the product of the transconductance gm and R, but in the f_T doubler at the input the signal can see the series of the C_{GS} producing in first approximation half of the capacitance. Of course we are adding a exact equal structure at the output node, it means that the output capacitance will be double. Other drawbacks are the double current through the load resistance that can need a higher supply voltage, that with the double of the current produce an increment of the power consumption. If we consider also the capacitances referred to ground at the source of the common source, in reality the capacitance is decreased at the input less than a factor of 2 [2.1].

2.3.6 Active Feedback

The Active Feedback is a technique based on the idea that two complex poles at the same frequency of two real poles, can provide a better bandwidth due to the peaking. To demonstrate it we can consider the

two normalized transfer functions (2.3.18) and (2.3.19) that represent respectively the frequency response of two real poles and of two complex poles [2.1].

(2.3.30) $$TF_{2Real} = \frac{1}{\left(1+\frac{s}{\omega_1}\right)\left(1+\frac{s}{\omega_2}\right)}$$

(2.3.31) $$TF_{2Cpx} = \frac{1}{1+\frac{s}{\omega_x Q}+\frac{s^2}{\omega_x^2}}$$

To compare the bandwidth, we make the hypothesis that the two real poles and the complex poles are at the same frequency $\omega_0 = 1/\tau_0$.

(2.3.32) $$\frac{1}{\left(1+\frac{\omega^2}{\omega_0^2}\right)\left(1+\frac{\omega^2}{\omega_0^2}\right)} = 1/2$$

(2.3.33) $$\frac{1}{\left(1-\frac{\omega^2}{\omega_0^2}\right)^2+\frac{\omega^2}{\omega_0^2 Q}} = 1/2$$

If we solve the equations (2.3.22) and (2.3.23) we obtain that the -3dB frequency are respectively equal to $BW_{2Real} = \omega_0\sqrt{\sqrt{2}-1}$ and $BW_{2Cpx} = \omega_0$ for $Q = \sqrt{2}$. It results into a bandwidth enhancement of a factor 1.55 for the complex solution [2.1]. To use this technique a feedback has to be introduced in order to make complex two poles, in general we will have a structure similar to Fig. 2.3.24.

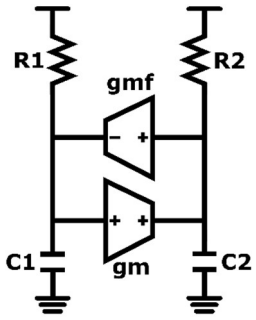


Fig. 2.3.24 Generic schematic of an Active Feedback circuit

2.3.7 Distributed Amplifier

The last technique that we want to proposed is the distributed structure. They are based on the use of transmission lines to split the amplifier in smaller amplifiers distributed along the input and output transmission lines. For properly functionality the lines must be terminated and in the design the delay of

the signal along the line must be studied in order to sum in the proper way the contribution from the different small amplifiers. Even if ideally it can provide a bandwidth enhancement of a factor of N, where N is the number of partitions realized on the amplifier, in reality it has also other issues, starting from the complexity of the design in particular if there are many stages, the parasitic losses of the transmission lines that degrade the bandwidth, possible mismatches in the output and input transmission lines, the complex realization of cascode structure and routing and of course the Miller effect between the input and output line [2.1].

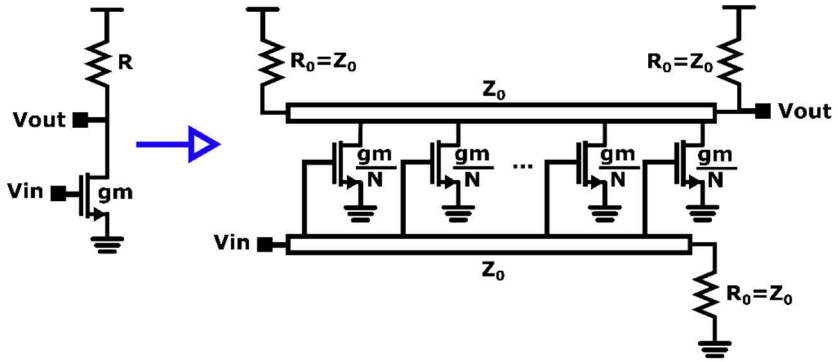


Fig. 2.3.25 Generic step from an amplifier to the schematic of a distributed amplifier

Chapter references

[2.1] Behzad Rezavi "Design of Integrated Circuits for Optical Communications" Second edition, john wiley & sons, inc. publication 2012.
[2.2] Eduard Sackinger "Broadband Circuits for Optical Fiber Communication" Copyright 0 2005 by John Wiley & Sons, Inc.
[2.3] E. Säckinger, "The Transimpedance Limit," in IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 57, no. 8, pp. 1848-1856, Aug. 2010, doi: 10.1109/TCSI.2009.2037847.
[2.4] G. Kalogerakis, T. Moran, T. Nguyen and G. Denoyer, "A quad 25Gb/s 270mW TIA in 0.13μm BiCMOS with <0.15dB crosstalk penalty," 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, 2013, pp. 116-117, doi: 10.1109/ISSCC.2013.6487661.
[2.5] Eduard Sackinger "The Transimpedance Limit" Ieee Transections on Circuits and Systems-i:Regular papers vol. 57, no. 8, august 2010
[2.6] S. B. Amid, C. Plett and P. Schvan, "Fully differential, 40 Gb/s regulated cascode transimpedance amplifier in 0.13 μm SiGe BiCMOS technology," 2010 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM), 2010, pp. 33-36, doi: 10.1109/BIPOL.2010.5667977.
[2.7] Sung Min Park and Hoi-Jun Yoo, "1.25-Gb/s regulated cascode CMOS transimpedance amplifier for Gigabit Ethernet applications," in IEEE Journal of Solid-State Circuits, vol. 39, no. 1, pp. 112-121, Jan. 2004, doi: 10.1109/JSSC.2003.820884.
[2.8] R. D. Middlebrook, V. Vorperian and J. Lindal, "The N Extra Element Theorem," in IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 45, no. 9, pp. 919-935, Sept. 1998, doi: 10.1109/81.721258.
[2.9] M. Tian, V. Visvanathan, J. Hantgan and K. Kundert, "Striving for small-signal stability," in IEEE Circuits and Devices Magazine, vol. 17, no. 1, pp. 31-41, Jan. 2001, doi: 10.1109/101.900125.
[2.10] S. Daneshgar, H. Li, T. Kim and G. Balamurugan, "A 128 Gb/s PAM4 Linear TIA with 12.6pA\√Hz Noise Density in 22nm FinFET CMOS," 2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), 2021, pp. 135-138, doi: 10.1109/RFIC51843.2021.9490496.
[2.11] G. Coudyzer, P. Ossieur, J. Bauwelinck and X. Yin, "A 25Gbaud PAM-4 Linear Burst-Mode Receiver With Analog Gain- and Offset Control in 0.25μm SiGe:C BiCMOS," in IEEE Journal of Solid-State Circuits, vol. 55, no. 8, pp. 2206-2218, Aug. 2020, doi: 10.1109/JSSC.2020.2987680.
[2.12] Ke Li, Shenghao Liu, Xiaoke Ruan, Dave J. Thomson, Yang Hong, Fan Yang, Lei Zhang, Cosimo Lacava, Fanfan Meng, Weiwei Zhang, Periklis Petropoulos, Fan Zhang, and Graham T. Reed," Co-design of a differential transimpedance amplifier and balanced photodetector for a sub-pJ/bit silicon photonics receiver," in Optic Express, vol. 28, no. 9/27, April 2020, doi: 10.1364/OE.389889
[2.13] K. Lakshmikumar, A. Kurylak, M. Nagaraju, R. Booth and J. Pampanin, "A process and temperature insensitive CMOS linear TIA for 100 Gbps/λ PAM-4 optical links," 2018 IEEE Custom Integrated Circuits Conference (CICC), 2018, pp. 1-4, doi: 10.1109/CICC.2018.8357007.

[2.14] M. G. Ahmed, T. N. Huynh, C. Williams, Y. Wang, P. K. Hanumolu and A. Rylyakov, "34-GBd Linear Transimpedance Amplifier for 200-Gb/s DP-16-QAM Optical Coherent Receivers," in IEEE Journal of Solid-State Circuits, vol. 54, no. 3, pp. 834-844, March 2019, doi: 10.1109/JSSC.2018.2882265.
[2.15] H. Li, G. Balamurugan, J. Jaussi and B. Casper, "A 112 Gb/s PAM4 Linear TIA with 0.96 pJ/bit Energy Efficiency in 28 nm CMOS," ESSCIRC 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC), 2018, pp. 238-241, doi: 10.1109/ESSCIRC.2018.8494285.
[2.16] I. García López, A. Awny, P. Rito, M. Ko, A. C. Ulusoy and D. Kissinger, "100 Gb/s Differential Linear TIAs With Less Than 10 pA\sqrt(Hz) in 130-nm SiGe:C BiCMOS," in IEEE Journal of Solid-State Circuits, vol. 53, no. 2, pp. 458-469, Feb. 2018, doi: 10.1109/JSSC.2017.2782080.
[2.17] A. Awny et al., "A dual 64Gbaud 10kΩ 5% THD linear differential transimpedance amplifier with automatic gain control in 0.13µm BiCMOS technology for optical fiber coherent receivers," 2016 IEEE International Solid-State Circuits Conference (ISSCC), 2016, pp. 406-407, doi: 10.1109/ISSCC.2016.7418079.
[2.18] S. Facchin, S. Zhou, M. Power, A. Jain, C. Scarcella, C. Antony, P. Townsend, P. Ossieur, "A 20Gbaud/s PAM-4 65nm CMOS optical receiver using 3D solenoid based bandwidth enhancement," 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS), 2017, pp. 723-726, doi: 10.1109/MWSCAS.2017.8053025.
[2.19] S. Shekhar, J. S. Walling and D. J. Allstot, "Bandwidth Extension Techniques for CMOS Amplifiers," in IEEE Journal of Solid-State Circuits, vol. 41, no. 11, pp. 2424-2439, Nov. 2006, doi: 10.1109/JSSC.2006.883336.
[2.20] S. Galal and B. Razavi, "40-Gb/s amplifier and ESD protection circuit in 0.18-/spl mu/m CMOS technology," in IEEE Journal of Solid-State Circuits, vol. 39, no. 12, pp. 2389-2396, Dec. 2004, doi: 10.1109/JSSC.2004.835639.
[2.21] C. Liao and S. Liu, "40 Gb/s Transimpedance-AGC Amplifier and CDR Circuit for Broadband Data Receivers in 90 nm CMOS," in IEEE Journal of Solid-State Circuits, vol. 43, no. 3, pp. 642-655, March 2008, doi: 10.1109/JSSC.2007.916626.
[2.22] L. Szilagyi, R. Henker and F. Ellinger, "An inductor-less ultra-compact transimpedance amplifier for 30 Gbps in 28 nm CMOS with high energy-efficiency," 2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS), 2014, pp. 957-960, doi: 10.1109/MWSCAS.2014.6908575.
[2.23] L. Szilagyi, R. Henker and F. Ellinger, "20 Gbit/s ultra-compact optical receiver front-end with variable gain transimpedance amplifier in 80 nm CMOS," 2016 IEEE MTT-S Latin America Microwave Conference (LAMC), 2016, pp. 1-3, doi: 10.1109/LAMC.2016.7851250.

CHAPTER 3

This chapter starts with the study of the used technology in order to determine the intrinsic performances of the MOSFET in function of its polarization. This analysis was realized to optimize the designs and know the tradeoff between the cut-off frequency, the noise and the transconductance gain per unit current of the used transistor. After the technology analysis, follow the presentation of the proposed oTIA designs.

3.1 Study of the technology

In order to optimize the designs, a previous analysis of the technology is fundamental because it defines the intrinsic limit of the transistors and it can be used to provide a good comparison with works realized in other technologies. In this section two variant of the same technology are presented, in particular we studied the 28nm CMOS HPC and the 28nm CMOS HPC+ developed by the Taiwan Semiconductor Manufacturing Company (TSMC). This technology can be used in a variety of applications such as high-speed networking chips, Central Processing Units (CPUs) and consumer electronics. It presents a reduction of the power consumption by lower leakage current, in particular 28HPC reduces the leakage by 20% respect the simple 28nm technology, while the 28NPC+ reduces it by 25% providing also 15% better performances. Moreover, its relaxed process rules can provide lower area and cost. The back end-of-the-line (BEOL) version that is chosen for this research work is characterized by 8 metal levels and the substrate is of P-type with no epitaxial layer, as common in bulk CMOS technologies. The metal stacks of the two technologies are equal and represented in Fig.3.1.1, where the layer thickness is not in scale [3.1].

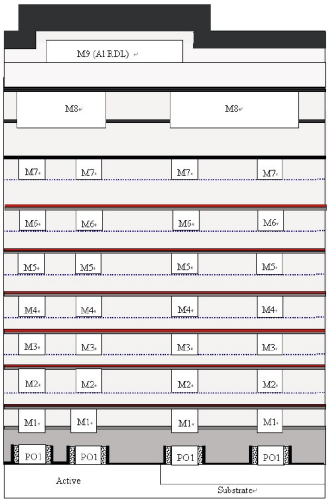


Fig.3.1.1 Metal stack not in scale of the 28nm CMOS HPC

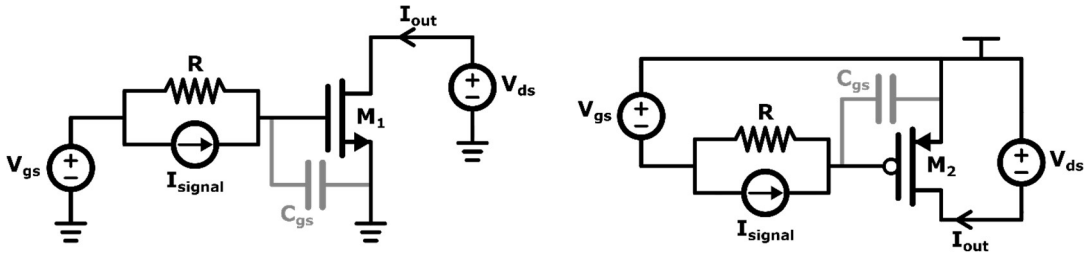
The analyzed technology parameters are

- Cut-off frequency (ft);
- Transconductance over current (gmOverId);
- Intrinsic gain (gmro);
- Thermal noise coefficient (γ)

The ft is the extrapolated frequency at which the small signal current of the MOSFET drops to unity, usually it is calculated for a common source configuration and it is related to the ratio between the transconductance gain and the parasitic capacitances of the MOSFET. The dominant capacitance of a transistor is the gate source capacitance, for that reason the cut off frequency is often approximated to the equation (3.1.1).

(3.1.1) $$f_t = \frac{gm}{C_{gs}}$$

In the preliminary analysis, we decided to perform a more precise computation of the cut-off frequency using the test bench illustrated in Fig. 3.1.2 respectively for the Nmos and for the Pmos.



(a) (b)

Fig.3.1.2 Circuit configuration for the simulation of the f_T for the NMOS (a) and the PMOS (b)

The polarization is determined by the voltage sources V_{gs} and V_{ds} while the signal current I_{signal} produces an output signal current I_{out}. As long as I_{signal} flows in the resistance R, I_{out} is constant in frequency, but when the parasitic capacitances of the MOSFET starts to absorb the input current, I_{out} starts to decrease as soon as its ratio with I_{signal} becomes equal to 1. In first approximation ft depends only to the overdrive voltage (V_{ov}) and to the carrier mobility, in particular it is equal to (3.1.2) [3.2].

(3.1.2) $$f_t = \frac{3\mu_{n/p}(V_{ov})}{4\pi L^2}$$

As we can aspect the cut off frequency increases with the overdrive voltage until the velocity saturation decreases the mobility. At that point ft reaches its maximum. Ideally we would like to polarize the MOSFET at this point to have maximum speed. In reality the cut-off frequency is not the only one that determine the performances of the circuit, but we have to take into account also the power consumption. In fact, if the chosen polarization is characterized by a low gmOverId, we have to increase the current to achieve the desired gain, determining wider routing and higher parasitic in the layout.
To compute the gmOverId, we used the same test bench and we report below the resulting plots of ft versus gmOverId for the two different version of the technology, simulated on a MOSFET with minimum length of 28 nm and width equal to 1.8 um.

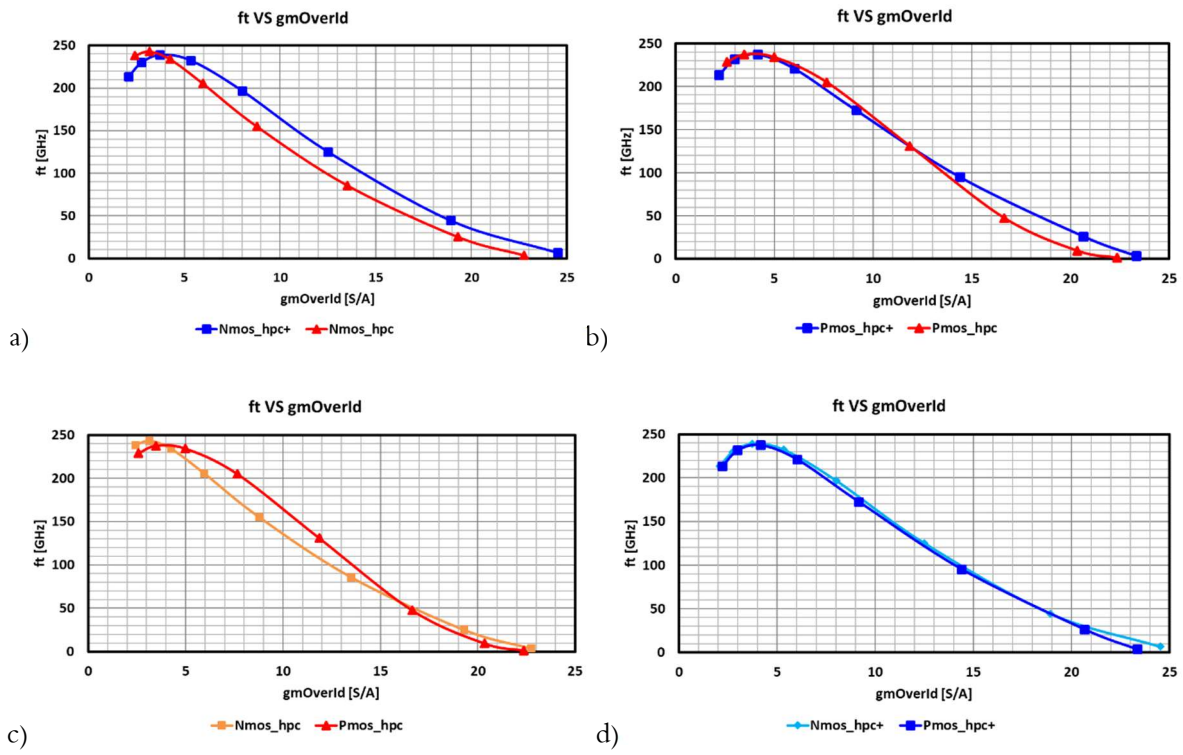


Fig. 3.1.3 Cut-off frequency versus gmOverId plots for Vds equal to 600mV: a) comparison between the 28nm HPC and HPC+ about the Nmos; b) comparison between the 28nm HPC and HPC+ about the Pmos; a) comparison between the Nmos and Pmos of the 28nm HPC and d) HPC+

As we can see from Fig. 3.1.3 the gmOverId in correspondence of the cut-off frequency peak is always around 5, it means that to achieve a transconductance gain of 50 mS we need a current equal to 10 mA, if we assume to use a metal x to carry the current, the minimum width capable to carry 10 mA in these technologies is equal to 1.4 um that can produce significant parasitic with the close metal routings or to the substrate. It suggests that even if the peak cut off frequency is around 250 GHz, in reality the chosen operating point corresponds closer to a ft between 150 GHz and 200 GHz in order to have a gmOverId closer to 10. Another important observation that we can do is that the Pmos ft-versus-gmOverId characteristics are almost equal in the two technologies, while the HPC+ present a better cut-off frequency for the Nmos. The Nmos and the Pmos are more similar in the HPC+ than in the HPC. We have to take

into account in this comparison also the voltage thresholds reported in the table below. As we can observe the HPC+ version presents a lower threshold voltage, it means that it can have lower power consumption in correspondence of the same overdrive voltage.

Vth (mV)	HPC	HPC+
Nmos	521	292
Pmos	579	374

Table 3.1.1 Voltage threshold of the Nmos and Pmos for the technology 28nm HPC and 28nm HPC+

Interesting is also the analysis of the intrinsic gain because it determines the maximum gain achievable by a single stage, if it is too limited we have to use cascoded structure or additional stages that add poles, limit the bandwidth and the power consumption. It is derived using the schematic in Fig. 3.1.2.

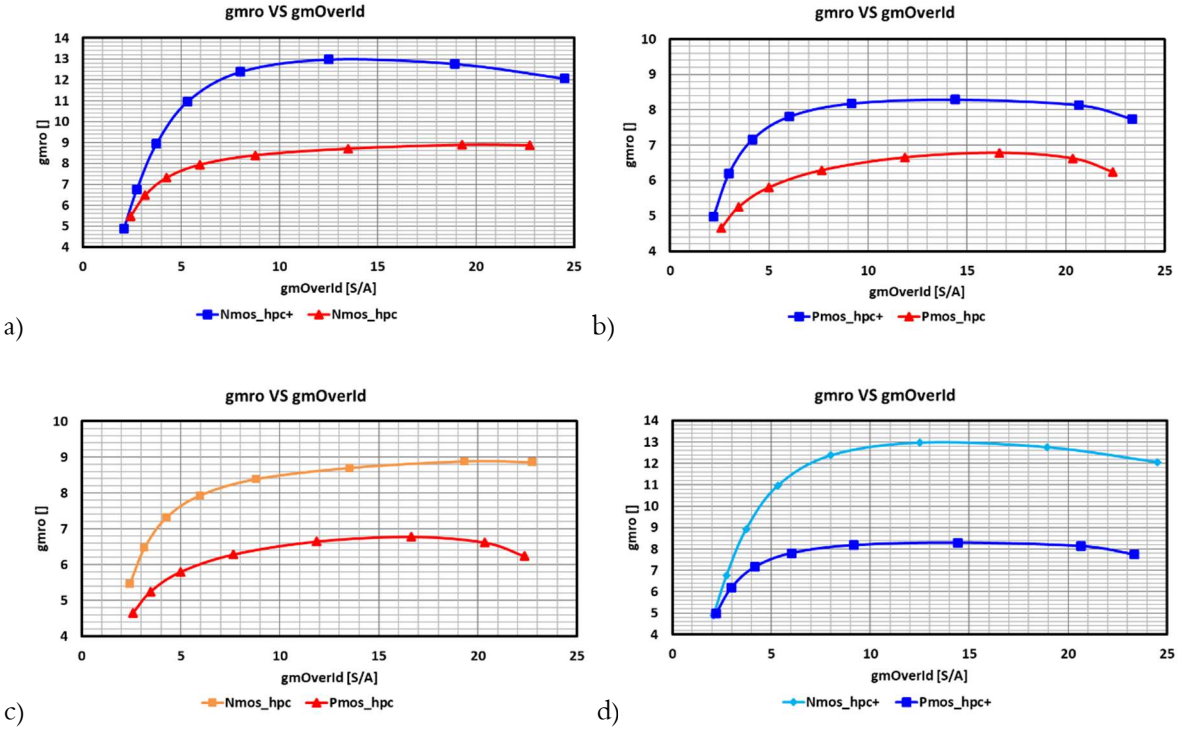


Fig. 3.1.4 Intrinsic gain versus gmOverId plots for Vds equal to 600mV: a) comparison between the 28nm HPC and HPC+ about the Nmos; b) comparison between the 28nm HPC and HPC+ about the Pmos; a) comparison between the Nmos and Pmos of the 28nm HPC and d) HPC+

We can observe from Fig. 3.1.4 that in the HPC+ technology the maximum intrinsic gain has been improved of a factor 1.5 and 1.24 respectively for the Nmos and the Pmos, even if the difference between the intrinsic gain of the Nmos and the Pmos changes from a maximum of 2.63 in the 28nm HPC to a maximum of 4.7 for the 28nm HPC+.

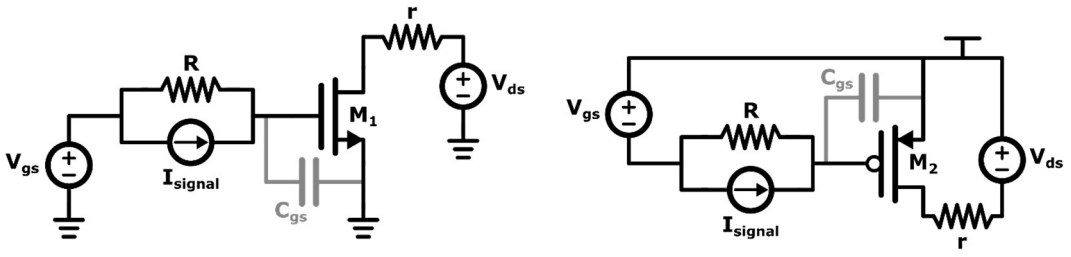
The last technology performance to evaluate is the noise of the MOSFET. If we analyze the noise of a transistor, we can identify three main contributions: the flicker noise, the noise of the gate resistance and

of the channel. Usually for high frequency applications the flicker noise is negligible while the expression of the other two contributions are equal to:

(3.1.3) $$\frac{I_{channel}^2}{Hz} = 4kT\gamma g_m$$

(3.1.4) $$\frac{V_{R_{gate}}^2}{Hz} = 4kTR_{gate}\frac{1}{3}$$

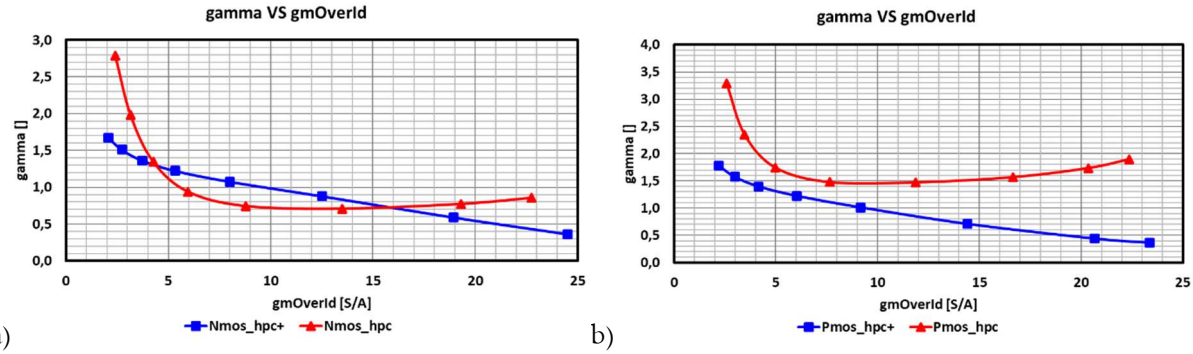
where k is the Boltzmann constant ($1.38\text{x}10^{-23}$ J/K), T is the absolute temperature (298K), R_{gate} is the physical resistance of the gate [3.3][3.4]. If the layout of the MOSFET is realized in order to have a small R_{gate}, the channel noise becomes dominant and we have to design the gm in order to satisfy the noise performances. As we can see from (3.1.3) in the noise expression there is also the thermal noise coefficient γ, that is a purely technological parameters that we studied for the 28nm HPC and HPC+ technologies. To obtain this value we used the test bench reported in Fig. 3.1.5 respectively for the Nmos and the Pmos.



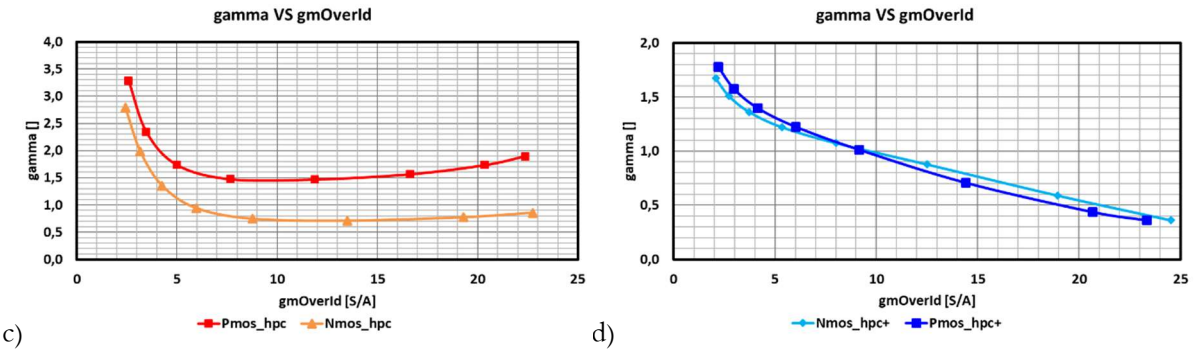
(a) (b)

Fig.3.1.5 Circuit used to evaluate the γ of the transistors for the NMOS (a) and the PMOS (b)

The resulting simulated values of γ in function of the gmOverId for the two technologies are reported in Fig.3.1.6.



a) b)



c) d)

Fig. 3.1.6 Thermic noise coefficent versus gmOverId plots for Vds equal to 600mV: a) comparison between the 28nm HPC and HPC+ about the Nmos; b) comparison between the 28nm HPC and HPC+ about the Pmos; a) comparison between the Nmos and Pmos of the 28nm HPC and d) HPC+

As we can observe, from the previous image, the noise was improved from the HPC to the HPC+ version, with the exception of the Nmos noise in the interval between 5 and 15 of the gmOverId, moreover in the HPC+ technology the noise performances of the Nmos and Pmos are more similar, in fact in the HPC we can observe that the Pmos result noisier than the Nmos of a factor close to 3 for gmOverId of 10.

3.2 Common Gate Cross Drain oTIA

The first design presented was realized in 28nm HPC CMOS technology and it is characterized by the schematic reported in Fig. 3.2.1 where we can identify a LNTA based on a Common Gate topology and characterized by a high transimpedance gain, three Variable Gain Amplifier, an Output Buffer that drives an output resistance load of 100Ω and a feedback for the DC Offset compensation.

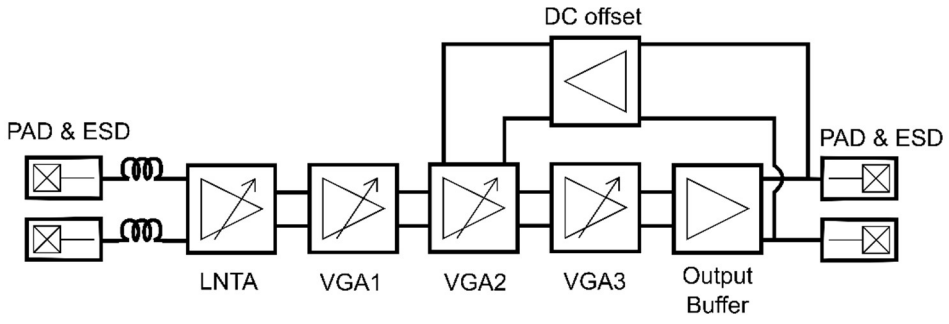
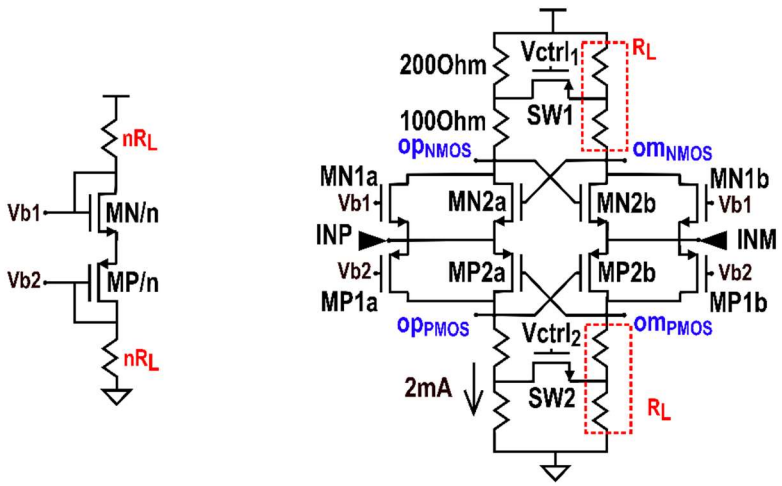


Fig.3.2.1 Schema of the Common Gate proposed TIA

The TIA was design in order to have a variable transimpedance gain that can be programmed from 78 dBΩ to 36 dBΩ maximizing the bandwidth and minimizing the noise and the distortion for an output differential peak-to-peak swing of 500 mV. In the next sections any stage will be described in detail and at the end the final results will be presented.

3.2.1 Low Noise Transimpedance Amplifier (LNTA)

As we have already introduced, the first stage is the most critical for the performances in particular of the noise and the bandwidth, but it becomes important also for the distortion at low gain condition. In chapter 2 we have highlighted the most common topologies present in literature and we have observed that usually the common gate present bad noise performances due to the contribution of the current mirror. In this design we decided to propose a variant of the common gate that overpasses this problem using the complementarity of the CMOS and a weak positive feedback is used in order to reduce the input impedance of the stage. The final schematic of the LNTA is presented in Fig. 3.2.2b. We can observe that current mirror was removed and replaced by a copy of the common gate realized by a Pmos. If, at the beginning, we ignore that MN2a-b and MP2a-b are crossed, we can observe that the signal current flows from the photodiode and is splitted in two portions to the Nmos and Pmos. If the resistances seen from the upper and lower paths are equal, the signal current is exactly divided by 2. The polarization of the circuit is provided by a scaled copy of the circuit that is represented in Fig. 3.2.2a. In particular, the MOSFET are polarized in order to have the gate voltage equal to the drain voltage, and the center node is auto-polarized in the middle of the supply voltage [3.5].



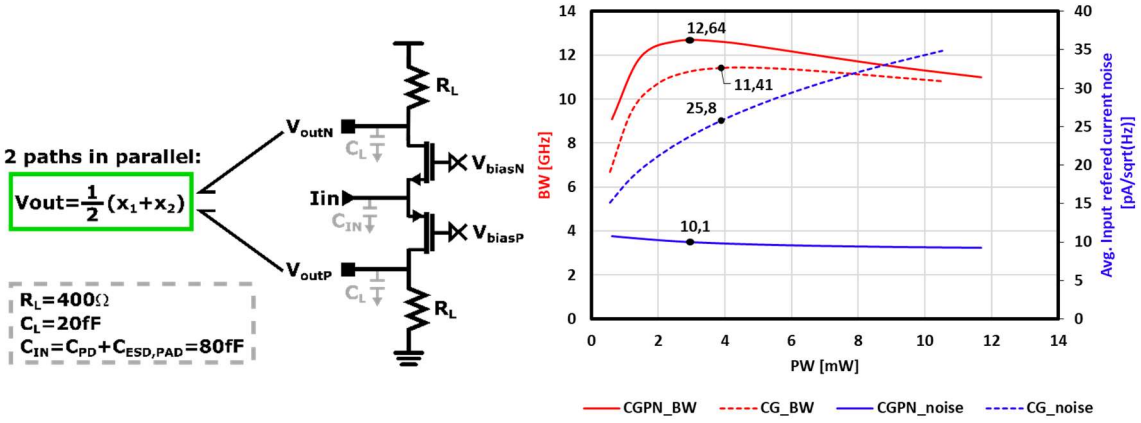
a) b)

Fig. 3.2.2 Schematic of the LNTA: a) the bias replica and b) the RF circuit [3.5]

To understand the functionality of a complementary Common Gate (CGPN), Fig. 3.2.3 represents a simplified schematic. The first observation is that it is characterized by an input impedance equal to the sum of the MOSFET transconductances, so we can have the same input impedance of a common gate with half of the current. The second observation is that the two paths (upper and lower paths) works as two independents common gate (Nmos and the other in Pmos), but without the current mirror, because the polarization is fixed by their gate-source voltage.

About the noise it also is exactly equal to the noise of a Common Gate but without the contribution of the current mirror [3.5].

To understand the advantage of the complementary common gate respect the simple common gate we can perform the same analysis realized in the chapter 2 for the common gate and the shunt feedback, also to the CGPN. To be coherent we realize this analysis with the 28nm HPC+ technology as we did for the other topologies and we evaluate the bandwidth optimum for a fixed gain of 200 Ohm, an input external capacitance of 80 fF and output external capacitance of 40 fF. We have to observe that in this topology we have two outputs that must be recombined, for this analysis we assume that the recombination is ideal, and the resistive load is set equal to 400 Ohm and the output external capacitance equal to 20 fF for any output. Fig. 3.2.3 presents the bandwidth and noise tradeoff of the CGPN and the simulated circuit compared to the results of the N-only common gate.

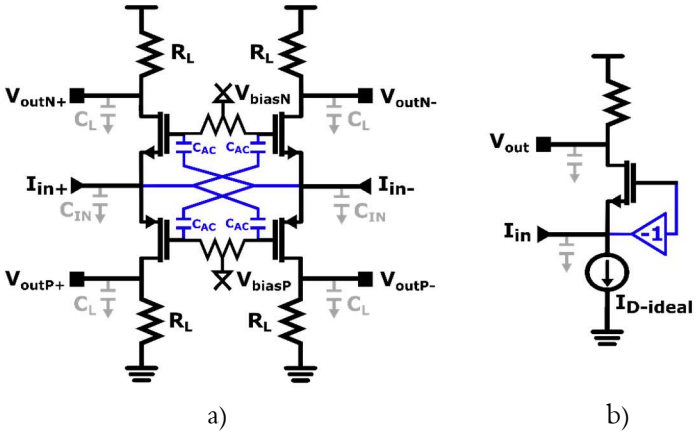


a) b)

Fig. 3.2.3 a) Schematic used to perform the bandwidth-noise tradeoff analysis, b) comparison between the bandwidth optimization of the simple common gate and the complementary common gate

As we can see from Fig. 3.2.3 the noise performance is highly improved from the simple Common Gate to the Common Gate PN as we expect without the current mirror. The power consumption in the optimum is improved of a factor 0.7. The optimum gm is very similar in the two cases (for the CG the optimum gm is equal to 18.9 mS, for the CGPN the sum of the Nmos and Pmos gm is equal to 18.4 mS) and the dc current is almost halved. Despite the reduction of the current, the power is not reduced by a factor of 2 due to the supply voltage, in fact for the simple CG the minimum V_{DD} corresponds to $V_{OV} + V_{DS} + RI_D$, while for the CGPN it corresponds to $2V_{DS} + 2RI_D$: the optimum CG has a supply voltage equal to 1.55 V while the CGPN is equal to 2.12 V. About the bandwidth it is extended by only 10% because the output pole is not significantly changed respect the common gate solution.

Different techniques to improve the complementary common gate performances are studied, the first technique is the Common Gate Cross Gate (CGxG). It is based on an AC loop used to double the equivalent input gm in order to be able to have the same input impedance with the half of the current, in order to reduce power consumption. The schematic of the CGxG is reported in Fig. 3.2.4.



a) b)

Fig. 3.2.4 a) Schematic of the differential CGxG, b) equivalent N-only single ended circuit of the CGxG in band

Also for that topology we can compute the tradeoff between noise and bandwidth keeping constant the gain of 200Ohm, an input external capacitance of 80fF and output external capacitance of 40fF. For coherency we realize this trade off in 28nm HPC+ technology and we decide to use an ideal C_{AC}.

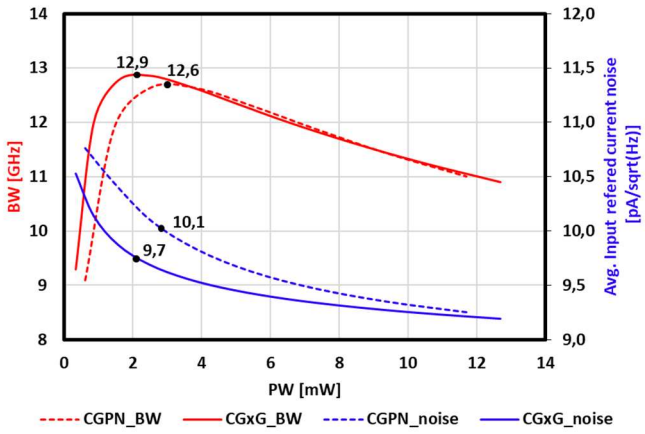


Fig. 3.2.5 Comparison between the bandwidth optimization of the complementary common gate and the common gate cross gate

As we can see from the plot the bandwidth is not significantly increase, but in the optimum point the power consumption is reduced by a factor 1.5. Also the noise is not significantly improved. An important observation about the CGxG is the Miller effect of the intrinsic capacitances of the MOSFET. Fig. 3.2.6 shows how the gate source capacitances are doubled due to two opposite voltages across them, while at the output the Miller effect on the C_{GD} and C_{DS} can be negligible if the two capacitors are equal. More over both the input nodes are loaded by all the four C_{GS} increasing the relative parasitics by another factor 2. It indicates that this solution can degrade the bandwidth if the intrinsic capacitances of the MOSFET are dominant to the external capacitance.

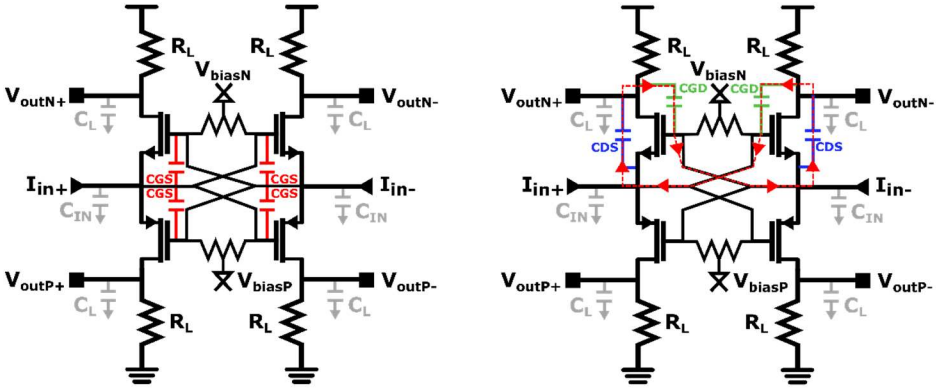


Fig. 3.2.6 a) Schematic of the differential CGxG where the Miller effect on the Cgs and b) on the Cds and Cgd are highlighted

The noise of the CGxG can be divided in resistance contribution and MOSFET contribution. As the resistive contribution is not changed by the CGPN and to the CG solutions, we can directly analyze the noise of the MOSFET and compare it to the CGPN solution. In the optimum bandwidth condition the relation $gm_{CGxG} = 1/2 * gm_{CGPN}$ is almost true, if the external capacitance dominates to the intrinsic capacitances of the MOSFET, it means that the MOSFET referred input current noise results equal to (3.2.1).

(3.2.1) $$S_{in_{CGxGMOS}} \approx 4KT\gamma g_{mCGxG} \left|\frac{sC_{in}}{2gm_{CGxG}}\right|^2 = \frac{4KT\gamma}{4g_{mCGxG}}|sC_{in}|^2 = \frac{4KT\gamma}{2g_{mCGPN}}|sC_{in}|^2 = \frac{1}{2}\frac{In_{in}^2}{Hz}_{CGPNMOS}$$

In reality this improvement is not so evident because in the bandwidth the dominant component is the resistive noise. If we look to the noise spectral density of the MOSFET contribution in the two cases, we observe instead this improvement in high frequency.

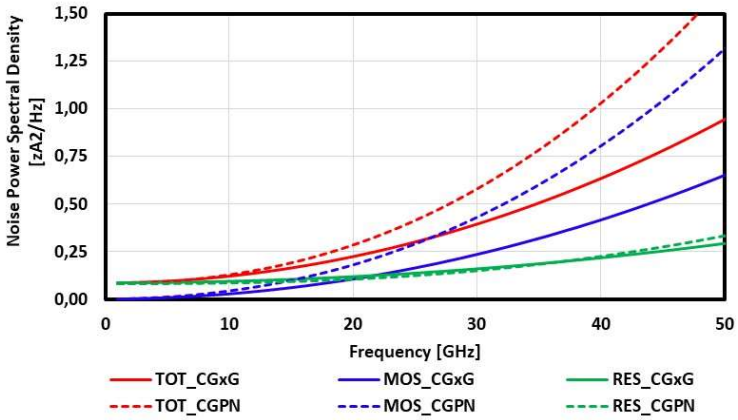


Fig. 3.2.6 Comparison between power spectral densities of the different contribution in the CGPN and in the CGxG

As we have observed the Common Gate Cross Gate doesn't provide advantages in the extension of the bandwidth. To extend the bandwidth, the Active Feedback technique was adopted in order to change

the two real poles in two complex poles and reduce the input impedance of the stage. The resulting topology was named Common Gate Cross Drain and its basic schematic is represented in Fig. 3.2.7 with its equivalent N-only, single ended representation.

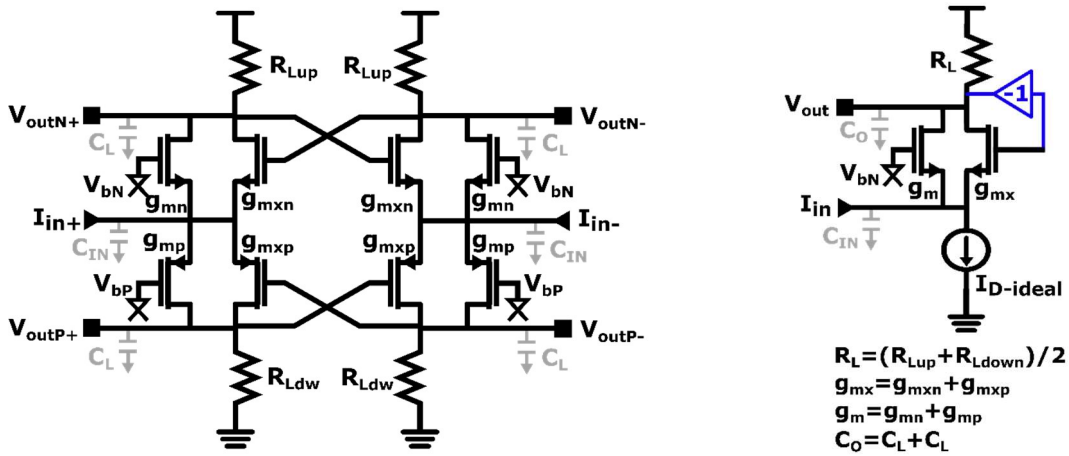


Fig. 3.2.7 Basic schematic of the Common Gate Cross Drain (a) and it equivalent N-only single ended circuit (b)

The Active Feedback is a weak positive feedback that connect the output of the stage to the gate of the MOSFET and boost the gm of the MOSFET in order to reduce the input impedance. In fact, if we refer to the equivalent N-only single ended structure, the equivalent input impedance becomes equal to (3.2.2), if we ignore the limited output resistance of the MOSFET [3.5].

(3.2.2) $$R_{INd} = \frac{2(1-\alpha)}{gm_{tot}}$$

Where gm_{tot} is the sum between the transconductance gain of the crossed and not crossed MOSFET ($gm_{tot} = gm_x + gm$) and α is equal to $gm_x R_L$, that is proportional to the maximum feedback loop gain [3.5].

The Gloop of the CGxD can be computed using the simplified circuit in Fig. 3.2.7 and computing theratio between Vo and Vo'.

(3.2.3) $$G_{loop} = \frac{Vo}{Vo'} = \frac{sgm_x R_L C_{in}}{gm_{tot}} \cdot \frac{1}{(1+sC_o R_L)\left(1+\frac{sC_{in}}{gm_{tot}}\right)}$$

It is characterized by two complex poles $p1 = \frac{1}{C_o R_L}$ and $p2 = \frac{gm_{tot}}{C_{in}}$ and a band pass gain.

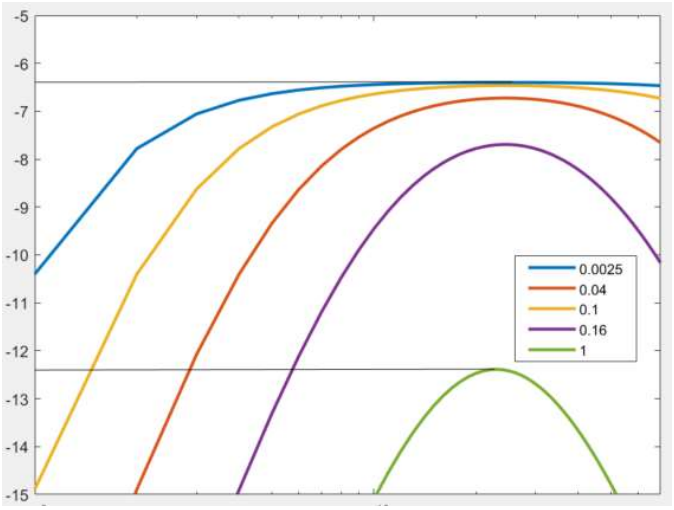


Fig. 3.2.8 Qualitative representation of the Gloop.

A qualitative representation is proposed in Fig. 3.2.8 for different values of $p1/p2$. Assuming that the pole $\frac{gm_{tot}}{C_{in}}$ is at lower frequency than $\frac{1}{C_o R_L}$, condition usually verified due to the big input photodiode capacitance, we can observe that the higher gain of the loop is equivalent to the gain after the dominant pole and, if we push the secondary pole at infinite (orange case), it corresponds to exactly $\alpha = gm_x R_L$. If the two poles are equal (green case), the maximum loop gain corresponds to $gm_x R_L/2$. In our design we decided to set $\alpha = gm_x R_L = 0.5$ to have enough gain margin for stability condition.

Fig. 3.2.9 represents the input impedance at the input node with and without the crossed transistors. We can observe that the impedance is effectively reduced, in this design it was not reduced by 6dB, but only 3dB due to the finite output impedance of the transistor which is not negligible in scaled technologies [3.5].

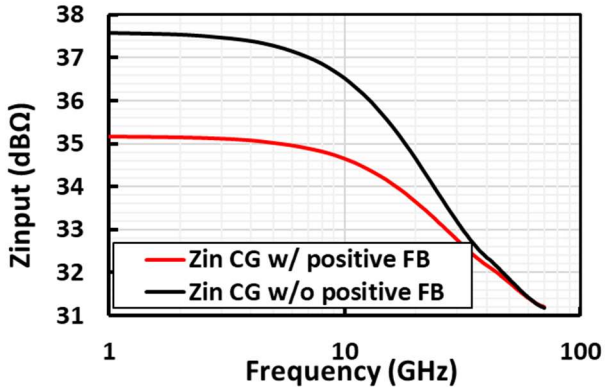


Fig. 3.2.9 Input impedance of the CGxD with and without positive feedback [3.5]

To design the CGxD we can consider the transfer function that results equal to

$$\frac{V_o}{I_{in}} = R_L \cdot \frac{1}{1 + s \cdot \left(\frac{C_{in}}{gm_{tot}} + C_o R_L - \frac{C_{in} R_L gm_x}{gm_{tot}} \right) + s^2 \cdot \frac{C_{in} C_o R_L}{gm_{tot}}} \tag{3.2.4}$$

Where we can identify two complex poles at $\omega_o = \frac{gm_{tot}}{C_{in}C_oR_L}$ with a quality factor of $Q = \frac{1}{\omega_0\left(\frac{C_{in}}{gm_{tot}}+C_oR_L-\frac{C_{in}R_Lgm_x}{gm_{tot}}\right)}$, it presents a negative contribution introduced by gm_x, that for stability issues must be small.

About the noise, if we consider the equivalent N-only schematic for simplicity, the equivalent input current noise is equal to (3.2.5) neglecting the output resistance of the transistors.

$$S_{in} = \frac{4KT}{R}\left|(1 + s\frac{C_{in}}{gm_{tot}})\right|^2 + 4KT\gamma gm_{tot}|1 + sC_{in}|^2 \quad (3.2.5)$$

It's the same as the noise contribution of the complementary common gate and the N-only common gate without the current mirror, therefore the previous considerations are still valid.

The complete schematic for the final design provided in [3.5] is shown in Fig. 3.2.2, where we can see a 300Ohm output resistance that can be changed by switches (SW1 and SW2) to limit the output voltage swing when the input signal grows and introduces distortion.

To minimize the load capacitance, the four outputs are recombined by a PN source follower. The combiner's schematic is shown in Fig. 3.2.10 [3.5].

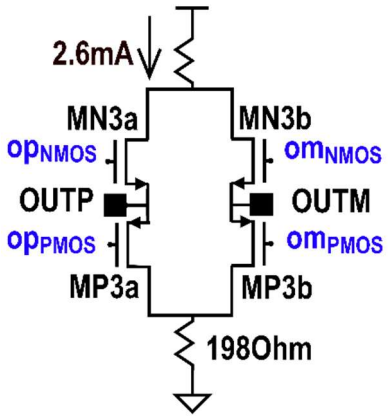


Fig. 3.2.10 Schematic of the complementary source-follower buffer [3.5]

3.2.2 Variable Gain Amplifiers (VGAs)

Variable Gain Amplifiers are made with a complementary transconductor and a folded cascode in Gilbert Cell configuration to achieve gain flexibility. To reduce the output capacitive load and allow DC coupling between the amplifiers, any VGA is followed by a source follower. Figure 3.2.11 illustrates its schematic [3.5].

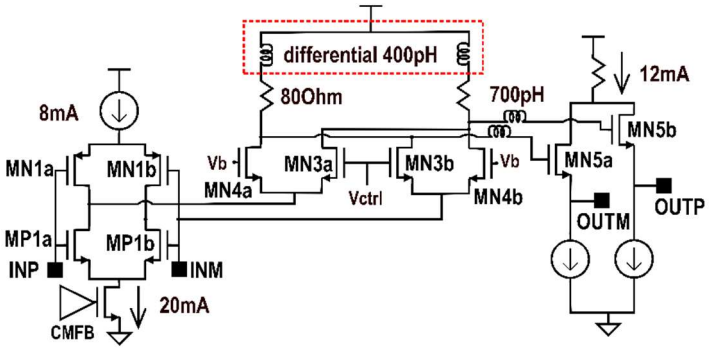


Fig. 3.2.11 Schematic of the Second Variable Gain Amplifier [3.5]

The complementarity of the transconductors allows to have the same gain to the N-only common source but half the current, resulting in a reduced voltage drop across the resistive load. As we have seen in Chapter 2, the size ratio between MN3 and MN4 determines the variability of a Gilbert Cell's gain. In this situation, the three VGAs are equal in polarizations and execute a gain variation of 12 dB starting from a maximum gain of about 8 dB, taking into consideration the source follower losses. In comparison to traditional Current Steering, the Gilbert Cell has better linear performance at low gain since the required current steering ratio is smaller [3.5]. The VGAs have a Triple Resonance Network used to recover the limited bandwidth of the LNTA. Any amplifiers have different inductances optimized for the bandwidth and the flatness of the transfer function. Large shunt inductances are used to achieve the necessary peaking. To minimize the coupling capacitances between the spirals and improve the resonance frequency of the large shunt inductances, they are realized using two distinct metal layers (M7 and M8). Metal 7's higher resistivity reduces the inductor's quality factor, but the series losses are absorbed by load the series resistance of the VGA [3.5]. Figure 3.2.12 illustrates the 3D layout of a 400 pH shunt inductor and the comparison of the resonance frequencies between the same inductance realized using just Metal 8 and using both the Metal layers, showing a factor 1.18 improvement [3.5].

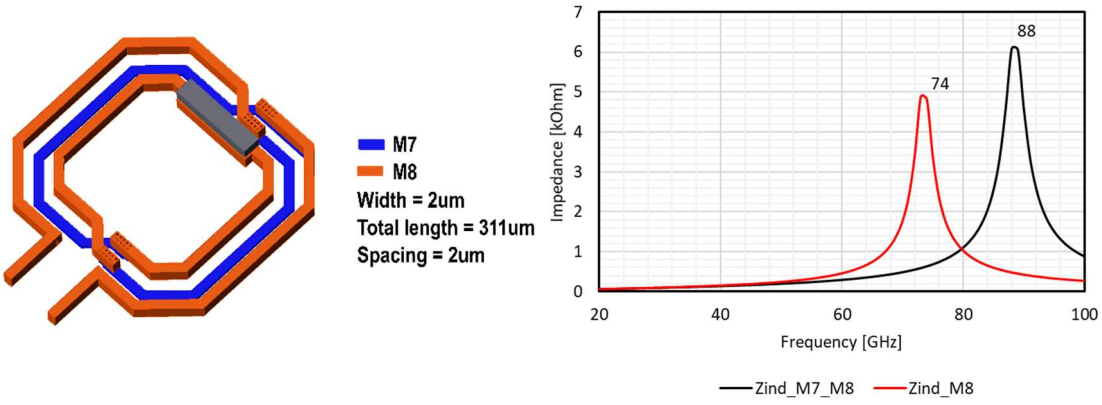


Fig. 3.2.12 a) Layout of the 400 pH shunt inductor, b) Impedance versus frequency of the 400 pH inductor realized in metal 8 (red curve) and in metal 7 and 8 (black curve) [3.5]

3.2.3 Output Buffer

At the output of the TIA there is an Analog Digital Converter that requires to be driven by 50 ohm single ended (100 ohm differential), this stage has to purpose to provide this matching. The Output Buffer must not limit the bandwidth of the TIA and must have acceptable linear performance since it is subjected to the largest voltage swings of the oTIA. As a result, this stage is intrinsically characterized by a tradeoff between bandwidth and distortion: higher gain means lower swing at its input, and thus lower distortion; however the gain is determined by transconductance, which means that to increase it, we must increase the gm, and thus the input capacitance, which limits the bandwidth of the last VGA. The tradeoff between distortion and bandwidth is somewhat overcome in the suggested design by utilizing an active load instead of an explicit 50 resistance, as shown in Fig. 3.2.13 [3.5].

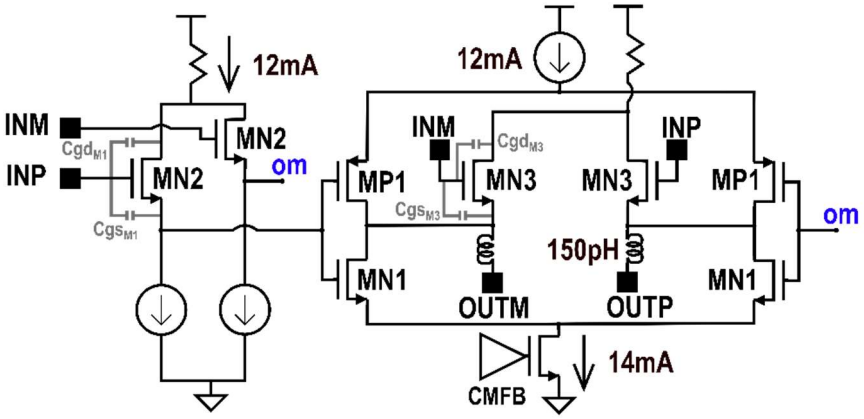


Fig. 3.2.13 Schematic of the Output Buffer [3.5]

The source follower at the output of the third VGA is identified by the MOSFET MN2, a transconductance realized by MP1 and MN1 and the active termination (MN3) that exhibits 1/gm as output impedance and it is design to be equal to 50Ω. The active termination has two main benefits. For start, it has a greater gain than a simple transconductance with a passive load with the same gm. In fact, instead of gm*50Ω/2, which is the gain of a passive resistive terminated common source [3.5], the active terminated stage gain is equal to (3.2.6).

$$Gain_{1G} = \frac{1 + gm * 50\Omega}{2} \tag{3.2.6}$$

The output buffer in the presented design has a gain of 5dB, which means that a standard resistive load stage would require 40% more transconductance and input capacitance, if we neglect the output impedance of the transistor in first approximation. The active termination's second advantage is the input capacitance provided by the active load; its gate-source capacitance is affected by a positive gain between, allowing it to load the previous stage with a negative capacitance and thus increase the bandwidth [3.5].

3.2.4 Input DC Current Cancellation

The signal is received by the TIA from a photodiode, which must be polarized in order to switch on and convert the signal light from the optical fiber to the electrical domain. The photodiode generates a DC current that must be absorbed by the TIA. To enable this extra feature, it usually incorporates a DC control loop. The complementarity of the proposed LNTA allows to directly absorb a DC current that is a fraction of its bias current. The CGxD has a current of 2 mA for every single ended branch in this design, and the Pmos may absorb up to 1 mA without reducing the transfer function appreciably [3.5]. The simulated post layout S21 with and without the DC photodiode current is shown in Fig. 3.2.14, and the two curves are well-matched [3.5].

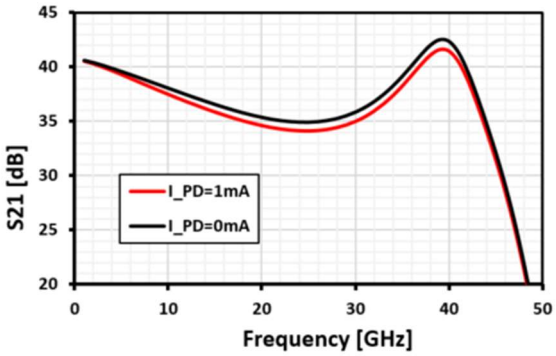


Fig. 3.2.14 S21 with and without the DC photodiode current [3.5]

3.2.5 Measurements Results

The prototype of the Common Gate Cross Drain Transimpedance Amplifier was realized in 28 nm CMOS TSMC-hpc technology in a chip area of 1mm x 1mm including the I/O pads. The photo of the chip is proposed in Fig. 3.2.15 where the stages are highlighted.

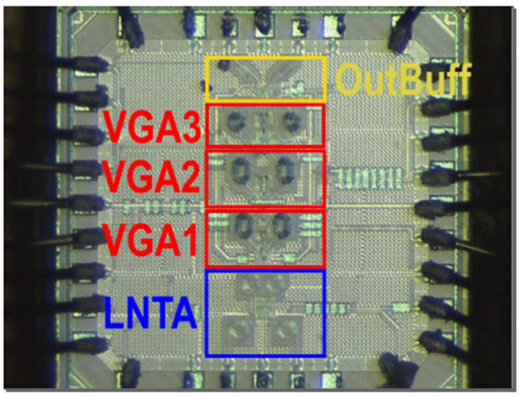


Fig. 3.2.14 Photo of the CGxD_TIA chip [3.5]

The circuit has a bias voltage of 2.4 V and a power consumption of 319 mW. The measurements are made at 20°C with a Keysight N5247B Network Analyzer, and the Four-port S-parameters are measured up to 50 GHz. The S-parameters measured and simulated are shown in Fig. 3.2.15, where we can see that the S22 results less than 10 dB across the whole bandwidth and match the simulations. The S21 also matches the simulations well at maximum and minimum gain.

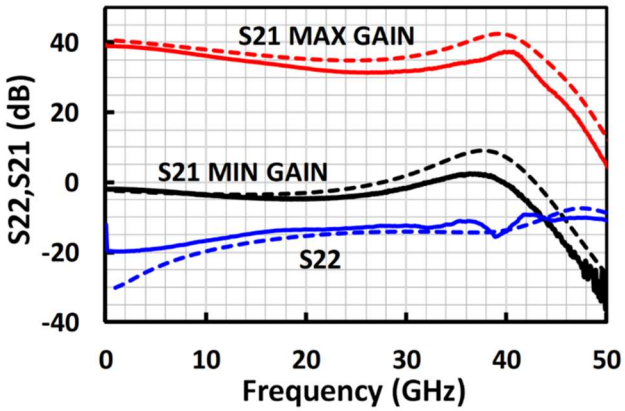
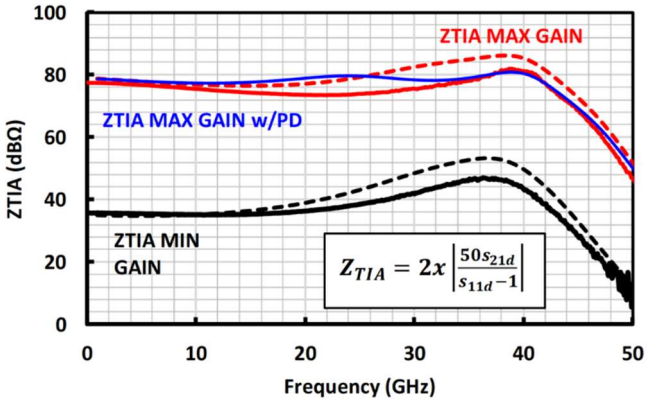
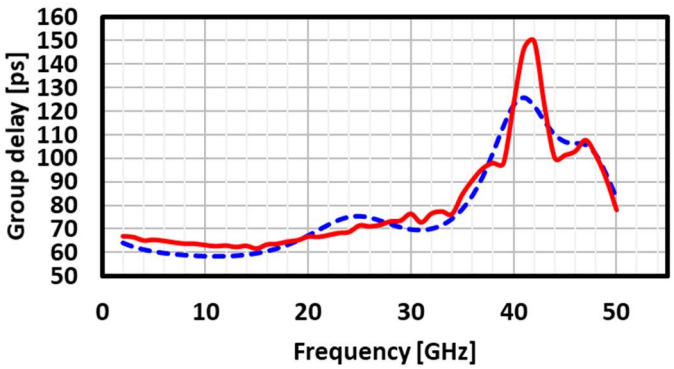


Fig. 3.2.15 Measured (solid line) and simulated (dotted line) S22 and S21 at maximum and minimum gain [3.5]

The TIA was designed for a typical photodiode with a 150 pH bondwire equivalent inductance. The measured $Z_{TIA_{tf}}$ transfer function with the photodiode and the simulated Z_{TIA} computed from the S-parameters by the formula $Z_{TIA} = 2x\left|\frac{50S_{21d}}{S_{11d}-1}\right|$ are shown in Fig. 3.2.16 (a), where we can see that the gain can vary from 78dB to 36dB, The transimpedance amplifier with the photodiode has a bandwidth of 42GHz and an in band ripple of 1.7dB, with an average group delay of 105ps and a ripple of 52.5ps (Fig. 3.2.16 (b)) [3.5].



(a)



(b)

Fig. 3.2.16 (a) Measured (solid line) and simulated (dotted line) Z_{TIA} and $Z_{TIA_{tf}}$, (b) Simulated (dotted blue line) and measured (solid red line) group delay

The noise was measured by integrating the output single ended noise from 1 GHz to 50 GHz, producing 0.176mVrms2, and the average input current noise was calculated using the formula (3.2.7). The average input noise density is 18pA/Hz.

(3.2.7)
$$I_{n,in,avg} = \sqrt{\frac{\int_0^{BW} \frac{V_{out,SE}^2}{Hz} df}{|Z_T@1GHz|^2 \cdot BW}} = \frac{V_{n,out,SE}}{3981\Omega * \sqrt{42GHz}}$$

The LNTA is the largest source of noise, accounting for 60% of the total, according to the simulation, while the first VGA contributes 30%. For different gain settings, distortion was measured at 1GHz while keeping an output swing of Vppd=500mV constant. The THD is plotted as a function of the equivalent input current using the formula in Fig. 3.2.17. (3.2.8).

(3.2.8)
$$I_{NNPP} = 2\sqrt{\frac{2P_{IN,SE}}{50}}|1 - s_{11d}|$$

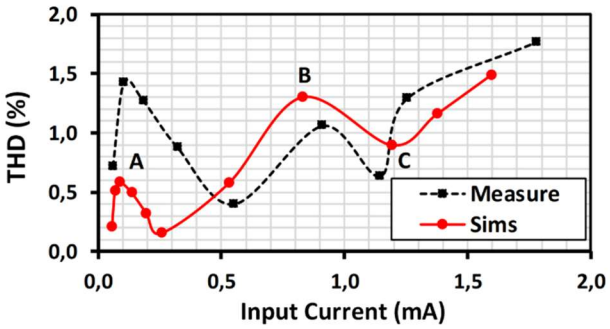


Fig. 3.2.17 Total Harmonic Distortion in function of the equivalent Input Current for different gain setting and a constant output swing of 500 mVppd

Due to second harmonic distortion caused by layout mismatches and asymmetries, the measured THD is larger than the simulated. The THD's non-monotonic behavior, which can also be seen in simulation, can be explained: the distortions of the Gilbert Cell's cascode in the week inversion region describe point A; when the transistors are fully turned on, the THD returns to low values. Because the VGA gains are all reduced at the same time, this phenomenon only occurs once. As the input signal is increased, the swings become more pronounced, and the THD rises. When the VGAs' gain is 0 dB, the LNTA lowers its gain, lowering the signal swing at the three VGAs and causing the second local minimum in the THD from B to C. At minimum gain, the maximum measured Total Harmonic Distortion is equivalent to 1.77%. The authors of [3.6] conducted an investigation to establish the compatible QAM modulation in relation to the THD and Optical Power Penalty, resulting that a THD of 1.77% should be able to sustain a 16 QAM modulation. We used the Matlab software described in Chapter 1 to check the Bitrate by inserting the E/O transfer function, the SNR, and the distortion. The SNR is calculated using the Vrms noise and the signal single ended at the output, resulting in a value of 16dB. An example of a noisy sent bit filtered by the transimpedance amplifier is shown in Fig. 3.2.18. We can see that the signal is substantially delayed by Δt and is also low pass filtered. The ringing after any pulse response, which might affect the readability of the bit stream, is another crucial consequence to notice of the not constant group delay.

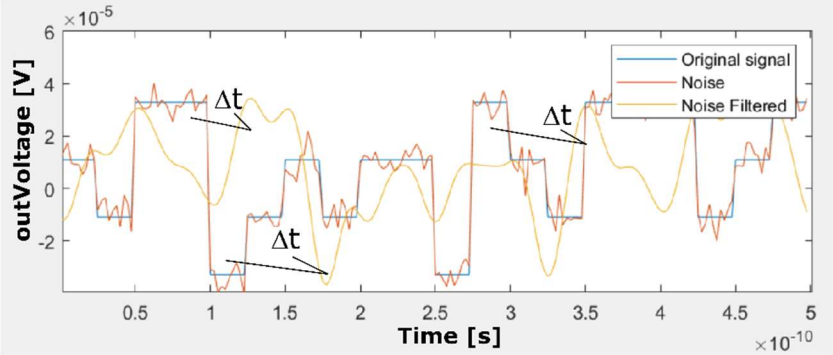


Fig. 3.2.18 bit stream of the In-phase component of a 16 QAM signal sended with 40 GBaud/s. In blue there is the original bit, in orange a white noise with 16 dB SNR, the yellow line is the noisy signal filtered by the transfer function.

We can acquire a BER equivalent to the results in the following table using the proposed TIA for 16, 32, and 64 QAM modulations. Because the suggested program can only evaluate BER values greater than 10^{-6}, the BER for 16 QAM and 160 Gbit is equivalent to 0.

Modulation	16 QAM	16 QAM	32 QAM	64 QAM
N. of sended Bit	4000000	4000000	5000000	6000000
BER	$< 2{,}5\text{x}10^{-7}$	$2{,}55\text{x}10^{-2}$	$2{,}302\text{x}10^{-4}$	0.0113
BaudRate	40GB/s	50GB/s	40GB/s	40GB/s
BitRate	160Gb/s	200Gb/s	200Gb/s	200Gb/s

Table 3.2.1 Modulation and BitRate achievable by the Common Gate Cross Drain TIA

	JSSC '18 [3.7]	**ISSCC '16 [3.8]**	**JSSC '19 [3.6]**	**ESSCIRC '18 [3.9]**	**JSSC '19 [3.10]**	**This work**
Linear/Limit.	Linear	Linear	Linear	Linear (PAM-4).	Linear (PAM-4)	Linear
Technology	0.13um BiCMOS	0.13um BiCMOS	0.13um BiCMOS	28 nm CMOS	16 nm FinFET	28nm CMOS
Supply Voltage	3.3 V	3.3 V	3.3 V	2.5 V / 1.2 V	1.8 V	2.4 V
SE/S2D/Diff.	Diff.	Diff.	Diff.	S2D	S2D	Diff.
Gain	65 dBΩ	80 dBΩ	73 dBΩ	65 dBΩ	78 dBΩ	78 dBΩ
Bandwidth	66 GHz	53 GHz	27 GHz	60 GHz	27 GHz	42 GHz
Noise	7.6 pA/ √ Hz	24.86 pA/ √ Hz	20 pA/ √ Hz	19.3 pA/ √ Hz	16.7 pA/ √ Hz	18 pA/ √ Hz
THD	<5% For max gain and 800 mVppd	4.13% for 900 mVppd	1.5% For 500 mVppd	<5% for 300mVppd	1.8% For 600 mVppd	1.77% For min gain and 500 mVppd
Power	150 mW	277 mW	313 mW	107 mW	60.8 mW	319 mW
BitRate	100 Gb/s	128 Gb/s	200 Gb/s	112 Gb/s	100 Gb/s	200 Gb/s
Energy per bit	1.5 pJ/bit	2.16 pJ/bit	1.5 pJ/bit	0.96 pJ/bit	0.6 pJ/bit	1.59 pJ/bit
Overload current compensation	-	YES	YES	YES	YES	YES
FOM [GHz·Ω/mW] *	782.4	1913.4	385.3	997.2	3527.4	1045.8

Table 3.2.2 TIA performance comparison and summary

Table 3.2.2 reports a summary of the performances of the proposed oTIA to some works present in literature. More in detail, respect to [3.7], which reports the lowest noise and the widest bandwidth, our design has 2.3 time higher average noise and 30% lower bandwidth, but a THD that is better by a factor of 3. Improving linearity inevitably leads to noise and bandwidth penalties. However, thanks to the

possibility of high order modulations, higher data rates could be achieved [3.6]. Comparing the Common Gate Cross Drain oTIA to [3.8] noise and distortion are improved by lowering 18% the bandwidth. Compared to the work of G. Ahmed and co. [3.6], which is also compatible with 16 QAM, they presents a bit higher noise, but a lower distortion, while the proposed TIA presents a bandwidth higher by 59%. The TIA of H. Li and co. [3.9] is also implemented in 28 nm CMOS but was optimized for PAM4, our noise is slightly better whereas bandwidth is lower by 28%. However, THD is 3 times lower for an output swing that is almost double, important to remember is that the THD is strongly affected by the type of modulation used, and it also affects the power consumption. In fact, [3.9] is the most power efficient solution also thanks to specific design choices, such as the lower transresistance gain (13 dB less), and the lower voltage output dynamic, that allows to use a 1.2 V supply for the power hungry portion of the circuit. If we compare to [3.10], which presents the highest figure of merit (FOM), defined in [3.11], the proposed TIA presents almost the same THD and transresistance gain and has 1.5 times more bandwidth. The higher FOM is justified by the more power efficiency, also thanks to the lower supply voltage, in the proposed design we couldn't reduce it due to the chosen topologies, in particular of the LNTA.

3.3 Shunt Feedback Feed Forward oTIA

The second design was implemented in 28nm HPC+ CMOS technology, and its block diagram is shown in Figure 3.3.1.

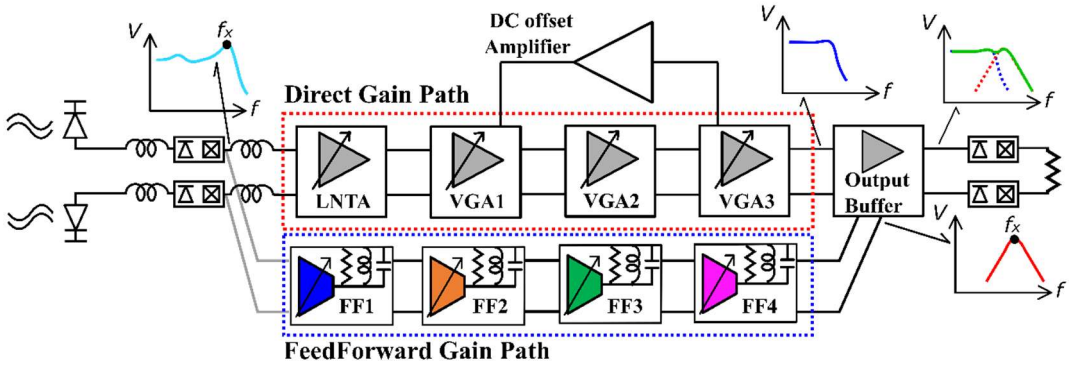


Fig.3.3.1 Scheme of the Shunt-Feedback Feed-Forward TIA

We can see two parallel paths: one is focused on amplifying the signal at low frequency (Direct path, DG), while the other is focused on amplifying the signal at high frequency (Feed-Forward path, FFWG). At the end of the oTIA, there's an Output Buffer that sums the two parallel paths and provides an output impedance of 100Ohm differential. The sum of the two gains is the final transfer function. The sum of the two paths is quite hard to analyze, but we can distinguish two main cases: in the first example the slopes of the two transfer functions are moderate; this is the most difficult scenario to investigate since

there is a large frequency interval when the two paths have comparable gain. The second example is when the two tracks have a high roll off and roll on slope; this is the simplest since one path is negligible in comparison to the other for the vast majority of the frequencies, and they are only comparable for a narrow interval of frequencies. The last scenario is similarly to the case of the proposed design, where the direct and feed forward paths are defined by eighth-order denominators.

The first situation, on the other hand, is more difficult to study and construct, but we can begin to comprehend it using a simple example in which the direct path is of first order and the feed forward path is defined by a second order transfer function.

(3.3.1) $$TF_{Dir} = \frac{1}{1 + s/\omega_0}$$

(3.3.2) $$TF_{FF} = \frac{\tau s}{(1 + s/\omega_1) \cdot (1 + s/\omega_2)}$$

Where $\omega_2 > \omega_1 > \omega_0$. The total transfer function is equal to:

(3.3.3) $$TF_{FF} = \frac{1}{1 + s/\omega_0} + \frac{\tau s}{(1 + s/\omega_0) \cdot (1 + s/\omega_2)} = \frac{1 + s(1/\omega_2 + 1/\omega_1 + \tau) + s^2(1/\omega_1\omega_2 + \tau/\omega_0)}{(1 + s/\omega_0)(1 + s/\omega_1)(1 + s/\omega_2)}$$

We can see that the three poles of the original transfer functions are the same, but at the frequency $\omega_z = \sqrt{\omega_0\omega_1\omega_2/(\omega_0 + \tau\omega_1\omega_2)}$ two complex zeros are introduced. If we assume that $f_0 = \omega_0/2\pi$ is the bandwidth that we wish to increase and that $f_2 = \omega_2/2\pi$ is the maximum bandwidth that the Feed-Forward link can provide, we must design ω_1 and τ to minimize the ripple.

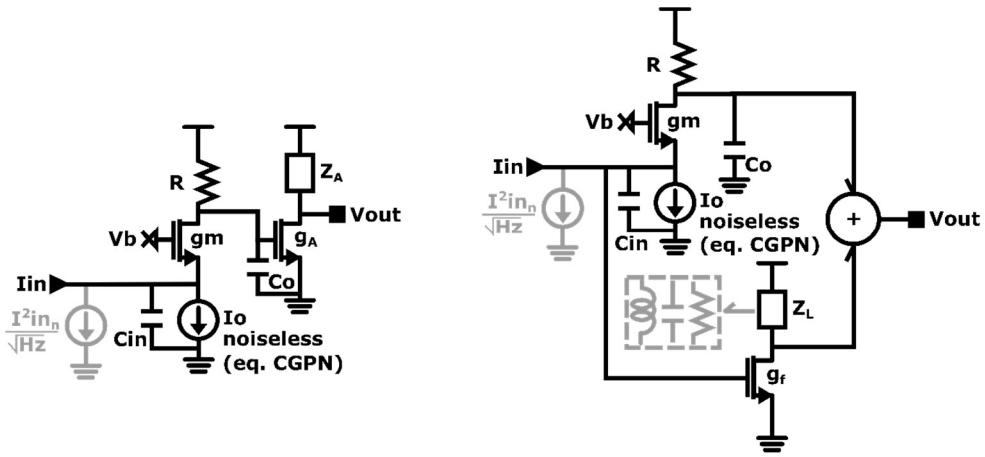


Fig.3.3.2 Basic schematic of the bandwidth extension by peaking provided by next stages (a) and by feed forward techniques for noise study

To investigate the benefits of feed forward in terms of noise, we looked to Fig.3.3.2 and begin by examining a standard condition with a TIA (for simplicity, imagine a common gate) cascaded by an

amplifier (a). When we look at the noise reported at the input node, we get (3.3.4), ignoring the noise from the amplifier load resistance Z_A.

$$S_{in} = \frac{4kT\gamma}{gm} \cdot s^2 C_{in}^2 + 4kTR + \frac{4kT\gamma}{g_A} \left| \frac{(1 + sRC_o)\left(1 + \frac{sC_{in}}{gm}\right)}{R} \right|^2 \tag{3.3.4}$$

Where the red part is the common gate contribution already studied and the blue part is the amplifier noise reported to the input by the TIA transfer function. If the amplifier provides some peaking, the bandwidth is extended, and the noise has to be integrated above the frequencies $\frac{1}{2\pi RC_o}$ and $\frac{gm}{2\pi C_{in}}$, where the noise of the amplifier can be approximated in first order to $\frac{4kT\gamma}{g_A} \cdot \frac{s^4 C_o^2 C_{in}^2}{gm^2}$, which increases with s^4, that appends because its noise is divided by the TIA gain that is going down with the frequency. This implies that the bandwidth extension of the VGA must be limited for noise performances. If we referred instead to Fig.3.3.2b, it is the same Common Gate but the bandwidth extension is realized by a Feed Forward stage that takes the signal before the TIA. Because the two paths transport the noise to the output with a gain, but then it is reported back to the input with the total of the two paths, the tradeoff between bandwidth and noise is different. The noise density of the whole input current is equal to (3.3.5).

$$S_{in} = \frac{V_{TIA_{nout}}}{\left| \frac{R}{(1 + sRC_o)\left(1 + \frac{sC_{in}}{gm}\right)} + g_f Z_{inCG} Z_L \right|^2} + \frac{4kT\gamma g_f |Z_L|^2}{\left| \frac{R}{(1 + sRC_o)\left(1 + \frac{sC_{in}}{gm}\right)} + g_f Z_{inCG} Z_L \right|^2} \tag{3.3.5}$$

Where $V_{TIA_{nout}}$ is the noise of the common gate at its output, Z_{inCG} is the gain from the photodiode current to the input of the FFW and $g_f Z_{inCG} Z_L$ is the transfer function of the feed forward path that is represented as a common source with a resonance LCR tank, or Z_L. Assume you're in the second condition, where the slopes of the Direct and Feed-Forward paths are sufficiently steep that we can identify two intervals of frequencies where one path is negligible in comparison to the other one.
At low frequencies, when $g_f Z_L = 0$, the noise becomes:

$$S_{in} \approx \frac{V_{TIA_{nout}}}{\left| \frac{R}{(1 + sRC_o)\left(1 + \frac{sC_{in}}{gm}\right)} \right|^2} = \frac{4kT\gamma}{gm} \cdot s^2 C_{in}^2 + 4kTR \tag{3.3.6}$$

It is exactly equal to the noise of only the Common Gate.
While at high frequency we assume $\frac{R}{(1+sRC_o)\left(1+\frac{sC_{in}}{gm}\right)} \ll g_f Z_{inCG} Z_L$ and the noise expression is:

$$S_{in} \approx \frac{V_{TIA_{nout}}}{\left| g_f Z_{inCG} Z_L \right|^2} + \frac{4kT\gamma}{g_f} \cdot \frac{1}{|Z_{inCG}|^2} \tag{3.3.7}$$

We can suppose that $g_f Z_{inCG} Z_L$ is a band pass transfer function that must meet the requirement $g_f Z_{inCG} Z_L = R$ in the middle of the band to keep the overall transfer function flat. Furthermore, $V_{TIA_{nout}} = \frac{4kT\gamma}{gm} \cdot \frac{s^2 R^2 C_{in}^2}{\left|(1+sRC_o)\left(1+\frac{sC_{in}}{gm}\right)\right|^2} + \frac{4kTR}{\left|(1+sRC_o)\left(1+\frac{sC_{in}}{gm}\right)\right|^2}$. Under these assumptions and considering that we are looking outside the TIA bandwidth, the expression (3.3.7) can be simplified to

(3.3.8) $$S_{in} \approx \frac{4kT\gamma gm}{s^2 R^2 C_o^2} + \frac{4kT gm^2}{s^4 R^3 C_o^2 C_{in}^2} + \frac{4kT\gamma}{g_f} \frac{1}{|Z_{incG}|^2}$$

To summarize, at low frequencies, the respective noise in the peaking bandwidth extension and feed forward extensions are (3.3.9) and (3.3.10).

(3.3.9) $$S_{in} = \frac{4kT\gamma}{gm} * s^2 C_{in}^2 + 4kTR + \frac{4kT\gamma}{g_A R^2}$$

(3.3.10) $$S_{in} = \frac{4kT\gamma}{gm} * s^2 C_{in}^2 + 4kTR$$

Taking into account that in the reality we could have a next stage also the feed forward case, we can adfirm that the two solutions presents the same low frequency noise as long as the input capacitance C_{in} is not significantly raised by the feed forward.

For the high frequency noise outside the TIA bandwidth, the peaking and feed forward noise are equal to (3.3.11) and (3.3.12).

(3.3.11) $$S_{in} = \frac{4kT\gamma}{gm} \cdot s^2 C_{in}^2 + 4kTR + \frac{4kT\gamma}{g_A} \cdot \frac{s^4 C_o^2 C_{in}^2}{gm^2}$$

(3.3.12) $$S_{in} = \frac{4kT\gamma}{gm} \cdot \frac{gm^2}{s^2 R^2 C_o^2} + 4kTR \frac{gm^2}{s^4 R^4 C_o^2 C_{in}^2} + \frac{4kT\gamma}{g_f} \cdot \frac{1}{|Z_{incG}|^2}$$

As can be seen, the noise contribution due to the common gate (red color) increases with frequency as s^2 in the inductive peaking [3.15], while on the contrary it is attenuated in the feed forward solution. Simultaneously, we add feed forward noise (green color), which is dependent to the selected input node of the Feed-Forward. While the noise of the amplifier in the peaking bandwidth extension solution (blue color) grows as s^4 because it is divided by the transfer function of the common gate, that after the two real poles, continues to decrease.

As previously described, the feed forward is a stage that is added to the input node, which means it loads the input, raising the input capacitance and limiting the LNTA's bandwidth. It implies that feed forward bandwidth extension techniques can be effective if the bandwidth limitation is lower than the final extension. Furthermore, the signal received from the feed forward must be carefully selected in order to benefit of noise advantages. The selected node in this design, for example, is the one between the

photodiode, ESD, and Pad and the input inductance such that the input feed forward signal is boosted by the resonance between the inductance and capacitance.

3.3.1 Low Noise Transimpedance Amplifier (LNTA)

The shunt-feedback topology is employed for this design, and its complete schematic is provided in Fig. 3.3.3, where we can identify the topology previously discussed in Chapter 2, the shunt feedback inverter based.

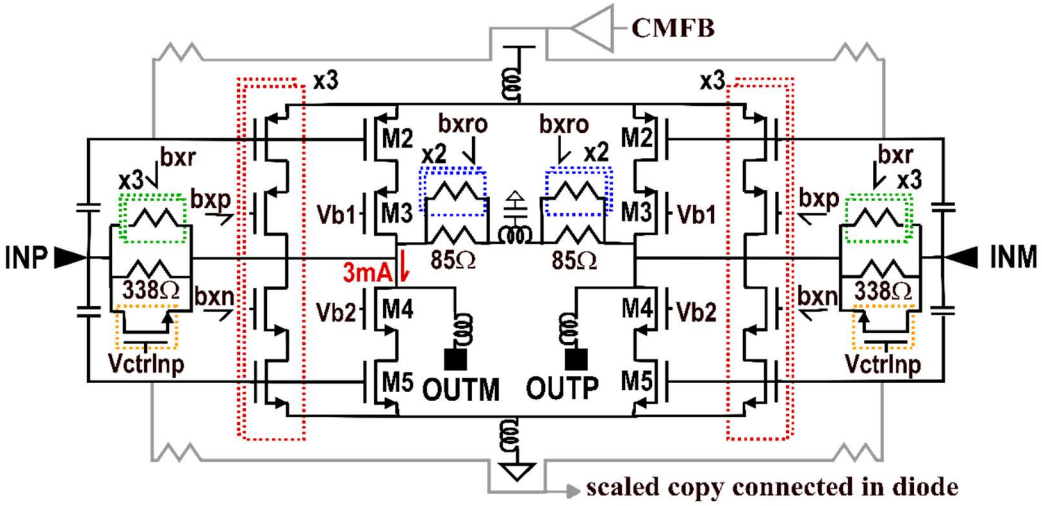


Fig. 3.3.3 Complete schematic of the LNTA shunt feedback

We opted to use a variable resistance in feedback to avoid distortion at low gain. The difficulty with this approach is the shape of the transfer function, which is strongly influenced by the ratio gm/R_f. We must consider the closed loop transfer function expression in order to understand how to avoid excessive shape variation.

$$H_{Rf}(s) = Closed\ Loop\ = \frac{RoRf\left(gm - \frac{1}{Rf}\right)}{1 + gmRo} * \frac{1}{1 + s\left[\frac{CoRo}{1 + gmRo} + \frac{Cin(Ro + Rf)}{1 + gmRo}\right] + \frac{s^2 CoCinRoRf}{1 + gmRo}} \tag{3.3.13}$$

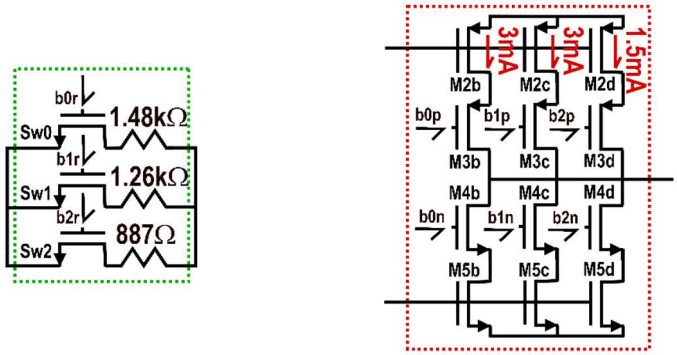
If we assume that $gmRo \gg 1$ and replaced R_f with R_f/N we get the following expression:

$$H_{Rf}(s)\ = \frac{\frac{Rf}{N}gm - 1}{gm} * \frac{1}{1 + s\left[\frac{Co}{gm} + \frac{Cin(Ro + Rf/N)}{gmRo}\right] + \frac{s^2 CoCinRf}{gmN}} \tag{3.3.14}$$

We can see that the gain is significantly influenced by N, and that the two complex poles are equal to $\sqrt{gmN/C_oC_{in}R_f}$. We decided to keep as constant as possible the poles' frequency by changing the gm with the feedback resistance. The closed loop transfer function under this assumption becomes:

(3.3.15) $$H_{Rf}(s) = \frac{\frac{R_f gm}{N^2} - 1}{gm/N} * \frac{1}{1 + s\left[\frac{NCo}{gm} + \frac{NC_{in}}{gm} + \frac{C_{in}R_f}{R_o gm}\right] + \frac{s^2 CoCinRf}{gm}}$$

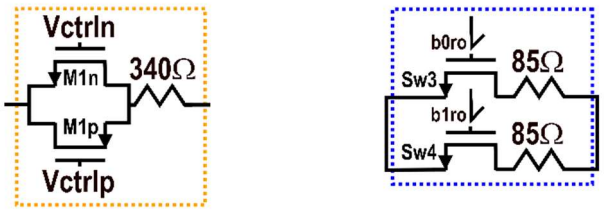
In this situation, the quality factor is $\sqrt{\frac{C_o C_{in} R_f}{gm}} \cdot \left(\frac{gmR_o}{C_{in}R_f + NR_o(C_{in}+C_o)}\right)$, and is thus dependent on the gain variation, with Q becoming lower for lower gain settings. To allow an additional degree of freedom to adjust the quality factor for different gain setting, we decided to provide the possibility to change also R_o. The variable gm is achieved by turning on and off the component of the inverter, while the variability of the resistance is implemented in part digitally (connecting and disconnecting resistors by switches) and in part analogically with two MOSFET in triode region. In parallel there is a fixed discrete resistance.



a) b)

Fig. 3.3.4 Implementation of the variable digitally resistance (a) and of the variable digitally gm (b)

The analog variable resistance is represented in Fig. 3.3.5. It is implemented by two parallel Pmos and Nmos regulated together in series with a resistance connected toward the output of the stage where there is the higher swing, in order to reduce the voltage swing across the MOSFETs and decrease their non-linear contributions.



a) b)

Fig. 3.3.5 Implementation of the variable analog feedback resistance (a) and of the variable output resistance (b)

Due to the pseudo-differential topology and the differential resistive load, the differential gain is lower than the common gain, to reduce it an inductive degeneration is used at the inverter's sources, and a capacitance is added to the shunt inductance's center tap to show the differential load in the common mode as well (Fig. 3.3.3).

3.3.2 Variable Gain Amplifiers (VGAs)

Different topologies for realizing VGAs can be found in the literature, however the most frequently used topology is the same utilized for the Common Gate TIA discussed above, the Gilbert Cell. A complementary PN version of the standard common source is proposed in this design, and its schematic is shown in Fig. 3.3.6.

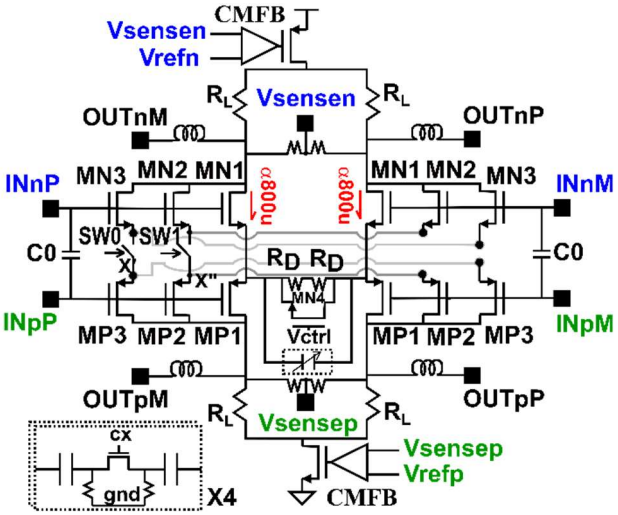


Fig. 3.3.6 Complete schematic of the VGA

The Pmos and the Nmos act as a shared source loaded both by a resistance R_L. Since the nodes X, X', and X'' are virtual ground for the signal, the VGA behaves as two parallel common sources, one in Pmos and the other in Nmos, with a shared DC current. A copy of the VGA provides the bias for this structure. The biasing of the stage is shown below; in particular, the Pmos and Nmos have a fixed sum of gate-source voltages taken from a scaled copy in diode connection highlighted by the dotted rectangle. It allows to set the current in the same way that a current mirror does. The regulators connected to the supply regulate the voltages at the drains of the two MOSFETs fixing to the same voltage of their gates (Fig. 3.3.7).

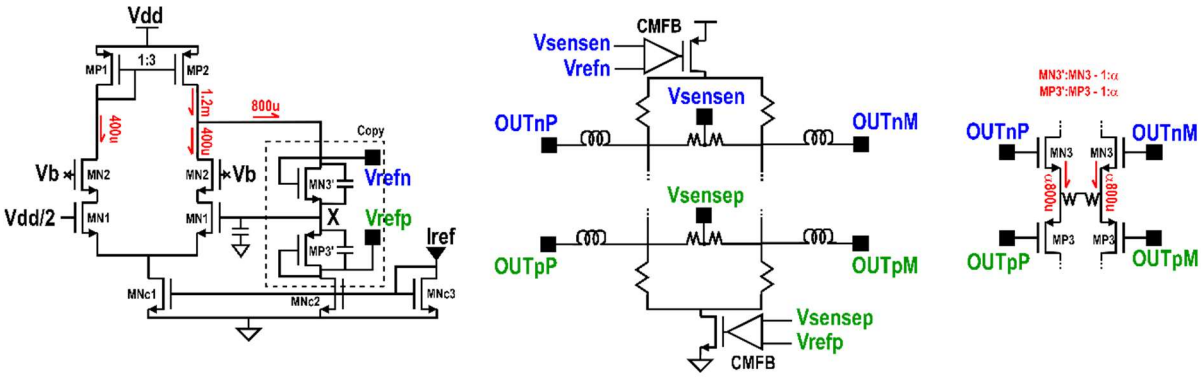


Fig. 3.3.7 Bias functionality of the VGAs

We can evaluate the stage as an amplifier with two input ports and two output ports, as shown in Fig. 3.3.8, to better understand how it operates.

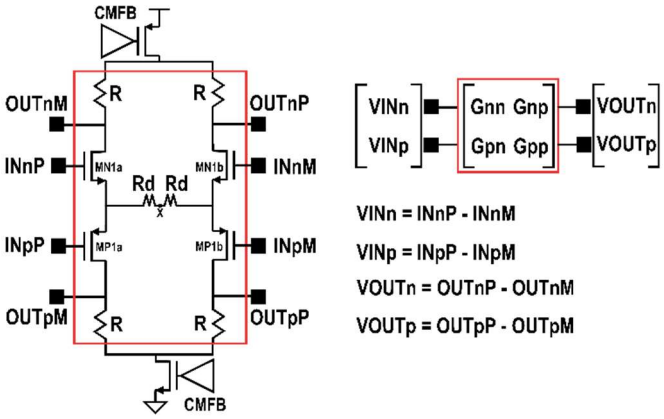


Fig. 3.3.8 Simplified VGA schematic and its matrix representation

Where the relative expressions are:

(3.3.16) $$Gnn = -\frac{Rgn(1 + R_d gp)}{1 + R_d(gn + gp)}$$

(3.3.17) $$Gnp = \frac{RR_d gngp}{1 + R_d(gn + gp)}$$

(3.3.18) $$Gpn = \frac{RR_d gngp}{1 + R_d(gn + gp)}$$

(3.3.19) $$Gpp = -\frac{Rgp(1 + R_d gn)}{1 + R_d(gn + gp)}$$

Take into account this expression, the output voltages result equal to:

(3.3.20) $$V_{on} = -\frac{Rgn(1 + R_d gp)}{1 + R_d(gn + gp)} V_{inn} + \frac{RR_d gngp}{1 + R_d(gn + gp)} V_{inp}$$

(3.3.21) $$V_{op} = -\frac{Rgp(1 + R_d gn)}{1 + R_d(gn + gp)} V_{inp} + \frac{RR_d gngp}{1 + R_d(gn + gp)} V_{inn}$$

Because the current that flows in the degeneration is provided by both the MOSFETs, the resulting gain is similar to a classic degraded common source, with a double degenerated Gloop (if $gn = gp$), that because through the degenerated resistance flows the current of both the branches. The resulting output voltages are equal to (3.3.22) if the two input voltages are different by a factor $\alpha = \frac{V_{iny}}{V_{inx}}$.

(3.3.22) $$\frac{V_{ox}}{V_{inx}} = -\frac{Rgx}{1 + R_d(gn + gp)} + \frac{(\alpha - 1)RR_d gngp}{1 + R_d(gn + gp)}$$

If the two input voltages are different, the Nmos and Pmos paths have different gains. A capacitance is utilized to equalize the two inputs voltages at high frequency. Because the gain provided by only the degenerative portion of the structure is about half the one of a N-only degenerated common source, two additional complementary common sources are added to the structure that can be turned on and off digitally by SW1 and SW0 to provide gain variability. Additionally, a variable degenerated resistance and capacitance are used to control the low frequency and high frequency gain of the transfer function. The advantages of tis topologies are the intrinsic DC compatibility between stages that permits to remove the source follower, and the reuse of the current for the Nmos and Pmos paths. While the main drawbacks are the managing of two input and two outputs and the headroom limitations.

3.3.3 Feed Forward Stages

Any Feed-Forward stage is implemented using the same circuit, as shown in Fig. 3.3.9.

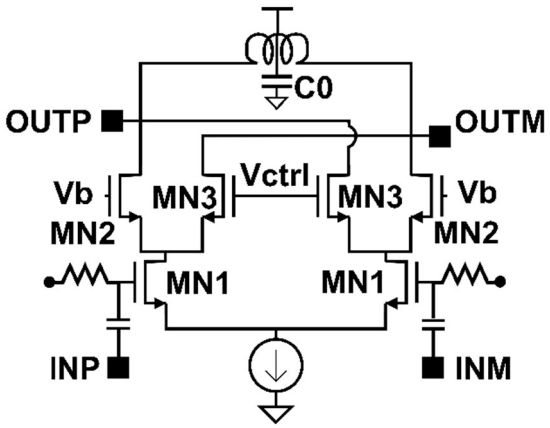


Fig. 3.3.9 Schematic of Feed-Forward stage

It is a N-only common source with a Gilbert Cell that provides gain variability and a resonance tank to realize the band pass transfer function. Any stage's gain is equal to the following expression:

$$\frac{V_{ox}}{V_{in}} = -gm \cdot \frac{sL}{1 + \frac{sL}{R} + s^2LC} \tag{3.3.23}$$

C and R, respectively represent the parasitic capacitances at the output node and the parasitic equivalent parallel resistance of the inductance. Due to the limited bandwidth of any Feed-Forward stage, a staggered approach is employed to expand the overall bandwidth of the passband Feed-Forward path, in which any stage is designed to resonate at different frequencies to have a wider overall bandwidth, as shown in Fig. 3.3.10.

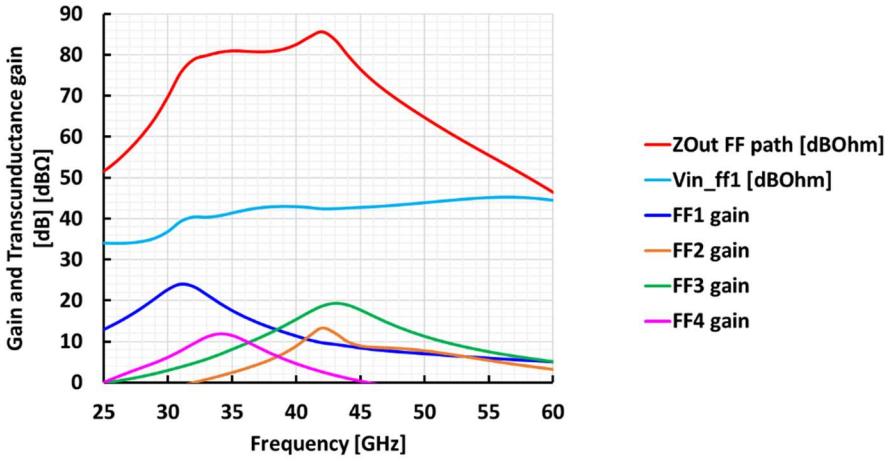
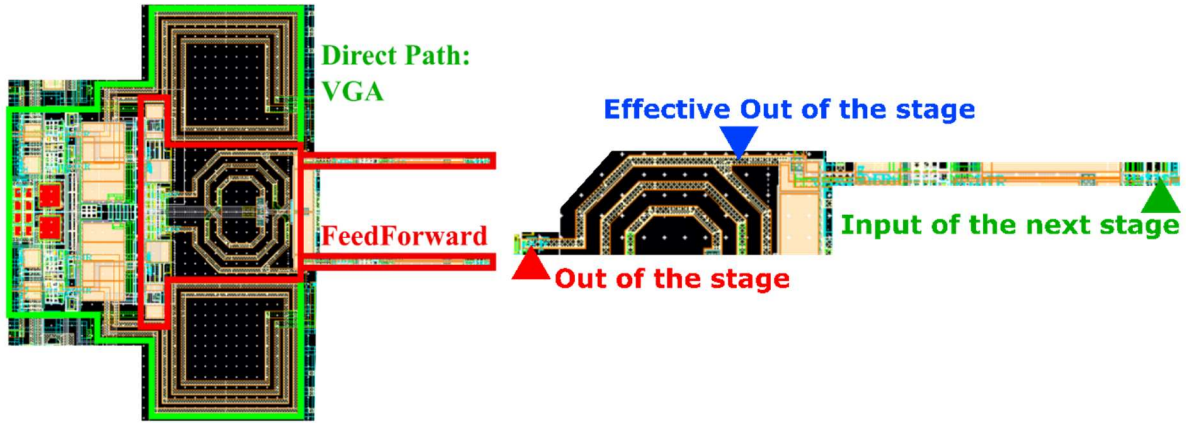


Fig. 3.3.10 Staggered structure of the Feed-Forward paths

In the image is also represents the input voltage taken from the FF1 in light blue, where we can see that the signal stays high also for high frequencies as it is necessary for this technique as already explained. The gain of any stage is varied in contemporary in order to keep as constant as possible the shape of the FFW Gain Path (red line). About the common mode gain, it is limited by the degeneration given by the current source, but its load includes not only the inductor, but also the stage's supply lines, for that reason a large capacitance is added at the shunt inductance's tap (C_0). Another challenge of this design is the layout, as we have to carry two paths that must keep a symmetry for the positive and negative signals. To preserve the symmetry, any stage of the direct path and of the Feed-Forward paths are concatenated into one other as shown in Fig. 3.3.11a.



a) b)

Fig. 3.3.11 a) Layout of a direct stage and a feed forward stage, b) detail of the Feed Forward inductance

As shown in the previous image any Feed-Forward stage is affected by a parasitic series inductance in order to be connected to the following stage. To minimize this effect, we decided to connect the following stage above the inductance's input node rather than directly after the stage's output, as shown in Fig. 3.3.11b.

3.3.4 Output Buffer and the Recombination of the two paths

Finally, we must recombine the two signals, which was performed in current mode with two parallel inverters as shown in Fig. 3.3.12.

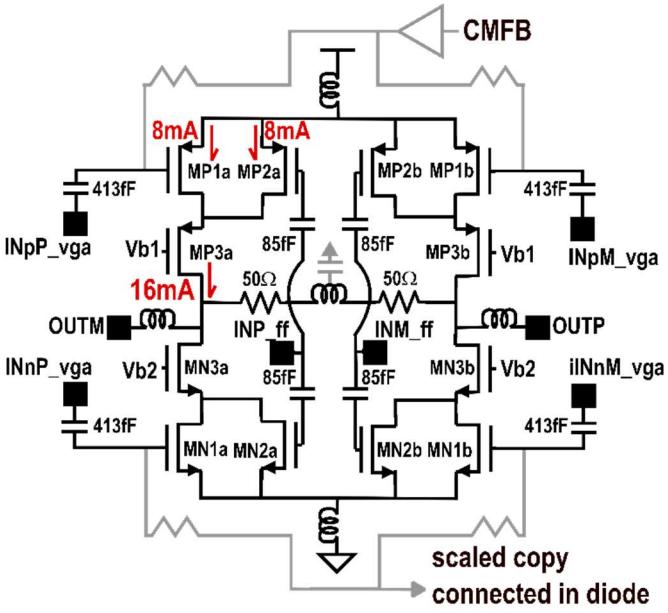


Fig. 3.3.12 Schematic of the output buffer and combiner

Because the feed forward must operate at a higher frequency, the AC coupling capacitance of the feed forward is lower than that of the direct path. A discrete differential resistance of 100 Ohm is utilized to achieve the matching, however it is limited by the triple resonance approach employed to improve the bandwidth. Also in this stage, as for the LNTA, the differential load provides limited gain in differential mode, but it is not seen in common mode, for that reason also in this stage we added degenerated inductors at the sources of the inverters and a capacitance at the center tap of the shunt inductance.

3.3.5 Measurements Results

The prototype of the Shunt-Feedback Feed-Forward Transimpedance Amplifier was realized in 28 nm CMOS TSMC HPC+ technology in a chip area 1mm x 1mm including the I/O pads. The photo of the chip is shown in Fig. 3.3.13, where we can identify the direct path and feed forward path stages

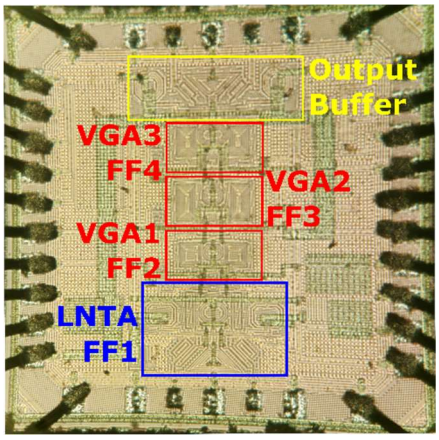


Fig. 3.3.13 Photo of the chip

The supply voltage is different for the Direct and the Feed-Forward paths, respectively it is 2.4 V and 1.8 V with current consumptions equivalent to 96 mA and 60 mA at maximum gain and 77 mA and 60 mA at minimum gain. A Keysight N5247B Network Analyzer with four ports and a maximum frequency of 67 GHz was used to do the S-parameter measurements. Figure 3.3.14 depicts the simulated and measured S/parameters, allowing us to determine the overall gain, the gain of only the direct path, and the gain of only the feed forward path.

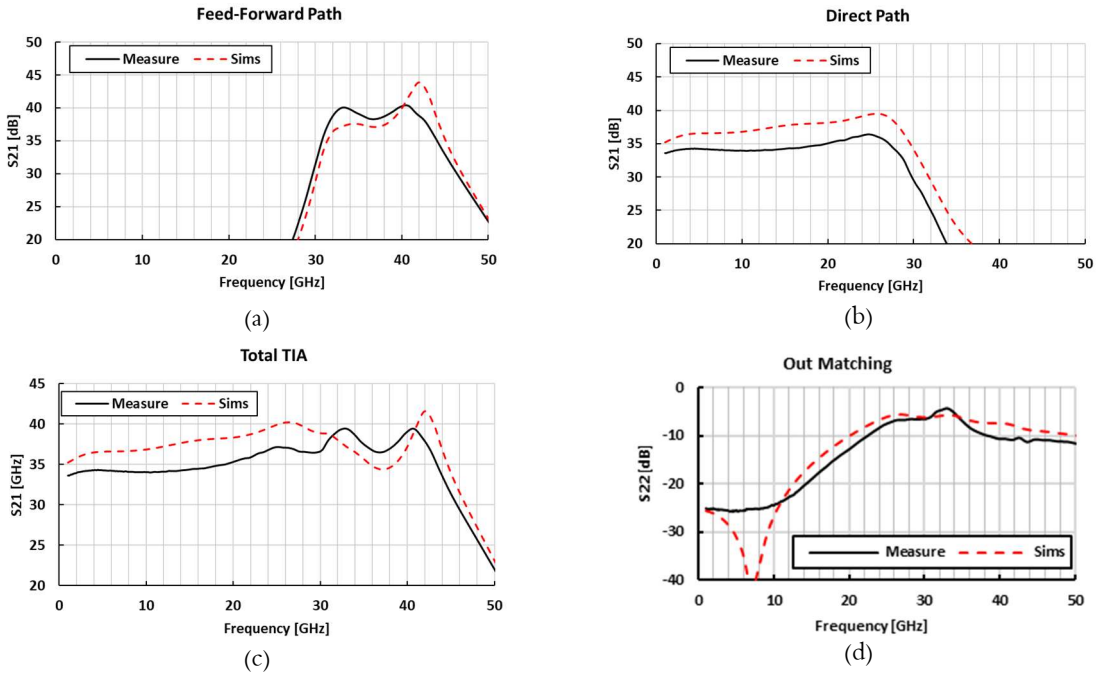
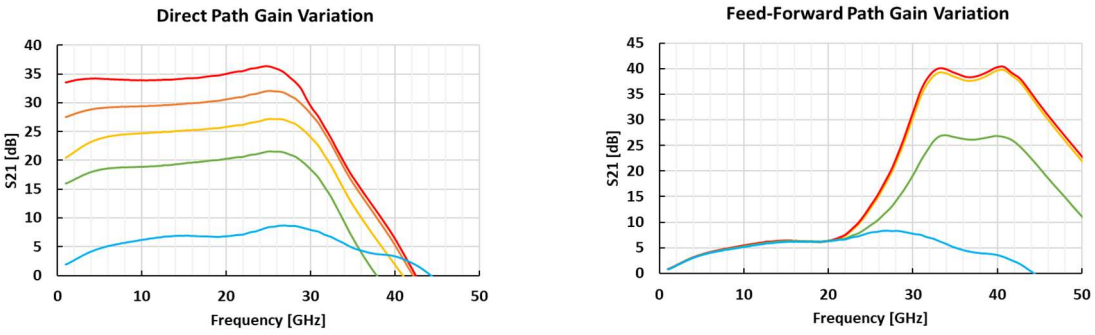


Fig. 3.3.14 S-parameters only the feed-forward path (a), of only the direct path (b), of the total oTIA (c), S22 simulated and measured (d)

The Feed-Forward path (a), with the exception of a lower quality factor at high frequency, matches the simulations well, however the Direct Path (b) has 1.6 dB less gain at 1 GHz and also has a lower gain around 15 GHz due to the direct path's lower quality factor of resonances. Due to these mismatches, when we compare the S21 of the total oTIA, we can see that the two parallel paths are not added as predicted.

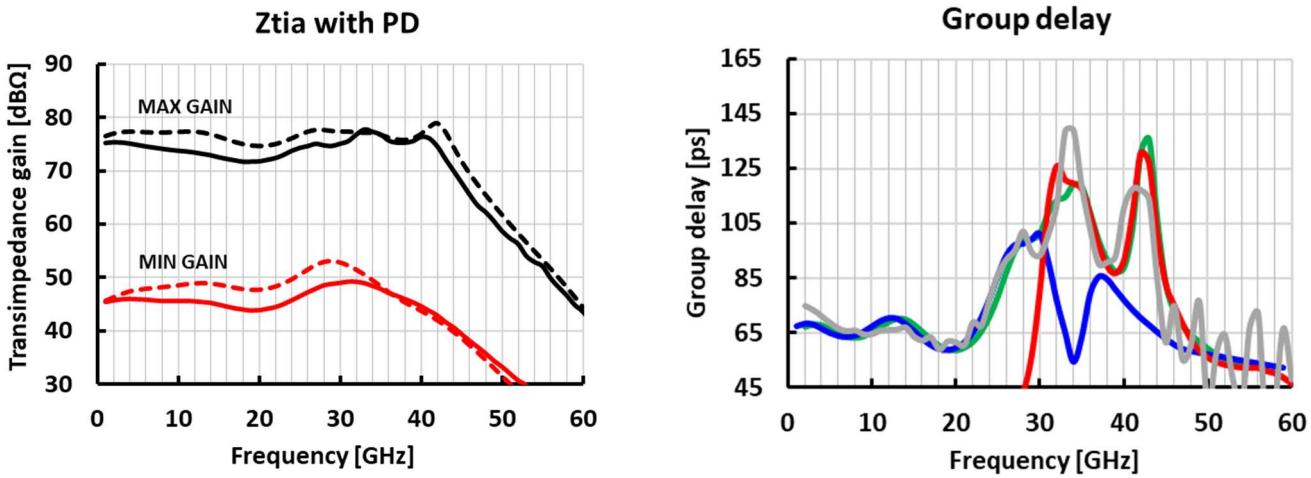
This can be partially solved by lowering the gain of the feed forward path to compensate the direct path's lower gain. This is possible thanks to the independent regulation of the two paths' gain, as shown in Fig. 3.3.15. The comparison of the S22 simulated and measured curves is also shown in Fig. 3.3.14 (d), where we can see a good fit between the curves, as we can expect due to the triple resonance network of the output buffer the matching around 30 GHz is compromised.



(a) (b)

Fig. 3.3.15 S21 of the TIA with the Feed-Forward path at minimum gain and changing the gain of the Direct Path (a), and with the Direct path at minimum gain and changing the gain of the Feed-Forward Path (b)

A standard model photodiode and a bondwire equivalent inductance of 150 pH were used to drive the TIA. In Fig. 3.3.16 (a), the simulated and measured Ztia using the photodiode model at maximum and minimum gain are shown, with a gain change of 30 dB from 76 dBΩ to 46 dBΩ. As we can see, dua to the additional losses of the direct path, the Ztia presents a drop in the gain in the middle of the bandwidth, a possible solution to recover it could be to implement an advanced version of the feed forward that can shift in frequency, and not only in gain.



(a) (b)

Fig. 3.3.16 (a) Transimpedance gain of the TIA with the Photodiode simulated (dotted line) and measured (solid line), (b) group delay of the TIA with the Photodiode measured (grey line), simulated (green line) and the contribution of only the Direct Path (blue line) and of the Feed-Forward Path (red line)

The group delay of the final electro optical transfer function is the results of the composition of the group delay of the two paths, in particular we have in Fig. 3.3.16 (b) the measured group delay in grey and the

post layout simulated group delay in green, but they are highlighted also the group delay of only the Direct Path in blue and of only the Feed-Forward Path in red, where we can see that the green is almost always above the highest group delay between the two. It suggests that a possible optimization of the group delay variation could be realized optimizing independently the group delay of the two path. The sum between the two paths doesn't produce any strong variation of the group delay probably because we are in the casuistic of one path always dominant to the other one.

The noise was measured taking the single ended output and it was reported to the bandwidth with the electro optical transfer function and it was integrated from 1 GHz to 43 GHz. The resulting input average noise current densities are 12 pA/Hz for the simulated noise, and 13.6 pA/Hz for the measured noise.

According to the simulation, the LNTA dominates the noise contributions by 47.7% and the first Feed Foward stage by 37%. The input current spectral densities are shown in Fig. 3.3.17, where we can see that the Feed-Forward dominates noise at high frequencies, whereas the LNTA dominates noise at low frequencies. As we already explained the noise of the direct paths, doesn't increase with frequency as f^2 as usually happens, but it is attenuated by the gain of the feed forward.

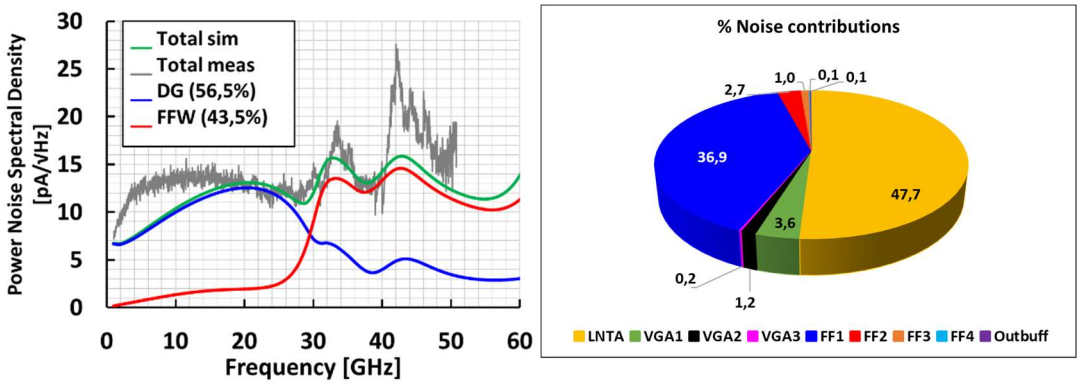


Fig. 3.3.17 Noise input current density and percentage of the contributions

The distortion is measured using a tone at 1 GHz keeping constant the output swing of 500 mV peak-to-peak differential, and it is realized for different gain levels. The THD is plotted in Fig. 3.3.18 as a function of the equivalent input current, which was calculated using the formula 3.3.26.

$$I_{NNPP} = 2\sqrt{\frac{2P_{IN,SE}}{50}}|1 - s_{11d}| \tag{3.3.26}$$

At minimal gain, the maximum Total Harmonic Distortion is equivalent to 1.83 %.

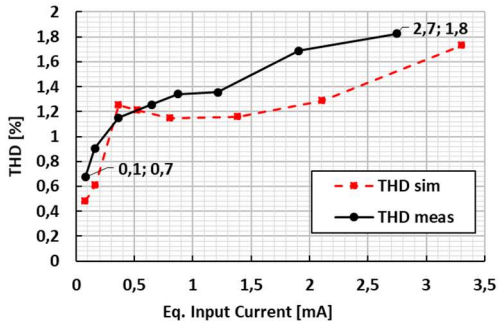


Fig. 3.3.18 Total Harmonic Distortion in function of the equivalent input peak to peak current (INNPP)

	This work	MTT '18 [3.14]	JSSC '19 [3.6]	JSSC '18 [3.7]	ESSCIRC '18 [3.9]	SSCL '20 [3.5]	RFIC '21 [3.12]	CICC '18 [3.10]
Tech	28nm CMOS	0.13um BiCMOS	0.13um BiCMOS	0.13um BiCMOS	28nm CMOS	28nm CMOS	22nm FinFet	16nm FinFet
Linear	YES	YES	YES	YES	YES	YES	YES	YES
Modulation	Coherent 16QAM	Coherent tested PAM4	Coherent 16QAM	PAM4	PAM4	Coherent	PAM4	PAM4
Voltage supply [V]	2.4 -1.8	3.3	3.3	3.3	2.5 – 1.2	2.4	0.8	1.8
SE/S2D/Diff.	Diff.	Diff.	Diff.	Diff.	S2D	Diff.	SE	S2D
ZT [dBΩ]	76	74	73	65	65	78	59.3	78
Dynamic Range [dB]	30	~24	43	15	Not specified	42	/	21
BW [GHz]	43	33	27	66	60	42	46.1	27
Input Ref. Noise [pA/√Hz]	13.6	12.2	20	7.6	19.3	18	12.6	16.7
THD [%]	1.83 0.5Vppd	4.2 0.5Vppd	2.2 0.5Vppd	<5 0.8Vppd	<5 0.3Vppd	1.8 0.5Vppd	<5 0,33mA (~304mVpp)	1.8 0.6Vppd
Power [mW]	232+108=340	218	313	150	107	319	11.2	60.8
FOM [8] [ZtxBW/Pdc]	798	759	385	782	997	1046	3789	6273

Table 3.3.1 TIA performance comparison and summary

Table 3.3.1 compares the Shunt-Feedback Feed Forward TIA with other linear TIA. Respect the previous design [3.5]. They have almost the same gain, THD and bandwidth, but thanks to the Feed Forward techniques we could reduce the high frequency noise contribution resulting in 44% less noise, at the cost of a higher power consumption that was increased by 6%.

Table 3.3 compares also the proposed TIA performance with previously published works. BiCMOS implementations for coherent modulations ([3.14]-[3.6]), and PAM4 designs have been reported ([3.7]-[3.12]). Looking at the Figure-Of-Merit (FOM), the proposed solution achieves a gain*bandwidth product normalized to power comparable to previous BiCMOS differential linear TIAs [3.14][3.6][3.7]. In terms of bandwidth/noise performance, this design comes closer to BiCMOS implementations, it is comparable to solutions in more advanced FinFET technologies [3.10][3.12] and improves compared with previous CMOS linear TIAs in the same technology node [3.5][3.9] The proposed TIA shows equal or better distortion compared with the other designs. To estimate the achievable Data Rate, a non-linear model, including noise and measured S-parameters, was implemented in Matlab. Simulations shows that this work can achieve 200Gb/s using 16QAM with 1.15x10-5 BER after group delay compensation. The presented TIA proposes a design technique to improve the tradeoff between bandwidth and noise at the

cost of higher power consumption. It can be used also in advanced technology as FinFET to achieve even better performances and to achieve higher levels of integration with the DSP in coherent optical receivers.

Chapter references

[3.1] http://www.tsmc.com/english/dedicatedFoundry/technology/28nm.htm

[3.2] Richard S. Muller, Theodore I. Kamins and Mansun Chan "Device Electronics for Integrated Circuits" third edition, John Wiley and Sons.

[3.3] B. Razavi, "Impact of Distributed Gate Resistance on the Performance of MOS Devices", IEEE Trans. Circuits and Systems- Part I, vol. 41, pp. 750-754, Nov. 1994.

[3.4] M. Manghisoni, "Gate Current Noise in Ultrathin Oxide MOSFETs and Its Impact on the Performance of Analog Front-End Circuits", IEEE Trans. Nucl. Sci. , vol. 55, no. 4, pp. 2399-2407, Aug. 2008.

[3.5] L. Aschei, N. Cordioli, P. Rossi, D. Montanari, R. Castello and D. Manstretta, "A 42-GHz TIA in 28-nm CMOS With Less Than 1.8% THD for Optical Coherent Receivers," in IEEE Solid-State Circuits Letters, vol. 3, pp. 238-241, 2020, doi: 10.1109/LSSC.2020.3012691.

[3.6] G. Ahmed, T. N. Huynh, C. Williams, Y. Wang, P. K. Hanumolu and A. Rylyakov, "34-GBd Linear Transimpedance Amplifier for 200-Gb/s DP-16-QAM Optical Coherent Receivers," in IEEE Journal of Solid-State Circuits, vol. 54, no. 3, pp. 834-844, March 2019.

[3.7] I. G. López, et al., "100 Gb/s Differential Linear TIAs With Less Than 10 pA/$\sqrt{}$Hz in 130-nm SiGe:C BiCMOS," in IEEE Journal of Solid-State Circuits, vol. 53, no. 2, pp. 458-469, Feb. 2018.

[3.8] A. Awny, et al., "A dual 64Gbaud 10kΩ 5% THD linear differential transimpedance amplifier with automatic gain control in 0.13μm BiCMOS technology for optical fiber coherent receivers," 2016 IEEE Int. Solid-State Circ. Conf. (ISSCC), San Francisco, CA, 2016, pp. 406-407.

[3.9] H. Li, G. Balamurugan, J. Jaussi and B. Casper, "A 112 Gb/s PAM4 Linear TIA with 0.96 pJ/bit Energy Efficiency in 28 nm CMOS," ESSCIRC 2018, Dresden, 2018, pp. 238-241

[3.10] K. R. Lakshmikumar, et al., "A Process and Temperature Insensitive CMOS Linear TIA for 100 Gb/s/ PAM-4 Optical Links," in IEEE J. of Solid-State Circuits, vol. 54, no. 11, pp. 3180-3190, Nov. 2019.

[3.11] G. Kalogerakis, et al., "A quad 25Gb/s 270mW TIA in 0.13μm BiCMOS with <0.15dB crosstalk penalty," 2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers, San Francisco, CA, 2013, pp. 116-117, doi: 10.1109/ISSCC.2013.6487661.

[3.12] S. Daneshgar et al., "A 128 Gb/s PAM4 Linear TIA with 12.6pA\√Hz Noise Density in 22nm FinFET CMOS," 2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), 2021, pp. 135-138, doi: 10.1109/RFIC51843.2021.9490496.

[3.13] Y. Liu et al., "A 50Gb/s PAM-4 Optical Receiver with Si-Photonic PD and Linear TIA in 40nm CMOS," 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020, pp. 1-4, doi: 10.1109/ISCAS45731.2020.9180413.

[3.14] A. Awny et al., "A Linear Differential Transimpedance Amplifier for 100-Gb/s Integrated Coherent Optical Fiber Receivers," in IEEE Transactions on Microwave Theory and Techniques, vol. 66, no. 2, pp. 973-986, Feb. 2018, doi: 10.1109/TMTT.2017.2752170.

[3.15] Eduard Sackinger "Broadband Circuits for Optical Fiber Communication" Copyright 0 2005 by John Wiley & Sons, Inc.

CHAPTER 4

This chapter presents a secondary work realized to analyze possible alternative way to realize wideband Variable Gain Amplifiers with low linearity [4.1].

4.1 Closed Loop Variable Gain Amplifier

We have already analyzed the importance to have high linearity for coherent communications, and that ideally we want a transfer function that doesn't change with the gain settings keeping constant the bandwidth. The stages that have a key role in the variability of the gain are the VGAs that has also a fundamental role for the distortion at medium and low gain settings. Usually the most used topology is the Gilbert Cell or the Current Steering. How schematic is also reported in Fig. 4.1.1 [4.1].

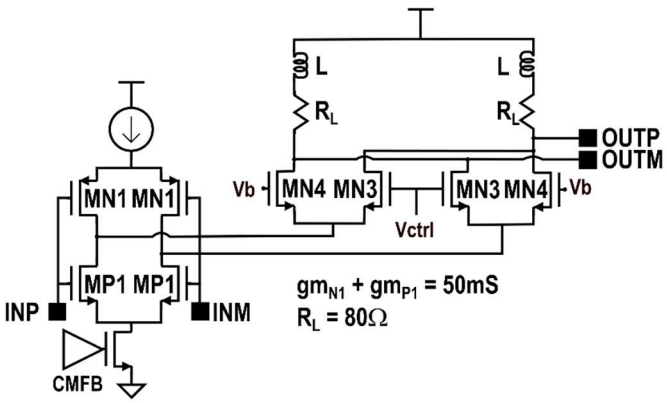


Fig. 4.1.1 Gilbert Cell VGA schematic [4.1]

This topology is characterized by a poor linearity due to the nonlinear transconductance and current divider. Another drawback of the Gilbert Cell is the Miller multiplication effect in the output conductance of the cascode that produces different impedances for different gains, it produces a variation of the transfer function shape for different gain conditions [4.1]. The block scheme of the proposed VGA is represented in Fig. 1.1.2 [4.1].

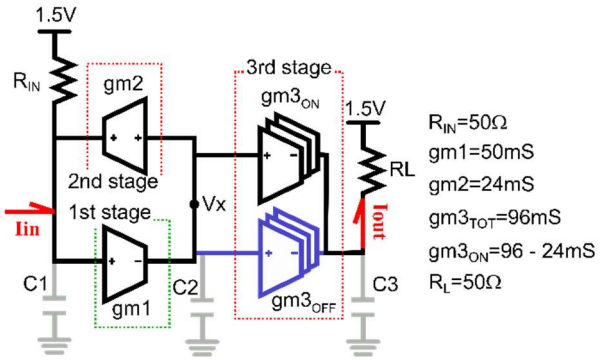


Fig. 4.1.2 Scheme of Closed Loop VGA

The VGA operates in current mode: the input signal is absorbed by the second stage in feedback and produce a voltage signal at the node x. Meanwhile the first stage amplify the input voltage and allows to produce the voltage swing Vx . At the end the third stage takes the voltage signal and produce an output current that is the input signal amplified by theratio between the transconductances of the third and second stage. The current mode amplification allows to reduce distortion that is also minimized by the closed loop principle and the gain of the first stage that minimize the swing at the input node. The resistance R_{IN} represents the output resistance of the previous stage and influences the quality factor of the closed-loop poles [4.1]. About the implementation of the third stage is divided in sixteen identical slices that can be independently turned on and off, with the same topology also the second stage is realized with a scaled down version. The schematic of any slice and of the second stage is represented in Fig. 4.1.3. It is a cascoded inverter, where for the third stage the cascoded MOSFETs can be turned off [4.1].

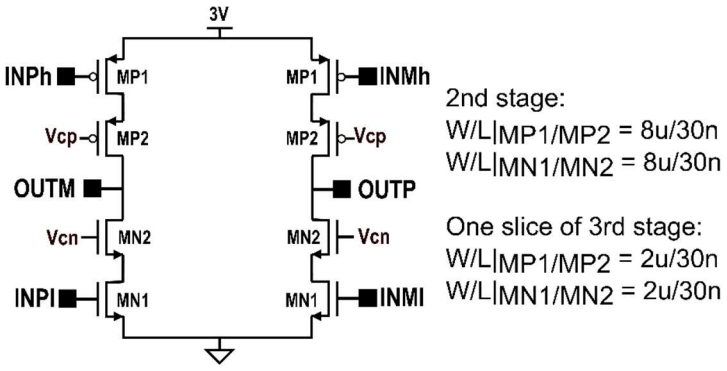


Fig. 4.1.3 schematic of the second stage and of a single slice of the third stage

Analyzing the closed loop transfer function, it is characterized by two complex poles produced by the feedback [4.1]. In particular, it is equal to:

$$\frac{I_{out}}{I_{in}} = \frac{gm3_{ON}}{gm2} \cdot \frac{1}{1 + \frac{s}{\omega_0 Q} + \frac{s^2}{\omega_0^2}} \cdot \frac{1}{1 + sC_3R_L} \tag{4.1.1}$$

Where

(4.1.2) $$\omega_0 = \sqrt{\frac{gm1gm2}{C_1C_2}}$$

(4.1.3) $$Q = \frac{R_{in}\sqrt{gm1gm2}}{\sqrt{\frac{C_2}{C_1}}}$$

We want that the transfer function shape doesn't change with the gain, it means that we have to consider the equations (4.1.1-3) substituting the parasitic capacitances with the equivalent expressions in function of the cut-off frequency of the MOSFETs and taking into account the turned ON and OFF slices od the third stage where $gm3_{TOT} = gm3_{ON} + gm3_{OFF}$ is constant [4.1]. In detail we have that for any transistor we define an input and an output parasitic capacitance equal to:

(4.1.4) $$C_{in} = \frac{gm}{\omega_T}$$

(4.1.5) $$C_{out} = \alpha \cdot \frac{gm}{\omega_T}$$

And the relation between the cut off frequencies for a turned off and on MOSFET is:

(4.1.6) $$\omega_{T_{ON}} = \beta \cdot \omega_{T_{OFF}}$$

Where α and β are two technology parameters. Under these assumptions (4.1.2-3) become

(4.1.7) $$\omega_0 = \sqrt{\frac{gm1gm2}{w_T((gm1 + \alpha gm2)\left(\alpha gm1 + gm2 + gm3_{ON} + \frac{gm3_{OFF}}{\beta}\right))}}$$

(4.1.8) $$Q = R_{in}gm1 \cdot \frac{\sqrt{1 + \frac{\alpha gm2}{gm1}}}{\sqrt{1 + \frac{\alpha gm1}{gm2} + \frac{gm3_{ON}}{gm2} + \frac{1}{\beta} \cdot \frac{gm3_{OFF}}{gm2}}}$$

Observing the equations, we can conclude that the variation of Q and ω_0 depend to the value of β that in this technology is equal to 1.1, that allows to guarantee a sufficiently constant value of Q and ω_0. About the shape of the transfer function, another important parameter to take into account in these applications is the peaking that depend on Q. It is strongly affected by the input resistance R_{IN}, it means that changing its value, the shape of the transfer function can be design according to the necessity, while ω_0 is independent to R_{IN} [4.1].

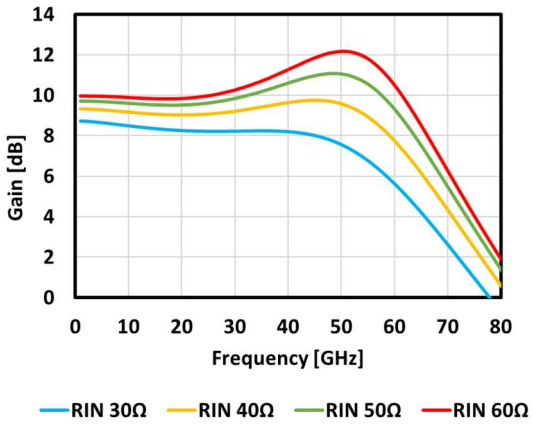


Fig. 4.1.4 Transfer function in for different values of R_{IN}

Fig. 4.1.4 shows the transfer function for different values of R_{IN} (from 30 Ω to 60 Ω), where we can observe that the peaking changes from 0 to 2.2 dB. In fact, if we consider the finite output resistances of the transconductors and we call them respectively r01, r02 and r03 for the three stages, the DC gain at the node VX becomes equal to (4.1.9) [4.1].

$$\frac{V_X}{I_{in}}|_{DC} = -\frac{gm1r_{01}R_{IN}^*}{1+gm1r_{01}gm2R_{IN}^*} \tag{4.1.9}$$

Where $R_{IN}{}^*$ is the parallel between R_{IN} and r02 [4.1].

In coherent optical TIAs, the bandwidth is importance as the distortion, in a closed loop circuit, it is strongly dependent to the Unity Gain Bandwidth (UGBW) of the Loop Gain. In the proposed VGA the Loop Gain (GL) is given by:

$$GL = \frac{gm_1r_{01}gm_2r_{02}R_{IN}}{r_{02}+R_{IN}} \cdot \frac{1}{1+sr_{01}C_2} \cdot \frac{1}{1+sC_1\frac{r_{02}R_{IN}}{r_{02}+R_{IN}}} \tag{4.1.10}$$

In order to improve the Unity Gain Bandwidth preserving the phase margin, we introduced two complex zeros in the first stage by a feed forward path. Fig. 4.1.5 presents the circuit implementation: it consists of two stages (n-type and p-type) in parallel with two outputs with same gain to accommodate DC requirements of the next stage. Each stage is presents a differential pair (MN1-2 and MP1-2) with an active load (MP3 and MN3) that is also used as a feed-forward path. The zeros are realized using the capacitors C0 for feed-forwarding the input signal. The capacitance resonates with the inductance L, thereby producing complex-conjugated zeros.

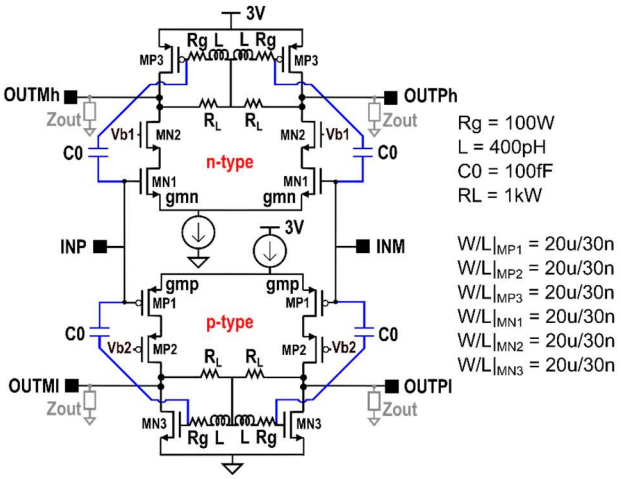
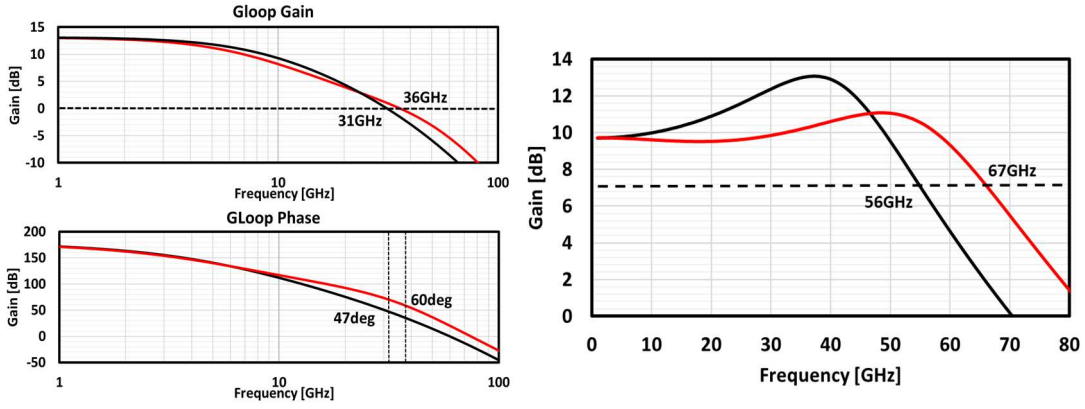


Fig. 4.1.5 Schematic implementation of the first stage [4.1]

The gain of the n-type stage is equa to (4.1.11) [4.1].

$$\frac{V_X}{V_{IN}} = A \cdot \frac{s^2 L(2C0 + C_{TOTgp}) + sR_g(2C0 + C_{TOTgp}) + 1}{s^2 L(C0 + C_{TOTgp}) + sR_g(C0 + C_{TOTgp}) + 1} \tag{4.1.11}$$

Where $A = gm_p Z_{out}//R_L = gm_n Z_{out}//R_L$ and C_{TOTgp} is the total capacitance at the gate of the PMOS load. It can be observed that we are introducing two complex zeros with also two complex poles, that usually are at high frequency. Designing properly the values of L and R_g, it is possible improve the bandwidth while improving the phase margin. If we compare the Open Loop Gain between the Closed Loop VGA with and without the zeros we can observe, from Fig. 4.1.6 (a), that the bandwidth is improved by a factor 1.2 and improving the phase margin from 47 degrees to 62 degrees. That bandwidth enhancement is also evident from the Closed Loop transfer function in Fig. 4.1.6 (b) [4.1].

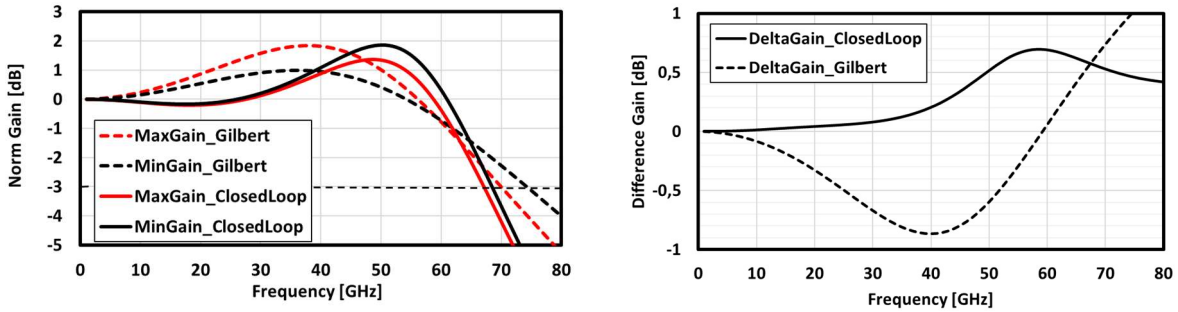


a) b)

Fig. 4.1.6 Open Loop Gain (a) and the transfer function with and without the complex zeros (b) [4.1]

4.2 Closed Loop VGA vs Gilbert Cell VGA

To evaluate the advantages of the Closed Loop VGA we have compared this design to the Gilbert VGA used in the Common Gate Cross Drain TIA design. Due to the difference between the voltage and the current amplification, we decided to drive both with a 50 Ω port, in particular in the Closed Loop solution the input 50 Ω resistance corresponds to R_{IN}. The first thing to compare is the shape of the transfer function for minimum and maximum gain, to make it more evident, in Fig. 4.2.1 are reported both the normalized transfer function for the two topologies and the difference between the maximum and minimum gain where is evident that in the Gilbert Cell VGA the difference between the two extreme conditions is higher than in the Closed Loop VGA [4.1].



(a) (b)

Fig. 4.2.1 Normalized transfer function at minimum and maximum gain (a), difference between the minimum and maximum gain (b)

The other performance to compare is the Total Harmonic Distortion (THD) that eas simulated with a tone at 1 GHz keeping constant the output swing at 500 mVppd. The distortion in the Closed Loop VGA grows linearly with the gain variation with a maximum value equal to 0.166 %, while the Total Harmonic Distortion in the Gilbert Amplifier is non-monotonic and has a peak value of 1%. In point A, the distortion grows due to the subthreshold biasing of MN3, while in point B the linearity is improved by linearization effect in the composition of the currents coming from MN3 and MN4 [4.1].

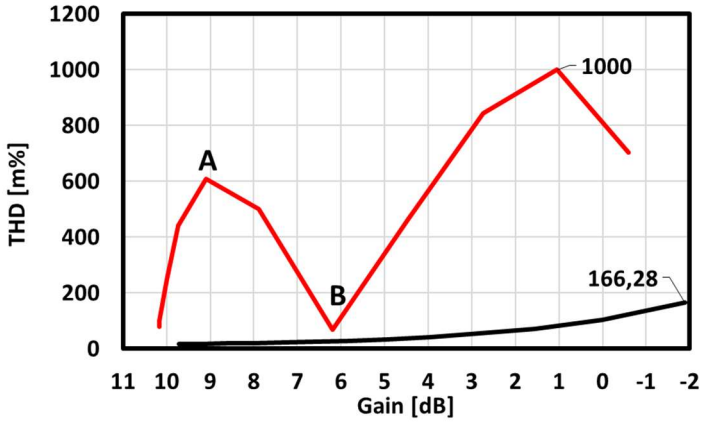


Fig. 4.2.2 THD of the Gilbert Cell VGA (red line) and of the Closed Loop VGA (black line)

The perfoemances are summarized in Table 4.1 where we can observe that in corrispondece of almost the same maximum gain and gain variation the THD of the Gilbert Cell is worse of a factor 6 respect the Close Loop VGA to the cost of a higher power consumption [4.1].

Table 4.1 Comparison with the Gilbert Cell VGA

\	**Tech.**	**Max gain**	**Min gain**	**Max THD @1 GHz 500 mVppd**	**Bandwidth**	**Voltage Supply**	**Power**	**Average Output Noise (Max Gain)**
Closed Loop VGA	28 nm hpc	9.7 dB	-1.9 dB	0.166 %	67 GHz	3 V	126 mW – 84 mW	6 nV/√Hz
Gilbert VGA	28 nm hpc	10 dB	-0.7 dB	1 %	70 GHz	2.4 V	51 mW	5 nV/√Hz

Chapter references

[4.1] In pubblication phase: L. Gnaccarini, L. Aschei, R. Castello and D. Manstretta, "A Low-Distortion 67 GHz Closed Loop Variable Gain Amplifier for Coherent-Opto Applications," in IEEE International Conference on Electronics, Circuits and Systems, 2021.
[4.2] L. Aschei, N. Cordioli, P. Rossi, D. Montanari, R. Castello and D. Manstretta, "A 42-GHz TIA in 28-nm CMOS With Less Than 1.8% THD for Optical Coherent Receivers," in IEEE Solid-State Circuits Letters, vol. 3, pp. 238-241, 2020, doi: 10.1109/LSSC.2020.3012691.

CONCLUSIONS AND FUTURE WORKS

The high data rates used to transmit vast amounts of data in wired data links is forcing more and more transmission systems to adopt optical links for their low losses at high frequency and coherent modulations for their high spectral efficiency. This solution has the consequence to dramatically change the transceivers performance requirements in terms of bandwidth, noise, distortion and power consumption with respect to NRZ or PAM4 modulations. The optical coherent devices are used mainly for long reach links, where the small number of devices relaxes the limit in power consumption. At the same time, due to the very dense constellations, the noise and linearity performances have to be good enough to meet the BER requirements. In these systems the Digital Signal Processing is fundamental to decode the symbol and it can be used to compensate, not only the optical impairments, but also some limitations of the electrical devices, such as group delay variation and even some types of distortions. In this scenario, usually the BiCMOS technology is used for its high cutoff frequency and low noise. nonetheless, the possibility to realize transceivers in CMOS technology is interesting in order to allow the integrability with the DSP and reduce system costs. My research was part of research project supported by Huawei Italia, aiming at the investigation of CMOS solutions for coherent optical communications. This thesis is focused on the design of two possible Transimpedance Amplifiers for coherent optical receivers: the first realized with a Common Gate topology and the second with the Shunt Feedback topology in 28nm CMOS technology.

The Common Gate TIA presents a final bandwidth of 42GHz with an input equivalent noise of 18pA/Hz and a maximum THD equivalent to 1.77% for 500mVppd at the output, equivalent to a SNR of 16dB. Even though an experimental verification of optical demodulation could not be performed in the lab due to equipment limitations, a system model was implemented in Matlab, incorporating the main TIA non-idealities (such as noise, bandwidth and distortion) as well as realistic photodiode model, demonstrating that the implemented TIA could sustain a 16QAM electro/optical communication with 40Gbaud sample-rate, corresponding to a data rate of 160Gb/s. Coherent links typically use dual polarization, which would bring the supported data rate to 320Gb/s for a single wavelength. Comparing the performances of the implemented CMOS TIA to the literature, we find that its performance is comparable to BiCMOS solutions used for the same application. The main limitation of this design was the moderate input noise. This was due to the capacitance of the photodiode that, loading the input of the TIA, limited the minimum gm of the common gate and consequently the -3dB frequency of the first stage. To compensate it, I used peaking in the next stages, thus amplifying also the noise. A possible improvement in the bandwidth/noise tradeoff is proposed in the second TIA.

The second design is the Shunt Feedback Feed-Forward TIA. It is based on two parallel paths: a low-frequency amplification path, that is designed for low noise and is characterized by a limited bandwidth, and a feed-forward path, with a band-pass transfer function, that amplifies the signal at high frequency and extends the bandwidth without amplifying the noise. This solution has the cost of higher power consumption, but it minimizes the high frequency noise contribution of the low frequency path, that is usually the dominant contribution. The final measurements present a bandwidth of 43GHz with and input equivalent noise of 13.6pA/Hz and a maximum THD of 1.83% for 500mVppd at the output. Also in this case, I tested the TIA with the Matlab script to verify the modulations that it can sustain. After a group delay compensation, we are able to use the Shunt Feedback Feed-Forward TIA to perform a 16QAM modulation with 200Gb/s data rate with a BER equal to 1.15x10-5. The necessity to compensate the group delay is due to the phase variation in the feed forward transfer function that could be optimized in a future version of the TIA. Another possible idea for future improvements of this design is the possibility to shift in frequency the feed forward transfer function to compensate additional losses in the low frequency path.

In both designs I used the advantages of the CMOS in term of linearity and complementarity, that helps to reduce the current consumptions. At the same time the limited fT and gmOverId of the technology limited the bandwidth of the proposed solutions. For this reason, possible improvements of the design performances could be achieved with more advanced technologies, characterized by higher cut off frequency.

APPENDIX

```
%>>>>>>>>>>>>>>>>>>> Matlab code for M ary-QAM modulation >>>>>>>>>>>>>>>>>>>%

clc;
clear all;
close all;

%PARAMETERS
%set all parameters of the amplifier (TIA)
amp.filename=input('->enter the filename of the csv file (transfer function from cadence) - save in your
folder - ex. prova.csv: ');
amp.filerow=input('->enter the row num. from which read (index start from 1) - cadence defoult 2: ');
amp.filecolumn=input('->enter the comun num. from which read (index start from 1) - cadence defoult 1: ');
amp.filemagcolumn=input('->enter the comun num. refered to the magnitude (index start from 1) - cadence
defoult 2: ');
amp.filemagdBflag=input('->enter 1 if the magnitude is in dB, otherwise 0 :');
amp.filephdegflag=input('->enter 1 if the phase is in deg, otherwise 0 :');
amp.mxtf=csvread(amp.filename,(amp.filerow-1),(amp.filecolumn-1));
if amp.filemagdBflag
    amp.s211GdB=amp.mxtf(1,amp.filemagcolumn);   %s21 db at 1GHz gain
    amp.s211G=10^(amp.s211GdB/20);  %s21 lin at 1GHz gain
else
    amp.s211G=amp.mxtf(1,amp.filemagcolumn); %s21 lin at 1GHz gain
    amp.s211GdB=20*log10(amp.s211G); %s21 db at 1GHz gain
end
amp.thdp=input('->enter the value of THD at the output (in percentuale %) : '); %thd %
amp.thd=amp.thdp/100;   %thd
amp.snrdB=input('->enter the value of snr (dB) : ');    %snr db
amp.snr=10^(amp.snrdB/10);    %snr lin
amp.voutppd=input('->enter the output pp diff at 1GHz (V) : ');    %vout ppdiff
amp.vinppd=amp.voutppd/amp.s211G; %vin ppdiff
amp.PoutdBm=10*log10(amp.voutppd*amp.voutppd/(4*2*100*0.001));
amp.vnoutppd=amp.voutppd/sqrt(amp.snr);
amp.PindBm=10*log10(amp.vinppd*amp.vinppd/(4*2*100*0.001));
amp.PNoutdBm=amp.PoutdBm-amp.snrdB;
amp.PNindBm=amp.PNoutdBm-amp.s211GdB;
amp.vninppd=amp.vinppd/sqrt(amp.snr);
amp.alpha1=amp.s211G-3*amp.thd*amp.s211G;   %distortsione y = alpha1*xcos(w) + alpha3*(cos(w))^3
amp.alpha3=4*amp.thd*amp.s211G/(amp.vinppd^2);  %distortsione y = a*xcos(w) + b*cos(3w)
disp('Amplifier parameters loaded');

%set all parameters of the data
sys.M=input('->enter the value of M array for QAM modulation : ');
sys.nsimbol=input('->enter the number of sended simbol : ');
sys.nbit=sys.nsimbol*log2(sys.M);
sys.br=input('->enter the bit rate : ');
sys.bp=1/sys.br;
sys.sr=sys.br/log2(sys.M);
sys.sp=1/sys.sr;
fprintf('Symbol rate equal to: %d\n', sys.sr);
fprintf('Sended bit equal to: %d\n', sys.nbit);
disp('System parameters loaded');


%-%-%-%-%-%-%-CHECKS%-%-%-%-%-%-%-
%check that M is a power of 2
ds=ceil(log2(sys.M));
dif=ds-log2(sys.M);
if(dif~=0)
   error('the value of M is only acceptable if log2(M)is an integer');
end
%FLAG for plot or not graphs
if sys.nsimbol>20
    param.plot = false;
else
    param.plot = true;
end

%-%-%-%-%-%-%-CREATE DATA%-%-%-%-%-%-%-
%BIT STREAM
%bit stream
databit=[];
if (sys.nbit < 1000000)
    databit=round(rand(sys.nbit,1)); % information generation as binary form
else
    sys.nbit = round(sys.nbit/10);
    for i=1:1:10
        databit=[databit; round(rand(sys.nbit,1))];
    end
end
sys.nbit = length(databit);

disp('Starting symbol generation...');
```

```
%SYMBOL STREAM
datasym=qammod(databit,sys.M, 'gray', 'InputType', 'bit');

%REFERENCE CONSTELLATION
x1=[0:sys.M-1];
p=qammod(x1,sys.M);

%SCALE THE SYMBOL STREAM RESPECT THE INPUT SIGNAL OF THE TIA
MAX_REAL = max(real(p));
MAX_IMG = max(imag(p));
MAX_VAL = max(MAX_REAL, MAX_IMG);
pnorm=p/MAX_VAL;    %normalized reference constellation
datasym = datasym/MAX_VAL;
datasym = datasym*amp.vinppd*0.5;
pin=pnorm*amp.vinppd*0.5;   %Vin reference constellation
Rdatasym=real(datasym);
Idatasym=imag(datasym);
% constDiagram = comm.ConstellationDiagram('ReferenceConstellation', pin, 'ShowGrid', true, 'XLimits', [-
amp.vinppd*2/3 amp.vinppd*2/3], 'YLimits', [-amp.vinppd*2/3 amp.vinppd*2/3]);
% v=[];   %from row to column
% for i=1:length(datasym)
%     v=[v;datasym(i)];
% end
% constDiagram(v);

%CREATE SIGNAL IN TIME FOR REAL AND IMAGINARY COMPONENTS
param.uppersample=10;    %uppersample del coseno portante fIF
t=0:sys.sp/(param.uppersample):sys.nsimbol*sys.sp-sys.sp/(param.uppersample);   %time vector totale dei
simboli
%real component
Rsignal=[];
for i = 1:sys.nsimbol
    Rsignal(param.uppersample*(i-1)+1:param.uppersample*i)=Rdatasym(i); %vettore segnale reale
end
if param.plot
    figure(1)
    subplot(2,1,1)
    plot(t, Rsignal); hold on;
end
%imaginary component
Isignal=[];
for i = 1:sys.nsimbol
    Isignal(param.uppersample*(i-1)+1:param.uppersample*i)=Idatasym(i); %vettore segnale reale
end
if param.plot
    figure(2)
    subplot(2,1,1)
    plot(t, Isignal); hold on;
end
disp('Ended symbol generation...');

disp('Starting model of the amplifier...');
%-%-%-%-%-%-%-RF AMPLIFIER%-%-%-%-%-%-%-
%ADDITIONAL WHITE NOISE Noisy_signal=White_gaussian_noise+Signal
Rnoise=wgn(length(Rsignal),1,amp.PNindBm, 100, 'dbm');
Inoise=wgn(length(Isignal),1,amp.PNindBm, 100, 'dbm');
Rsignal_n=Rsignal+Rnoise';
Isignal_n=Isignal+Inoise';
if param.plot
    figure(1)
    subplot(2,1,1)
    plot(t, Rsignal_n); hold on;
    title('Real signal')
    legend('Original signal','Noise', 'Noise Filtered')
    figure(2)
    subplot(2,1,1)
    plot(t, Isignal_n); hold on;
end
disp('Added noise...');

%FILTERING  SIGNAL_FILTERED=REAL(FFT'(TF*FFT(SIGNAL)))
%%%%%%%%%%definition of the time for the fft
param.slen=length(Isignal_n);
param.dt=sys.sp/(param.uppersample);
f_vect=-param.slen/2:param.slen/2-1;
f_vect=(f_vect/param.slen)/param.dt;
f_vect=fftshift(f_vect); %definition of the frequency vector for the fft (Hz)
%read the data file and storage in vectors
TFmag=amp.mxtf(:,amp.filemagcolumn);
if amp.filemagcolumn==2
    TFph=amp.mxtf(:,3); %phase
else
    TFph=amp.mxtf(:,2); %amplitude
end
if amp.filemagdBflag
    TFmag=10.^(TFmag/20);   %amplitude in linear
end
if amp.filephdegflag
    TFph=TFph*pi/180;   %phase in radians
```

```
end
TFfreq=amp.mxtf(:,1);   %frequency
TFmag=TFmag/TFmag(1);   %normalized gain, the gain will be added in the distortion equation
%%%%%%%%%%interpolation of the positive frequencies of magnitude and phase
f_vectpos=f_vect(1: param.slen/2);  %positive vector of the frequencies
%magnitude
[TFmagtemp, index]=unique(TFmag);   %delete repeted value of mag for interpolation error
TFmagInt=interp1(TFfreq(index), TFmagtemp,f_vectpos, 'linear','extrap');    %mag interpolation
%phase
[TFphtemp, index]=unique(TFph);   %delete repeted value of ph for interpolation error
TFphInt=interp1(TFfreq(index), TFphtemp,f_vectpos, 'linear','extrap');  %ph interpolation
%%%%%%%%%%complex transfer function
TF=TFmagInt.*exp(1j*TFphInt);
%%%%%%%%%%plot transfer function
figure(3)
subplot(2,1,1)
plot( TFfreq,mag2db(TFmag), 'r-',  f_vectpos,mag2db(TFmagInt), 'b-')
title('Magnitude')
legend('Starting transfer function','Interpolated transfer function')
xlabel('Frequency (Hz)')
ylabel('Magnitude (dB)')
figure(3)
subplot(2,1,2)
plot( TFfreq,TFph, 'r-',  f_vectpos,TFphInt, 'b-')
title('Phase')
legend('Starting transfer function','Interpolated transfer function')
xlabel('Frequency (Hz)')
ylabel('Phase (rad)')
%%%%%%%%%%create hermitian transfer function
TFphtemp=-TFphInt;  %conjugate phase
TFmagtemp=TFmagInt;
f_vectneg1=-f_vectpos;  %relative frequency vector of the conjugated values
f_vectneg=f_vect( param.slen/2+1: length(f_vect));  %frequency vector to interpolate
TFmagtemp1=interp1(f_vectneg1, TFmagtemp,f_vectneg, 'linear','extrap'); %mag interpolation for negative
frequencies
TFphtemp1=interp1(f_vectneg1, TFphtemp,f_vectneg, 'linear','extrap');   %ph interpolation for negative
frequencies
TFHph=[TFphInt TFphtemp1];  %complete hermitian ph vector
TFHmag=[TFmagInt TFmagtemp1];   %complete hermitian mag vector
TFH=TFHmag.*exp(1j*TFHph); %complex hermitian transfer function
%%%%%%%%%%plot hermitian transfer function
%magnitude
figure(4)
subplot(2,1,1)
plot( f_vect,mag2db(TFHmag), 'r.-.')
title('Magnitude')
xlabel('Frequency (rad/s)')
ylabel('Magnitude (dB)')
%phase
figure(4)
subplot(2,1,2)
plot( f_vect,TFHph, 'r.-.')
title('Phase')
xlabel('Frequency (rad/s)')
ylabel('Phase (rad)')
%%%%%%%%%%group delay calcolation for phase recovery
TFHphase=angle(TFH);  %phase of the transfer function in radians at low frequency
param.groupdelay=(TFHphase(2)-TFHphase(1))/(2*pi*(f_vect(2)-f_vect(1)));    %group delay
param.nshiftdt=round(abs(param.groupdelay)/param.dt);   %delay in index of time to shift for the phase
recovery
%%%%%%%%%%filtering of the signal
Rsignal_nf=real(ifft(TFH.*fft(Rsignal_n)));
Isignal_nf=real(ifft(TFH.*fft(Isignal_n)));
if param.plot
    figure(1)
    plot(t, Rsignal_nf); hold on;
    title('Real signal')
    legend('Original signal','Noise', 'Noise Filtered')
    figure(2)
    plot(t, Isignal_nf); hold on;
    title('Imaginary signal')
    legend('Original signal','Noise', 'Noise Filtered')
end
disp('Filtered signal...');

%DISTORTION y = alpha1*x + alpha3*x^3
Rsignal_nfd=[];
Isignal_nfd=[];
temp1=[];
Rsignal_nfd=amp.alpha1*Rsignal_nf+amp.alpha3*Rsignal_nf.*Rsignal_nf.*Rsignal_nf;      % inphase or real
component
Isignal_nfd=amp.alpha1*Isignal_nf+amp.alpha3*Isignal_nf.*Isignal_nf.*Isignal_nf;      % Quadrature or
imaginary component
if param.plot
    figure(1)
    subplot(2,1,2)
    plot(t, Rsignal_nfd); hold on;
    figure(2)
```

```
    subplot(2,1,2)
    plot(t, Isignal_nfd); hold on;
end
disp('Distorted signal...');

%RECOVERY PHASE OF THE SIGNAL
Rsignal_nfd_rf=circshift(Rsignal_nfd, -param.nshiftdt);
Isignal_nfd_rf=circshift(Isignal_nfd, -param.nshiftdt);
if param.plot
    figure(1)
    subplot(2,1,2)
    plot(t, Rsignal_nfd_rf); hold on;
    title('Real signal')
    legend('Noise, Filtered, Distorted/gain', 'Phase recoveried')
    figure(2)
    subplot(2,1,2)
    plot(t, Isignal_nfd_rf); hold on;
    title('Imaginary signal')
    legend('Noise, Filtered, Distorted/gain', 'Phase recoveried')
end
disp('Recovered phase...');

disp('Start sampling signal...');
%-%-%-%-%-%-%-SAMPLING SIGNAL AND DATA RECOVERY%-%-%-%-%-%-%-
%RECOVERY REAL AND IMAGINARY SYMBOLS
RealdataRX=[];
ImgdataRX=[];
temp=param.uppersample/2;
for i=1:length(Rdatasym)
    (param.uppersample/2)+param.uppersample*(i-1);
    RealdataRX(i)=Rsignal_nfd_rf((param.uppersample/2)+param.uppersample*(i-1));
end
for i=1:length(Idatasym)
    (param.uppersample/2)+param.uppersample*(i-1);
    ImgdataRX(i)=Isignal_nfd_rf((param.uppersample/2)+param.uppersample*(i-1));
end
clear j;

%PLOT SCATTER CONSTELLATION AND COMPARE WITH THE SENDED SIGNALS
dataRXsym=RealdataRX+j*ImgdataRX;   %symbol generation received
dataRXsymIdeal=datasym*amp.s211G;   %symbol generation sended
pout=pnorm*amp.voutppd*0.5;   %symbol reference constellation
constDiagram = comm.ConstellationDiagram('ReferenceConstellation', pout, 'ShowGrid', true, 'XLimits', [-
amp.voutppd*2/3 amp.voutppd*2/3], 'YLimits', [-amp.voutppd*2/3 amp.voutppd*2/3], 'EVMNormalization',
'Average constellation power');
v=[];   %from row to column
for i=1:length(dataRXsym)
    v=[v;dataRXsym(i)];
end
constDiagram(v);

disp('Conversion from symbol to bit...');
%RECOVERING THE BIT STREAM
dataRXsymTemp=(dataRXsym/(amp.voutppd*0.5))*MAX_VAL;    %recover inizial amplitude for using qamdemod
dataRXbitTemp=qamdemod(dataRXsymTemp,sys.M, 'gray', 'OutputType', 'bit');
[r, c]=size(dataRXbitTemp);
dataRXbit=[];   %from matrix to string
for i=1:c
    dataRXbit=[dataRXbit; dataRXbitTemp(:,i)];
end

disp('Compute BER...');
%BER
ErrorRate = comm.ErrorRate;
ber=ErrorRate(dataRXbit, databit);
disp('BER :');
ber(1)
disp('Number of error :');
ber(2)
disp('Number of sended bits :');
ber(3)
```

ACKNOWLEDGES

I would like to thank all of my collaborators, first of all to professors Manstretta Danilo and Castello Rinaldo for the opportunity of this research's work. I would like to thank also the team of Huawei Milano, in particular Paolo Rossi (ex. employee), Federico Vecchi, Luca Romanò, Danilo Caccioli, Federico Persia (ex. employee), and Marco Della Santa for their technical support.

Other acknowledges are for my family and my friend Giada Covini that knows always how to make me smile.

A very special acknowledgment is for Raul Ciuche, because even when the days were long and the spirit was down he always knew how to sustain me, make me laugh, and encourage me to do what was better for me.