Development of integrated emission architectures for THz applications in 28nm FD-SOI CMOS technology

Sébastien Sadlo

THÈSE PRÉSENTÉE

POUR OBTENIR LE GRADE DE

DOCTEUR DE

L'UNIVERSITÉ DE BORDEAUX

ECOLE DOCTORALE DES SCIENCES PHYSIQUES ET DE L'INGÉNIEUR
SPÉCIALITÉ : ELECTRONIQUE

Par Sébastien SADLO

"Development of integrated emission architectures for THz applications in 28nm FD-SOI CMOS technology"

"Développement d'architectures d'émission intégrées pour les applications THz en technologie CMOS avancées"

Thèse CIFRE Préparée en collaboration avec :
STMicroelectronics, Crolles, France
Laboratoire IMS, Bordeaux, France

Soutenue le : 30 novembre 2023

Membres du jury :

Mme.	Daniela DRAGOMIRESCU	Professeure des universités	INSA Toulouse	Rapporteure
M.	Ehsan AFSHARI	Professeur des universités	University of Michigan	Rapporteur
M.	Yann DEVAL	Professeur des universités	Université de Bordeaux	Président
M.	Andreas KAISER	Professeur des universités	ISEN	Examinateur
M.	Raphael GUILLAUME	Ingénieur-Docteur	STMicroelectronics	Invité
Mme.	Nathalie DELTIMPLE	Professeure des universités	Bordeaux INP	Directrice de these
Mme.	Andreia CATHELIN	Docteure HDR	STMicroelectronics	Co-directrice de these

Abstract

The THz part of the spectrum allows disruptive applications such as 100 Gb/s wireless communications, imaging, and spectroscopy and offers an interesting perspective for the 6G future standards. The need for this kind of application is rising and nanoscale VLSI CMOS technologies dispose of active devices with high enough ft and fmax to integrate Systems on Chips in this frequency range.
This work presents research results towards the design of amplification circuits above fmax/2 in 28nm FD-SOI CMOS technology. This thesis enriches state of the art for amplifier gain boosting design methodology and extends it for large bandwidth operation. The theoretical study of this work compares the proposed solution with the classical approaches. In light of the state of the art, an analysis of the performance's limitations regarding the saturated output power, bandwidth, and efficiency is presented. Design methods to improve these metrics are proposed and the design and measurements of 200 GHz power amplifiers are presented.

Keywords: CMOS, FD-SOI, sub-THZ, 6G, power amplifier, silicon, integrated circuit

Résumé

La partie du spectre THz permet des applications disruptives telles que les communications sans fil à 100 Gb/s, l'imagerie, la spectroscopie et offre une perspective intéressante pour les normes futures de la 6G. Le besoin pour ce type d'applications est en hausse et les technologies CMOS VLSI à l'échelle nanométrique disposent de dispositifs actifs avec une f_t et une f_{max} suffisamment élevées pour intégrer des systèmes sur puces dans cette plage de fréquences.
Ce travail présente des résultats de recherche pour la conception de circuits d'amplification au-dessus de $f_{max}/2$ en technologie CMOS FD-SOI 28nm. Cette thèse enrichit l'état de l'art pour la méthodologie de conception d'amplification de gain et l'étend pour une grande opération de bande passante. L'étude théorique de ce travail compare la solution proposée avec les approches classiques. À la lumière de l'état de l'art, une analyse des limitations de performance concernant la puissance de sortie saturée, la bande passante et l'efficacité est présentée. Des méthodes de conception pour améliorer ces métriques sont proposées et la conception et les mesures d'amplificateurs de puissance de 200 GHz sont présentées.

Mots clés : CMOS, FD-SOI, sub-THZ, 6G, amplificateur de puissance, silicium, circuit intégré

Acknowledgments

Many people have contributed along the journey of this thesis. The support and guidance of my advisors, Nathalie Deltimple and Andreia Cathelin, kept me on course in the right direction. Their knowledge of the domain and the research world allowed them to identify this subject at the right time.

I would like to acknowledge professors Daniela Dragomirescu and Ehsan Afshari for accepting to review this PhD dissertation. I want to thank all the members of the jury, Yan Deval, Andreas Kaiser and Raphael Guillaume for participating to the PhD presentation and defense. All the feedback and the interesting discussions allowed to enrich the PhD defense.

My family has played a key role throughout the years, always supporting and encouraging me. My parents, Alina and Jan, and my sister, Patricia, were always attentive and had the right words to get through the difficult days. Special thanks to Stefania, I could not have done it without you.

The environment of this thesis between STMicroelectronics and the IMS laboratory was very rich, each person that I had the chance meet during these years contributed in a certain way to the thesis.

I would like to thank my ST colleagues from the R&D Team where everything started during my internship, prior to the thesis, and where everything continues today. Thank you Christophe for your formative supervision, your discipline and good mood. Thank you Antoine and David for the pleasant atmosphere during our PhD years and now our daily work.

I am thankful to the PhD team “Chip Decathlon” : Denis Flores, Andres Aprilla, Khalil Bouchoucha, Soufiane Mourrane, Julien Poupon, Virgile Colrat and Adrien Verplancke. We were on the same boat, some of us arrived at destination and others are still on their way. But the great time working by your side leaves nothing but pleasant memories despite the quantity of work, the stress and the fatigue that we endured.

I would like also to thank people from ST that had advised me along the thesis, Denis Pache, Frederic Paillardet, Philippe Cathelin, your experience and your kindness are a great example to follow.

Finally, I would like to thank the people that I met and worked with at the IMS lab, Antoine Lhomel, Colin Boulanger, Emmanuel Pages, Remi Queheille, Timothée Le Gall, Magali De Matos, Sebastien Fregonese, Marina Deng, Jean-Charles Henrion, Gwennael Diverrez, the experience and the memories from this period have greatly fed my PhD journey.

Table of Contents

List of figures

Chapter 1 Introduction

The design of Radio-Frequency (RF) circuits has first been developed with the main target of transmitting data wirelessly over long distances. As the field grew and the technologies and the circuit performances improved new applications were thought of. The use of higher and higher bandwidths and frequencies created opportunities to exploit wave-matter interactions. The millimeter-wave (mmWave) and Terahertz (THz) frequencies belong to the highest frequencies that can be generated today by RFIC technologies. The matter reacts differently over this large spectrum part and signals of these bands can be used for spectroscopy, imaging, or space radiation observation. In addition, the photon's energy is very low compared to other imaging techniques such as X-Rays, and consequently mmWave and THz frequencies do not ionize and damage the matter. The wavelength in free space is equal to 3mm at 100GHz and further decreases with higher frequencies. Thus, very precise distance measurements can be done using these frequency signals in industry, in radar applications, and in the biomedical field. Wireless communications also benefit greatly from the large available bandwidth in these bands. This enables data rates one order of magnitude higher than mmWave 5G applications, up to hundreds of gigabits per second. With the available low-cost silicon technologies, the development of RF front ends opens the door to numerous mass-market disruptive applications.

However, the implementation of such RFICs deals with components close to their limits and lumped element approximation is no longer applicable. Hence the device modeling, the passive elements design and simulation, and the circuit blocks architecture must be adapted to the constraints. This thesis work has the three following main goals to path the way towards the design of sub-THz RFIC systems:

-After evaluating the limitations of devices, propose a design methodology to increase the transistor and passive devices' performances

-Propose innovative circuit architectures adapted to the high operation frequencies by taking advantage of the small transmission line and passive elements due to the short wavelength

-Demonstrate the capability of the VLSI CMOS FD-SOI technology for future 6G and other over 100GHz applications.

This manuscript is composed of six chapters. The first one is dedicated to the general introduction with the presentation of the context, the main goals, and the outline of this thesis.

The second chapter presents the available opportunities in mmWave and sub-THz frequency bands. The state of the art in these bands is presented and the design challenges are discussed. Finally, it highlights the problems that this thesis wants to address.

The third chapter presents the integrated circuit (IC) technologies and their limitations to design at mmWave and sub-THz frequencies. The high-frequency transistor and passive device's behavior are presented.

The fourth chapter studies the Gain Boosting design technics to enable the design close to the f_{max}. It presents a design methodology for sub-THz Gain Boosted amplifiers and the implementation of an amplifier at 184 GHz in 28nm FD-SOI CMOS technology from STMicroelectronics.

The fifth chapter further investigates the gain boosting techniques for the design of sub-THz amplifiers having at the same time great performances in gain, saturation power level, and operation frequency bandwidth. The limitations of each metric are studied, and large bandwidth gain boosting, power combining, and large bandwidth impedance matching technics are presented. Finally, the design and measurement results of a 190GHz power amplifier are presented.

The last sixth chapter of this thesis concludes with the presented work and discusses the future research perspectives regarding the integration of systems in VLSI technologies operating at mm-Wave and THz frequency bands.

Chapter 2 Introduction to THz

The terahertz (THz) frequency band is located between the millimeter wave and the far-infrared spectrums, in the 300GHz to 3THz range (Figure 2.1). The wavelengths in vacuum at these frequencies are under the millimeter and range from 0.1mm to 10µm. At these frequencies, the short wavelength can be used for precise imaging or distance measurements for industrial or radar applications. Some materials are transparent for these frequency waves and these properties can be exploited to realize scanning for security or industrial purposes. Along the large THz frequency bandwidth, numerous molecular absorption frequencies are present and therefore it can be used for spectroscopy. These frequencies are neither harmful nor destructive as the X-Rays are and consequently are an interesting alternative. Finally, a relatively small bandwidth to the frequency of operation represents multiple gigahertz (GHz) that enables from multiple to hundreds of gigabits per second communication data rates. These frequency bands are disruptive for wireless communication given the achievable bandwidths.

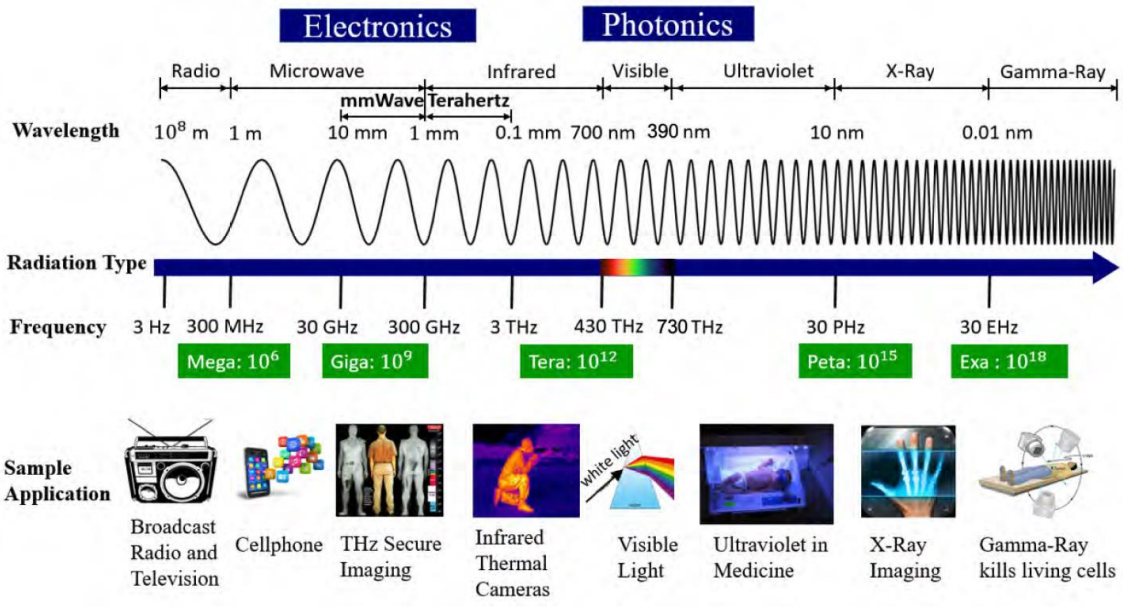


Figure 2.1 The electromagnetic spectrum, and various applications as a function of frequency [Rapp19]

The THz frequencies offer a large scope of new applications but the available components operating at these frequencies are limited. Frequencies above the THz bands are generated with optical components while below frequencies are in the millimeter wave range and are thus generated with electronic circuits. The lack of solutions in this frequency band is known as the THz gap and is illustrated in Figure 2.2.

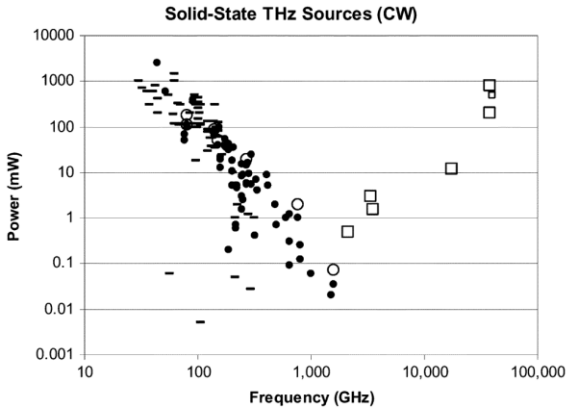


Figure 2.2 The THz gap with respect to source technology, including quantum cascade lasers (□), frequency multipliers (●), and other electronic devices (–). Hollow symbols represent cryogenic results [Crow05]

The sub-THz band is in the 100GHz-1THz range and still presents the advantages and opportunities of the THz band. However, it is in the scope of operation of recent electronic devices that present maximal operation frequencies of a couple of hundred Gigahertz. Therefore, research to integrate the previously described systems in silicon integrated circuits is increasingly present in the academic and industrial worlds. In addition, the low cost of silicon-integrated circuit technologies enables the deployment of sub-THz circuits for mass-consumer markets. However, the design of such circuits is challenging as the operation frequencies are close to the device's limits. New architecture and design methodologies must be developed to overcome these limitations.

The following parts of this chapter will first present the main possible applications at sub-THz frequencies. Then the need for high wireless communication data rates will be discussed. Following, possible communication architecture will be presented and a summary of the state of the art of the literature will be depicted. Finally, the actual limits of the circuits presented in the SOA will be highlighted and compared with the objectives of the thesis.

2.1 Physical opportunities of the sub-THz spectrum

Figure 2.3 illustrates the numerous possible applications that enable the sub-THz spectrum, the main applications are presented in the following sections.

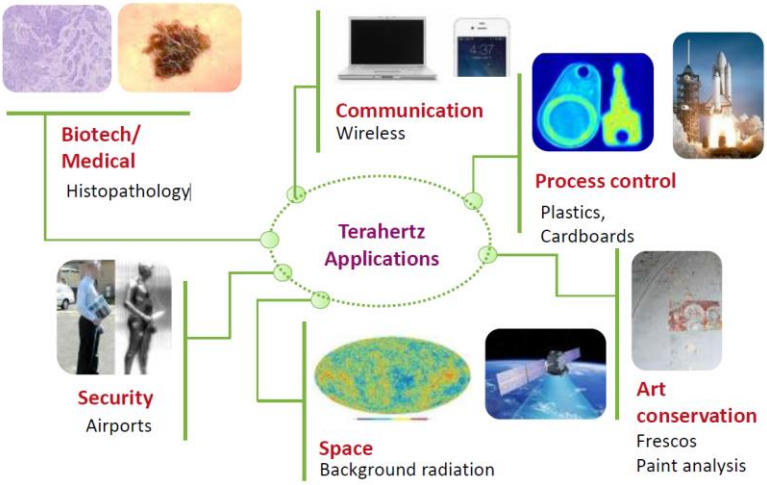


Figure 2.3 Main THz applications map [Pfei19]

2.1.1 Imaging

The transparency of some materials to THz waves opens the opportunity to realize imaging captors using these wavelengths. Multiple cases can leverage this technology.
Security imaging can greatly benefit from it as metallic materials are reflective at these frequencies and the detection of weapons can be done as depicted in Figure 2.4.

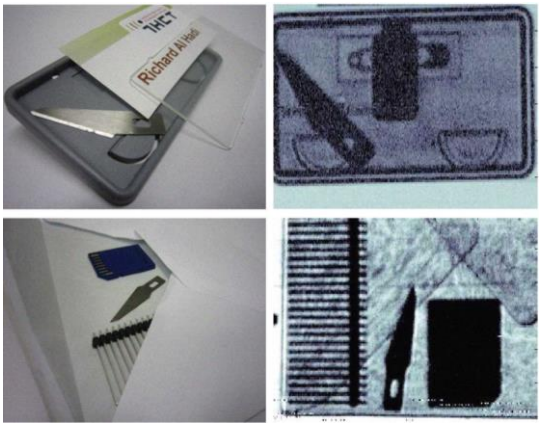


Figure 2.4 Transmission-mode images at THz of a badge and an envelope [Hadi11]

THz security scanning has the advantage of detecting metallic objects without harming people or altering materials. It has all the advantages of X-ray imaging as it depicts the objects' forms but without the inconveniences of the X-ray systems that require specific expensive radiation generators. Silicon solutions are handy and low-cost in comparison.
On the other side of imaging applications, biomedical imagery can benefit from THz as well. The organic matter reaction to the THz illumination depends on its chemical composition. Therefore, THz imaging can be used to discriminate abnormal tissues. Figure 2.5 depicts the imagery of a mouse brain where the cancerous parts appear on the THz images.

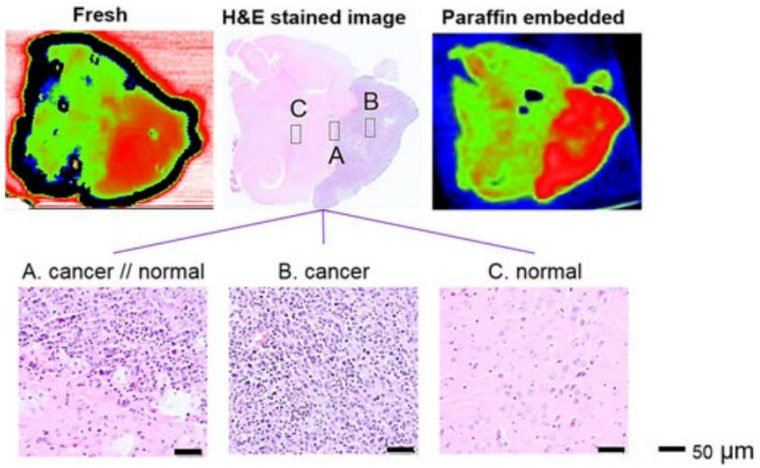


Figure 2.5 THz images of a mouse brain presenting cancerous parts [Cheo17]

2.1.2 Spectroscopy

Sub-THz and THz frequency bands are sufficiently large to present multiple molecular absorption frequencies. Spectroscopy can be realized using these frequencies to identify specific components. Hence, sub-THz circuits can be used for the pharmaceutics industry or gas detection. Figure 2.6 presents a 220 to 320GHz CMOS rotational spectrometer for molecular detection [Wang18].

The spectrometer can analyze a mixture of gazes and simultaneously identify multiple molecules from the absorption rays in the 220 to 320GHz band.

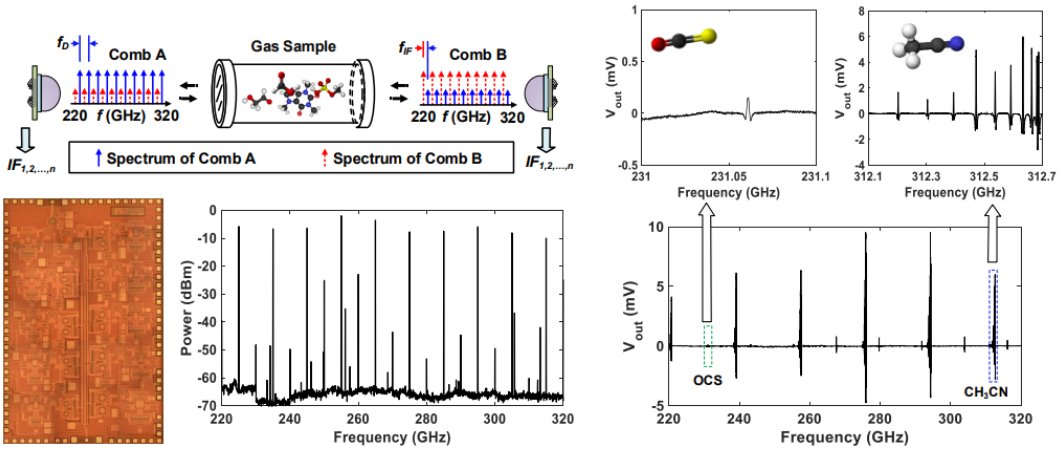


Figure 2.6 Dual-THz-comb spectrometer for gas sensing [Wang18a]

THz spectrum enables unprecedented wave-matter interaction applications using IC technologies. Molecular absorption frequency is a very precise data that can be used as an absolute reference. Another use of the wave-matter interaction is a chip-scale clock referenced on a molecular absorption frequency. By referencing the electronic clock oscillation to the rotational-mode transition of OCS molecules at the sub-THz, the clock achieves an “atomic-clock-grade” precision using a mainstream 65nm CMOS process IC and a THz waveguide [Wang18b].

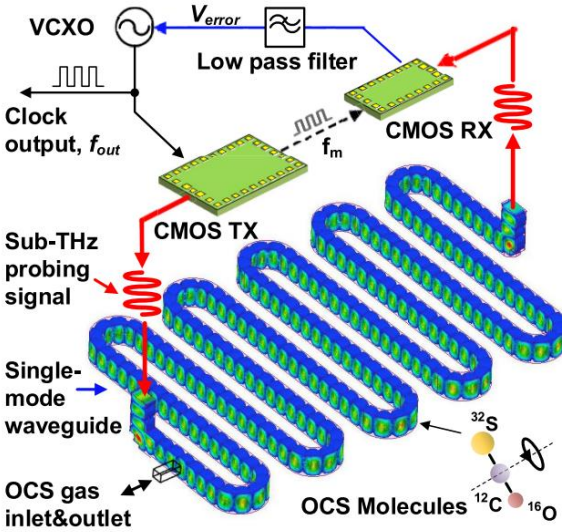


Figure 2.7 CMOS molecular clock implementation [Wang18b]

2.1.3 Radar

High-resolution integrated radars are crucial in nowadays’ automotive, vital sign, and security sensing applications. The sub-THz/THz spectrum shows great opportunities in both high-resolution and all-weather radar imaging capabilities.

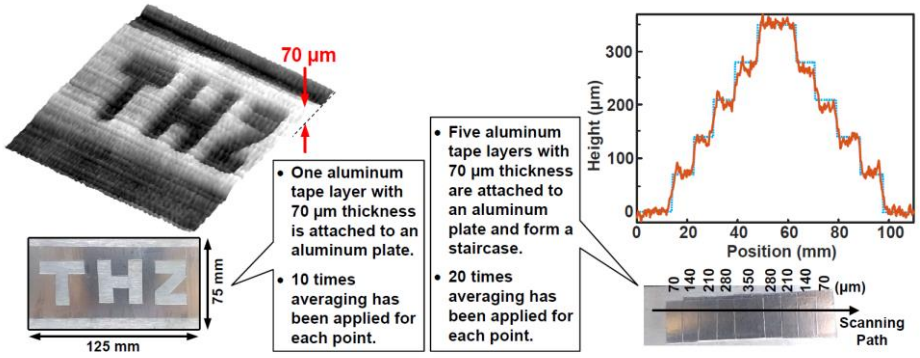


Figure 2.8 THz images [Hoss21]

Figure 2.8 presents the micrometer distance measurement capability of a 250GHz FMCW radar [Hoss21]. Thanks to the short wavelength a micrometer resolution can be achieved. It has promising applications in the industry for process and quality control.

2.1.4 Communication

The world is increasingly interconnected and every increase in available bandwidth is leveraged by actual applications and enables new disruptive usages. The number of connected devices increases rapidly as well, leading to an exponential increase in wireless communication capability demand. The spectrum is saturated up to 100GHz and a significative bandwidth increase is not possible in the frequency ranges used for 5G, 4G, or Wifi standards. The next section focusses on the motivations to use the sub-THz spectrum for the next generations of wireless communications.

2.2 Largely available bandwidths, increasing needs for data rates

The data rate of communication has been increasing exponentially year-by-year as presented in Figure 2.9 [Fuji18]. Wired communications were dominant regarding the data rates for decades. Wireless communications were in another class and the comparison wasn't pertinent. However, the demand for wireless data rates increased much faster than the wired one to the point that we can predict that in the next decade, the wireless needs will meet the optical wired data-rates. Conversely, in the optical domain, the wireless world requires disruptive innovations to follow the demand. Hundreds of Gigabits per second are conceivable with actual research results but the increase to the Terabit per second area still requires R&D.

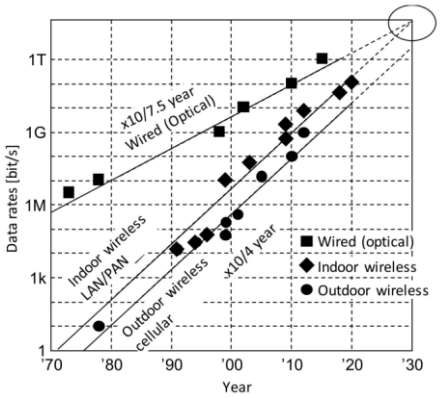


Figure 2.9 Evolution of data rate of wired communication and wireless communication over the years [Fuji18]

Successive generations of communication standards were developed by the industry, with a couple of Mb/s for the first indoor generations and only hundreds of kilobits per second for the first outdoor standards. In comparison to today's available rates, even with a 1.4-time increase year-by-year [Okad22], the first years' rates were still low, but each evolution opened the door for disruptive applications. Figure 2.10 presents the chronological evolution of these standards since 1995 and predicts by following the trend of an available 1Tbps link in 2035.

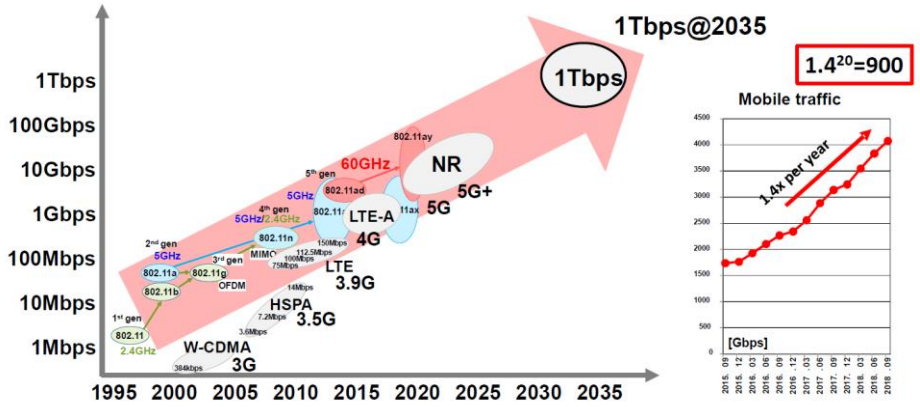


Figure 2.10 Evolution of wireless communication standards' data rates over the years [Okad22]

The rate of information that can be transmitted is bounded to two metrics, the used bandwidth and the signal-to-noise power ratio. Channel capacity C is the theoretical upper boundary on the information data rate that can be communicated for any error rate. The Shannon-Hartley theorem gives the following expression of C:

$$C = BW.\log_2\left(1+\frac{S}{N}\right) \tag{2.1}$$

Where *BW* is the signal bandwidth, S is the average signal power over the bandwidth, and N is the average noise power over the bandwidth.

The spectral efficiency can be defined as the maximal data rate per unit of frequency. Therefore, it is bounded as well with a quantity defined from (2.1):

$$\eta = \frac{C}{BW} = \log_2\left(1+\frac{S}{N}\right) \tag{2.2}$$

Equation (2.2) shows that complex signal modulations, which increase the spectral efficiency, require a high signal-to-noise ratio. Hence the spectral efficiency is limited by the signal power level. However, the signal power is limited for safety and practical reasons. For connected objects, battery life is crucial, and it is desired to reduce power consumption. Consequently, the bandwidth remains the only leverage to increase the channel capacity. For this reason, wireless standards tend to use higher frequencies to address the demand. Nowadays, the last generations of Wifi use 2.4GHz, 5GHz, and up to 10GHz frequency bands. For higher short-distance data rates, the WiGig uses the V-Band 60GHz available band. For outdoor technologies, the last 5G standard uses mm-Wave frequency bands mostly around 30GHz. Figure 2.11 presents the available spectrum and channel bandwidth in function of the carrier frequency. The link capacity is also depicted and illustrates the correlation between used bandwidth and capacity.

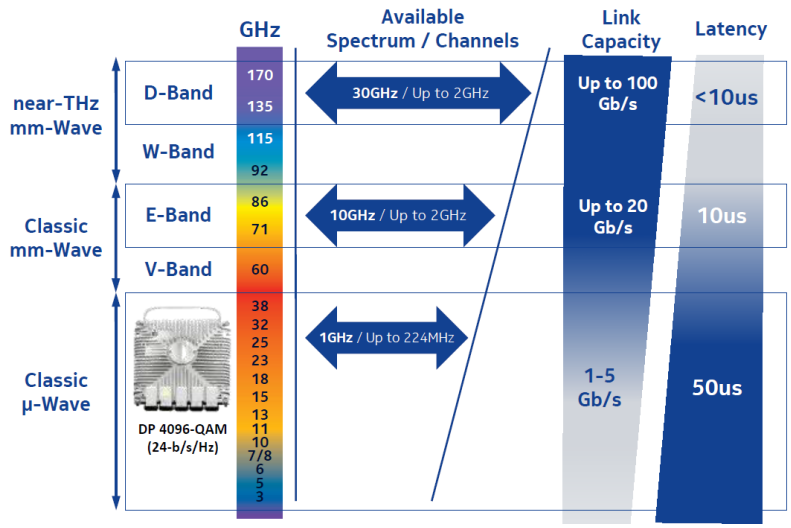


Figure 2.11 Available spectrum and channel bandwidths as a function of the carrier frequency [Shah22]

The D-Band which is in the 135-170GHz range is expected to be used in the future and offers all the necessary bandwidth to achieve >100Gb/s rates. The RF and mm-Wave design community put a great effort recently to work towards a solution in D-Band.

Several characteristics are already expected for the next wireless communication standard. The 6th generation needs to enable >100Gb/s link capacity with a low energy and cost per transmitted bit to be sustainable with the data traffic increase in a world that looks to reduce its energy consumption. The 6G will also have an important infrastructure evolution and numerous innovations to offer extremely low latency and high reliability. Figure 2.12 resumes the different expected evolutions of the 6G.

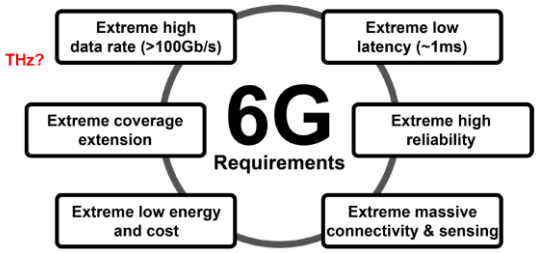


Figure 2.12 6G Requirements [Okad22]

The increase in carrier frequency comes with the drawback of a higher path loss while transmitting wireless signals. The Friss equation gives the relation between the transmitted signal power and the received one with the formula:

$$P_r = P_t D_t D_r \left(\frac{\lambda}{4\pi d}\right)^2 = P_t D_t D_r \left(\frac{c}{4\pi d f}\right)^2 \quad (2.3)$$

With,

P_r: the received power,

P_t: the transmitted power,

D_r: the receiver antenna gain,

D_t: the transmitter antenna gain,

λ: the wavelength of the signal

c: the speed of light in a vacuum

f: the frequency of the signal

d: the distance between the transmitter and the receiver

The received power level is proportional to the emitted power and the antenna's gain, which depends on how the antenna can concentrate the signal power in one direction. It considers only the attenuation due to the propagation in free space and represents the minimal loss that can be achieved. The free space path loss (FSPL) is equal to the remaining factor of (2.3),

$$FSPL = \left(\frac{c}{4\pi df}\right)^2 \tag{2.4}$$

The FSPL is inversely proportional to f², hence a 100GHz signal experiences a 20dB higher loss than a 10GHz signal. The other parameters of the Friss Equation (2.2) are frequency independent, consequently at equal emitted signal levels the received power is 20dB lower for a 10 times frequency increase. This represents an additional challenge for >100GHz transceivers and architectures must compensate for these increasing losses.

In addition, the atmosphere presents absorption frequencies that increase the propagation losses in some frequency ranges. Figure 2.13 (left) presents the atmospheric attenuation as a function of frequency. The 20dB per decade free space loss increase is visible in addition to the absorption losses. At 60GHz the additional attenuation is 15dB, there is another absorption around 120GHz, and then a 28dB additional attenuation at 183GHz. The absorption losses tend to increase with frequency. These bands can be used for short-range communications but moderate to long-range are not conceivable because of the additional losses. Figure 2.13 (right) illustrates this range of limitations by representing the communication distance from 100 GHz to 350 GHz. The communication distance is calculated from the distance at which the atmospheric attenuation reaches 10 dB. In the frequency band from 192 GHz to 298 GHz, the communication distance is over 2 km, and in the frequency band from 121 GHz to 154 GHz, the communication distance is over 8 km.

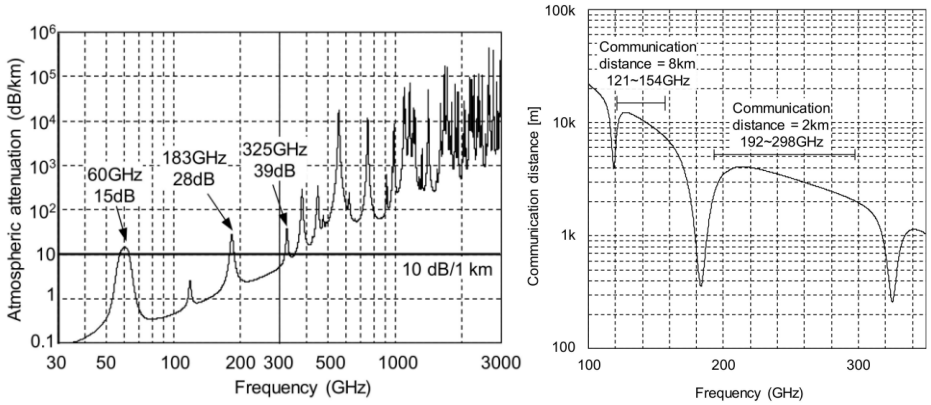


Figure 2.13 Atmospheric attenuation (left) [Fuji18]

Another phenomenon arises with frequency. The antenna size is proportional to the wavelength of the signal, hence an increase in frequency implies a smaller antenna size. Consequently, a 10 times increase in frequency leads to a 10 times smaller antenna. For a fixed antenna array surface as presented in Figure 2.14, 100 times more antennas fit in the same area at 300GHz than at 30GHz. For an antenna array, the gain directivity of the N antenna array is equal to N times the gain of one antenna. However, as the number of antennas increase, the generated beam become thinner. This opens the door for a precise beam steering that can be used in some cases of communication or radar systems but can also be a limitation. As the number of elements increases the precision of the phase shift between antennas has to increase to be able to target precisely a receiver.

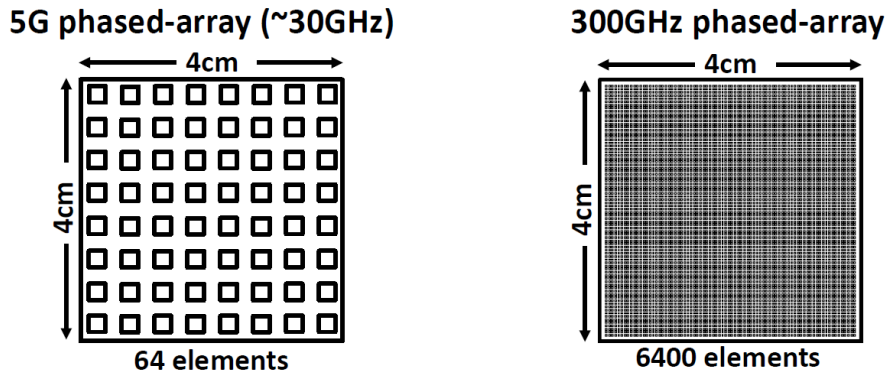


Figure 2.14 Comparison of 30GHz and 300GHz antenna arrays at constant surface [Okad22]

When the number of antennas is scaled to the wavelength on the emitter or the receiver only, the 20dB antenna array gain improvement due to 100 times more antennas compensating the 20dB additional path loss due to the frequency increase (for example 30GHz vs. 300GHz).

When the antenna array elements are multiplied by 100 on the transmitter and receiver ends, the antenna array gains are increased by 20dB on each side. Thus, 40dB compensates for the 20dB additional path loss due to the frequency and leaves 20dB more received power.

In conclusion, at a constant antenna array surface, for a 10 times frequency increase, the link margin is improved by 20dB. At constant total transmitted power P_t, the total power is combined from all the unitary elements. At 300GHz each element needs to provide 20dB less power than each element of the 30GHz array. For a constant distance link, because of the additional 20dB link margin at 300GHz, the required output power of each element is decreased by 40dB compared to unitary elements of a 30GHz array. This means that a directive 300GHz link is 20 times more efficient than a 30GHz link. As the frequency increases the achievable power levels that can be generated by active devices tend to decrease but because of this antenna array advantage, the constraints are much lower. Consequently, the design of such systems at these frequencies is not limited by the active devices' performance decrease. Figure 2.15 illustrates the previous demonstration by comparing the received power level from transceivers at 28GHz, 73GHz, and 140GHz over distance in the case of directional or not antenna arrays on the RX and/or TX side.

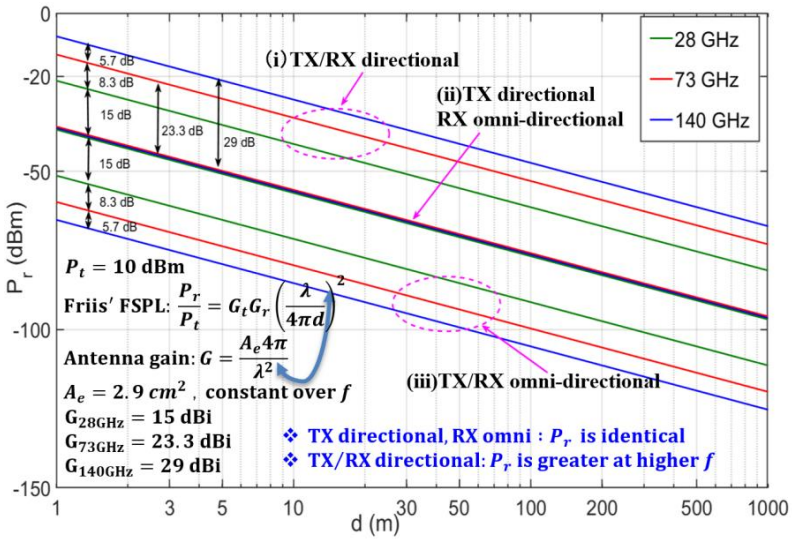


Figure 2.15 Received power vs. distance for different signal frequency and Tx/Rx antenna configurations [Xing18]

To pave the way for sub-THz wireless systems an amendment has been ratified to standardize and attribute frequency bands for such future systems [IEEE17]. The 250-320GHz frequency band has been chosen to allocate a 70GHz continuous band for a 100 Gb/s Wireless Point-to-Point Physical Layer. The available frequency range has been divided as depicted in Figure 2.16 into multiple channels [Petro20]. Different divisions are proposed, from 2.16GHz channel width to the complete 69.12GHz.

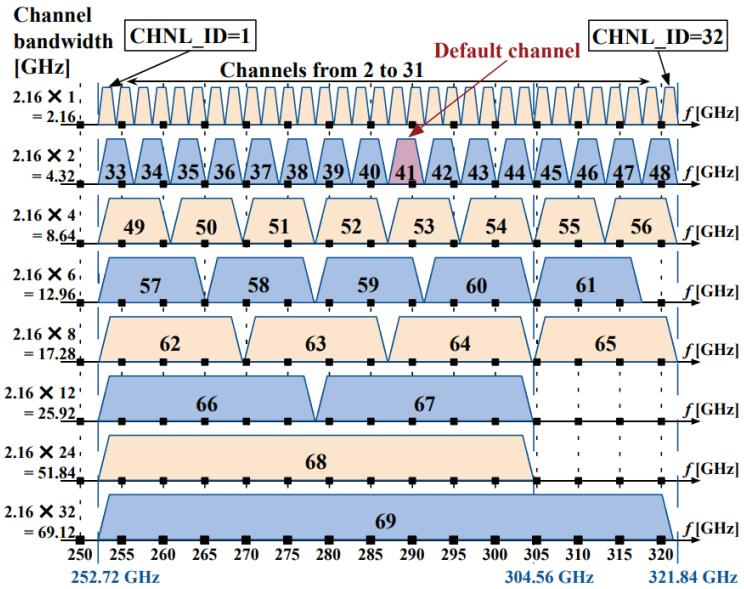


Figure 2.16 IEEE 802.15.3d channel plan [Petro20]

2.3 Architectures: Frequency upconverting, Harmonic multipliers, Fundamental

The motivations are numerous, and the frequency bands are available for sub-THz wireless communications. The design of RF front-ends must face the technological limitations of the IC components. A modulated signal that carries the data to be transmitted is at a given intermediate frequency (even baseband signal) and is up-converted to the carrier frequency using a mixer. The operation can be done in multiple manners, but the result is a modulated signal around a sub-THz

carrier frequency. Most of the modulation and operations can be done before the mixer at frequencies that are far below the maximal frequency of the active components and don't face difficulties. However, after the mixer, the operations have to be done near or even passed the component frequency limit. Multiple cases can be distinguished.

As presented in Figure 2.17, if the carrier frequency is below f_{max} the last front-end block is the Power Amplifier (PA) to increase the power level of the signal for efficient transmission. This architecture can be named PA last. However, the design of sub-THz amplifiers is challenging and results in moderate to poor efficiency and moderate power levels. The drawback of these schemes is that the maximal carrier frequency is limited by the components f_{max}, usually, the designs become very complicated from $2*f_{max}/3$.

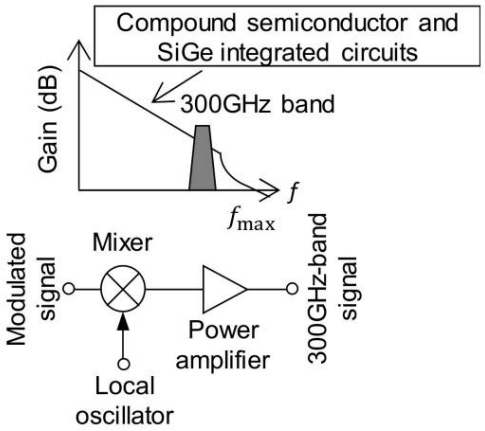


Figure 2.17 Schematic of PA last emitter architecture [Fuji18]

To achieve even higher frequencies, the signal from the PA can be multiplied with a non-linear block. This way frequencies well above f_{max} can be achieved. This architecture is depicted in Figure 2.18 and can be referred to as multiplier base architecture. Besides the advantage of the high output frequency, two limitations exist. The multiplication operation presents a consequent conversion loss, consequently, the output power level is greatly decreased by the multiplier. Usually, the higher the multiplication operation the higher the losses, only a fraction of the input power is available at the ouput. Therefore, it significantly decreases the efficiency of the transmitter front-end. Another drawback of the multiplication-based architecture is the multiplication of the signal bandwidth at the same time. A 10GHz wide signal at intermediate frequency will occupy N times larger bandwidth after an N times frequency multiplication. This operation diminishes greatly the spectral efficiency of the wireless communication link.

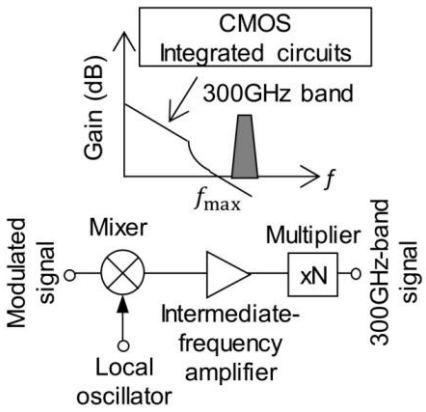


Figure 2.18 Schematic of multiplication-based emitter architecture [Fuji18]

An alternative architecture is the mixer last scheme, where the signal is upconverted after being amplified by the PA. The achievable output frequencies can be above $f_{max,}$ but the upper limit is $2f_{max}$ as the LO and the IF frequencies must be in the operation range of the devices. The output power level is low compared to the PA's last architecture because of the conversion losses of the mixer. Finally, the final bandwidth around the sub-THz carrier frequency is not multiplied and remains the same as in IF. Therefore, the spectral efficiency is not decreased with this method.

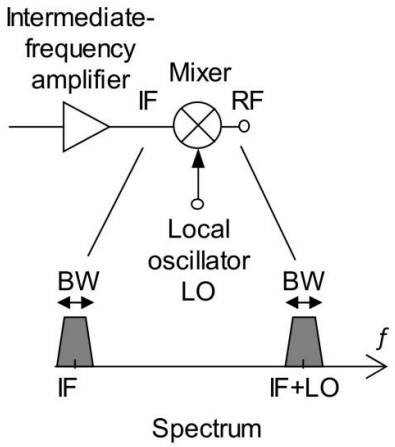


Figure 2.19 Schematic of mixer last emitter architecture [Fuji18]

The generation of sub-THz modulated signals can be done using one of the above methods with the presented advantages and drawbacks. But the receiver-side implications are important as well. The multiplication-based emitter cannot carry complex modulations as the complementary division operation is not possible. A power detector can be used but the choice of receiver architecture is limited. The PA last and mixer last architectures have similar signal characteristics at the difference of a lower power level for the latter. The complementary receivers can be mixer-first or LNA-first architectures. The mixer-first receiver is required for the mixer-last emitter and comes with a high NF. To complete the PA last emitter, both receivers can be used. The LNA first receiver offers a lower NF metric compared to the mixer first.

In summary, to achieve the best link budget and consequently be able to use complex modulations to increase spectral efficiency, the PA last and LNA first are preferred. The other architectures are interesting for imaging, spectroscopy, and radar sub-THz applications.

2.4 Transceivers over 100GHz

The development of integrated communication transceivers over 100GHz accelerated in the last couple of years. Multiple publications already presented demonstrators at these frequencies. The performances of such systems are not limited only by the RF front end, but the IF, LO, ADC, and DAC performances must support very high data rates. Therefore, transmitters and receivers can be analyzed separately and face different challenges.

The final transceiver performances are defined either by one end or the other. The transmitted data can be coded into the amplitude and/or phase of the RF signal.

To increase spectral efficiency, Quadrature Amplitude Modulation (QAM) is commonly used. The constellation that represents all the possible amplitude and phase of QAM modulations to code information into the RF signal is presented in Figure 2.20. The black dots in Figure 2.20 (a) depict a 4-QAM constellation as 4 possible states are possible. Therefore, the 4 states are coded into 2 bits: “00”, “01”, “10” and “11”. Each symbol codes for 2 bits, consequently, the spectral efficiency

of the 4-QAM is 2 bit/Hz. Higher order modulations are 16-QAM in Figure 2.20 (b) and 64-QAM in Figure 2.20 (c) and have spectral efficiency of 4 bit/Hz and 6 bit/Hz respectively. Generally, a 2^N-QAM modulation has an N bit/Hz efficiency. In comparison, the maximal amplitude of the three QAM modulations are equal, only the number and the distance between the points differ. The operations that are operated on the modulated signal are done with non-ideal blocks. Therefore, the consequence is an alteration of the modulated signal, leading to changes in the phase and amplitude. Consequently, the alteration causes a movement on the constellation plane and can lead to ambiguities or errors in the demodulation. The red dots on the constellations of Figure 2.20 are displaced by the same error vector represented with a red arrow. The errors are of the same amplitude and phase for the three different modulations. However, the consequences are different for each case.

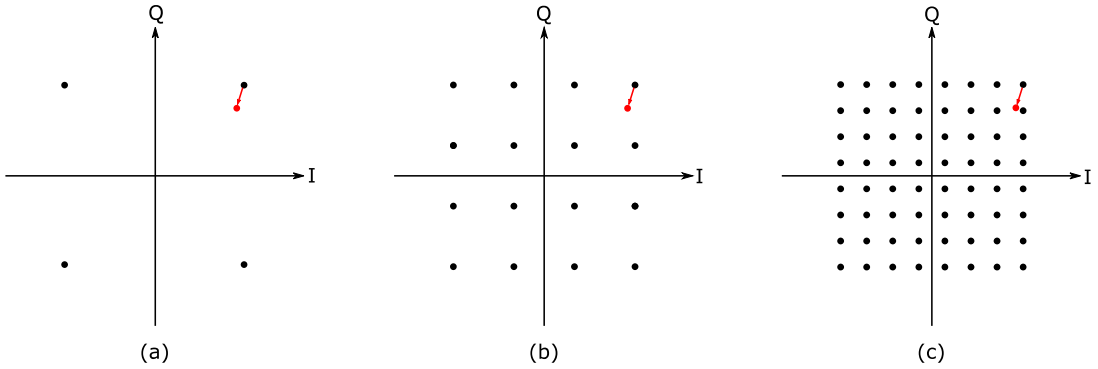


Figure 2.20 Representation of 4-QAM (a) 16-QAM (b) and 64-QAM (c) constellations in black, with a non-ideal symbol in red and its associated error vector

For 4-QAM modulation, the error doesn't prevent decoding the information. For the 16-QAM decoding is still possible, but the dot is close to the middle point between multiple possible black dots. However, for the 64-QAM modulation the same displacement causes a loss of information as the red dot is closer to other constellation dots. This means that a higher-order modulation is less robust. For a large amount of modulated data, a circuit can be characterized with the Error Vector Magnitude (EVM) which is defined as the root mean square (RMS) of the error vector amplitudes and can be expressed as a percentage as:

$$EVM_{\%} = 100\sqrt{\frac{\sum_{i=1}^{N_S}\frac{|R_i - S_i|^2}{|S_i|^2}}{N_S}} = 100\sqrt{\frac{\sum_{i=1}^{N_S}|E_i|^2}{N_S}} \qquad (2.5)$$

Where R_i is the received i^{th} symbol, S_i is the i^{th} ideal symbol, E_i is the i^{th} error vector, and N_S is the number of symbols.

It can also be expressed in decibels (dB) as:

$$EVM_{dB} = 20log_{10}\left(\frac{EVM_{\%}}{100}\right) \qquad (2.6)$$

The EVM depends on the type of transmitted data but for many random symbols, it gives a good evaluation of circuit performance. Therefore, it can be used to compare different circuits at the same data rates.

The alteration of the signal can be represented as a noise that is added to the signal. The higher the EVM, the higher the noise. Hence a relation between the EVM and the Signal to Noise Ratio (SNR) can be established [Shaf06]:

$$SNR = \frac{1}{EVM_{lin}^2} \tag{2.7}$$

Once the distortion is represented with noise, the communication theory gives the relation between the SNR and the Symbol Error Rate (SER) for an M-QAM modulation. From [Gold05] they are related with:

$$SER = 1 - \left(1 - \frac{2\left(\sqrt{M}-1\right)}{\sqrt{M}} Q\left(\sqrt{\frac{3SNR}{M-1}}\right)\right)^2 \approx \frac{2\left(\sqrt{M}-1\right)}{\sqrt{M}} Q\left(\sqrt{\frac{3SNR}{M-1}}\right) \tag{2.8}$$

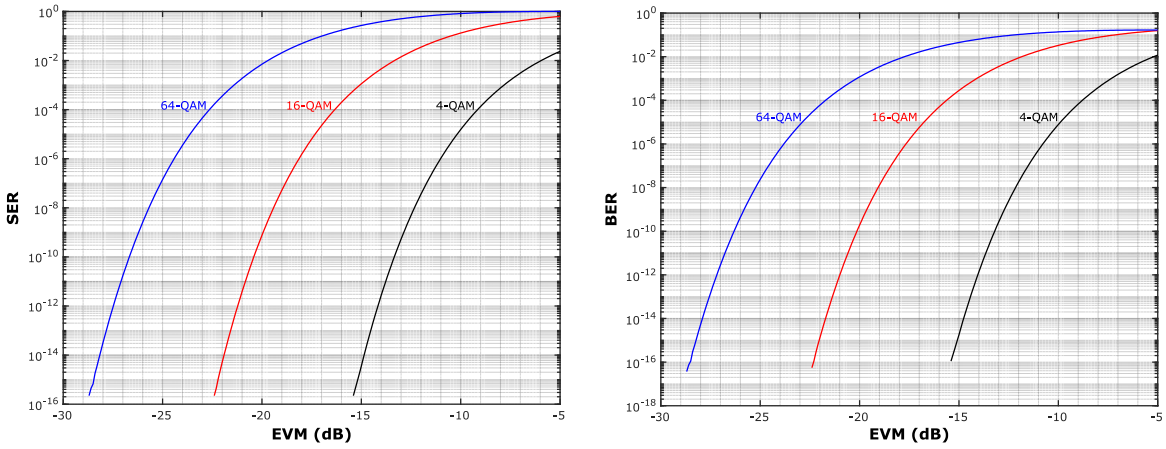
Where Q is the Gaussian co-error function:

$$Q(x) = \frac{2}{\sqrt{\pi}} \int_0^{\mathrm{x}} \mathrm{e}^{-t^2} \mathrm{dt} \tag{2.9}$$

Finally, by replacing (2.7) in (2.8) the SER can be expressed as a function of the EVM:

$$SER \approx \frac{2\left(\sqrt{\mathrm{M}}-1\right)}{\sqrt{\mathrm{M}}} Q\left(\frac{1}{EVM_{lin}} \sqrt{\frac{3}{\mathrm{M}-1}}\right) \tag{2.10}$$

Figure 2.21 (a) represents the SER as a function of EVM in dB for 4-QAM, 16-QAM, and 64-QAM from equation (2.10). As expected, the higher the modulation complexity the lower the tolerance to EVM. The three curbs are 6dB apart from each other. Because each symbol carries a given number of bits the Bit Error Rate (BER) can be calculated from the SER. It is considered that an error on a symbol occurs when it is decoded as its next neighbor symbol. Hence, one bit in coordinates is wrong.



(a) (b)

Figure 2.21 SER (a) BER (b) as a function of EVM

An N-QAM modulation carries $\log_2(M)$ bits per symbol, therefore the BER can be expressed as:

$$BER = \frac{1}{log_2(M)} SER \approx \frac{2\left(\sqrt{\mathrm{M}}-1\right)}{log_2(M)\sqrt{\mathrm{M}}} Q\left(\frac{1}{EVM_{lin}} \sqrt{\frac{3}{\mathrm{M}-1}}\right) \tag{2.11}$$

Figure 2.21 (b) represents the BER as a function of EVM. A certain number of errors are tolerated using corrector codes which are implemented by adding extra bits to the data stream and allow to check the transmitted data and correct errors. The higher the quantity of errors the more redundancy is necessary.

A practical limit to the acceptable BER is 10^{-3} which corresponds to an EVM of -20dB, -14dB, and -7.3dB for the 64-QAM, 16-QAM, and 4-QAM respectively. The gap between each modulation

requirement is large and an increase from a 4-QAM to 16-QAM or from 16-QAM to 64-QAM requires a consequent performance increase.

The achievable data rate is a product of the symbol rate that represents the signal bandwidth and the modulation complexity. An increase in symbol rate requires linear performances of the PA over a large bandwidth and degrades the EVM. Consequently, the maximally achievable data rate of a circuit is the best combination of symbol rate and modulation complexity.

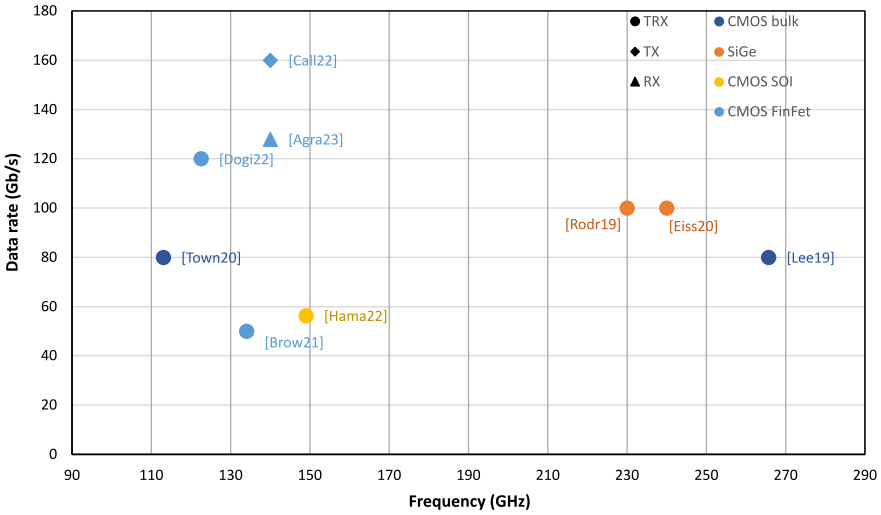


Figure 2.22 State of the art of transceivers, transmitters, and receiver over 100GHz

Figure 2.22 presents the published works on transceivers over 100GHz. The maximal data rates are represented as a function of frequency. The publications are recent, and most of them were published after the beginning of this thesis, which started in November 2019.

The two frequency bands of interest can be distinguished. As discussed before, the D-band is experimented with for high data rates and witnesses its use for possible future communication standards. There are also published transceivers in the 230GHz to 300GHz band to address the previously mentioned IEEE 802.15.3d standard.

The highest reported data rate in the D-band is 120Gb/s [Dogi22]. The architecture is presented in Figure 2.23 and is composed of two RF frontends, the first operating at 108GHz and the second at 135GHz. This way full-duplex is possible or the two bands can be used at once to extend the available bandwidth. It is composed of a classical RF frontend scheme with a two-step frequency up-conversion (down-conversion for Rx).

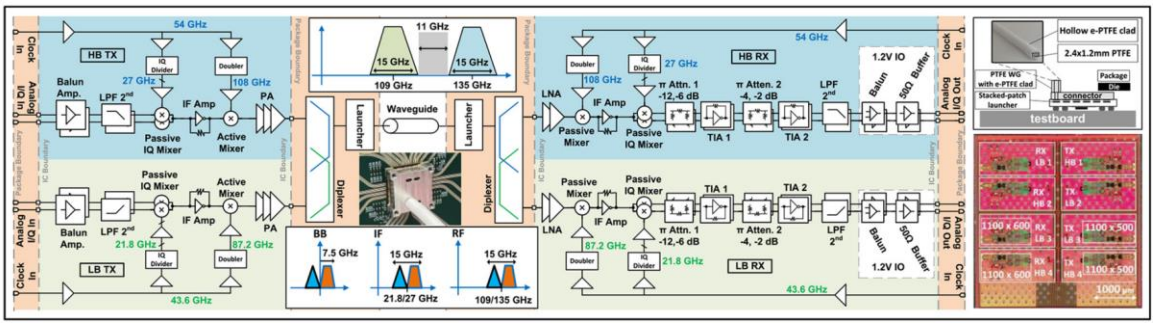


Figure 2.23 Transceiver architecture with illustrations of the frequency plan, waveguide connection, and die photograph [Dogi22]

Above 200GHz, two publications report data rates of 100Gb/s, [Rodr19] and [Eiss20]. Both are in SiGe technologies with reported f_{max} above 500GHz. The architecture of [Eiss20] is presented in Figure 2.24. It is a zero-IF architecture, which eases the overall scheme and reduces the number

of conversion steps. The high f_{max} of the technology allows the operation of the mixer and PA at the carrier frequency which presents a 12dBm $P_{sat.}$ At the Rx side, the LNA can operate at the carrier frequency as well to improve the sensitivity of the circuit.

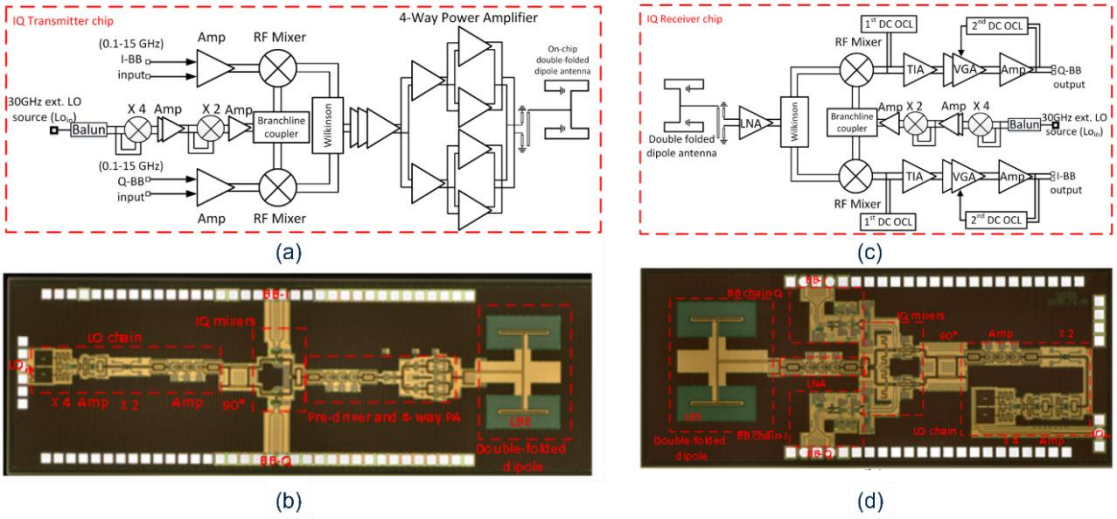


Figure 2.24 240GHz IQ transmitter schematic (a), chip microphotograph (b), and IQ receiver schematic (c) and chip microphotograph (d)

The design of transceivers is in theory not limited by the devices f_{max} and other architectures presented in the following section. For CMOS processes or less advanced SiGe processes the f_{max} is not high enough to operate between 250 and 300GHz. Innovative architectures are then necessary to operate at these carrier frequencies. [Lee19] presents a 266GHz CMOS transceiver achieving 80Gb/s. Its architecture and chip microphotograph are presented in Figure 2.25. It is composed of a square mixer that up-converts the RF signal to the second harmonic of the LO, thus the mixer must only operate at 133GHz to produce an output signal around 266GHz. The architecture uses the same principle on the Rx side. It is a type of mixer-last for Tx and mixer-first for Rx architecture that uses additional frequency multiplication in the mixer.

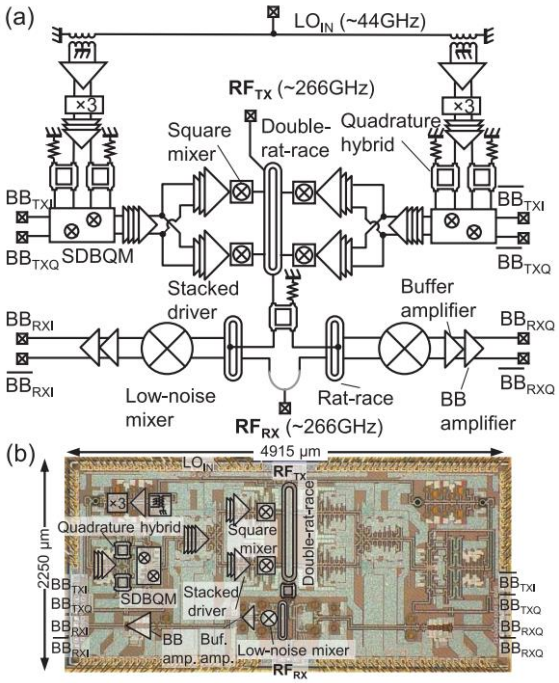


Figure 2.25 (a) Overall schematic of the 300-GHz-band CMOS TRX. SDBQM: Semidoubly balanced quadrature mixer. (b) Die micrograph [Lee19]

Figure 2.22 reports in addition to the transceivers a Tx [Call22] and an Rx [Agra23] at D-Band. The two circuits are presented in Figures 2.26 and 2.27 and are present in the SOA comparison because of their promising performances. The Tx can achieve 160Gb/s data rate while the Rx 128Gb/s. The two circuits are complementary to operate a link at 140GHz and will possibly demonstrate a D-Band wireless link with a data rate at or above 120Gb/s.

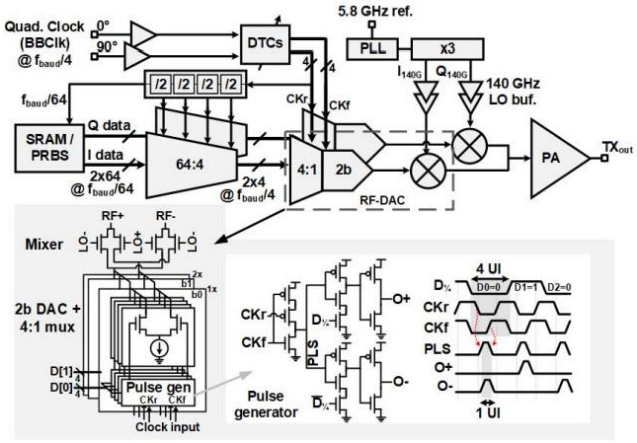


Figure 2.26 Top-level architecture of transmitter showing details of RF-DAC with 4:1 MUX [Call22]

In addition, the two circuits integrate the analog-to-digital converter (ADC) and digital-to-analog converter (DAC) which are essential but are also limiting factors regarding the data rates in communication systems. These circuits are necessary to enable such high data rates and these publications present data rates at the state of the art while integrating them.

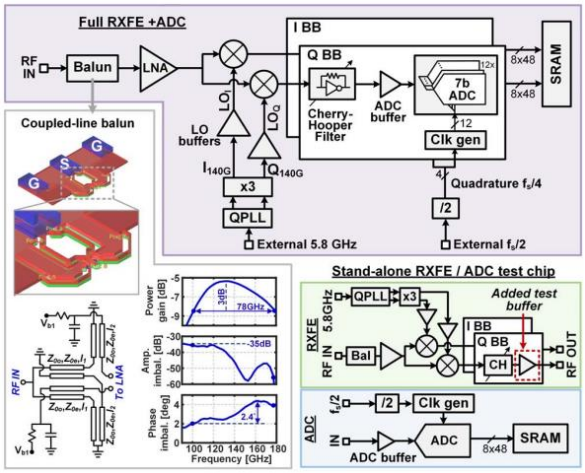


Figure 2.27: High-level architecture of the fully integrated receiver (top) and implementation of coupled-line-based Guanella balun (bottom left). Configuration of test chip for stand-alone characterization of receiver front-end (RXFE) and ADC with sampling rate fs (bottom right) [Agra23]

All of the reported data rates from Figure 2.22 are achieved for 16QAM modulations which suggest a bottleneck of spectral efficiency at these frequencies. This limitation is due to multiple factors, such as non-linearities over power amplitudes, group delays over the large used bandwidths, gain variations, and other phenomenons that distort the signals and thus prevent the use of higher-order modulations with acceptable EVMs. But it also represents an area of improvement that will allow in the future a multiple-time increase in the data rates. Figure 2.28 depicts the reported EVM measurements of the transceivers from Figure 2.22 as a function of their data rates. The reported EVM are in the -14dB to -20dB range which corresponds to the window for 16-QAM modulation with a BER below 10^{-3}.

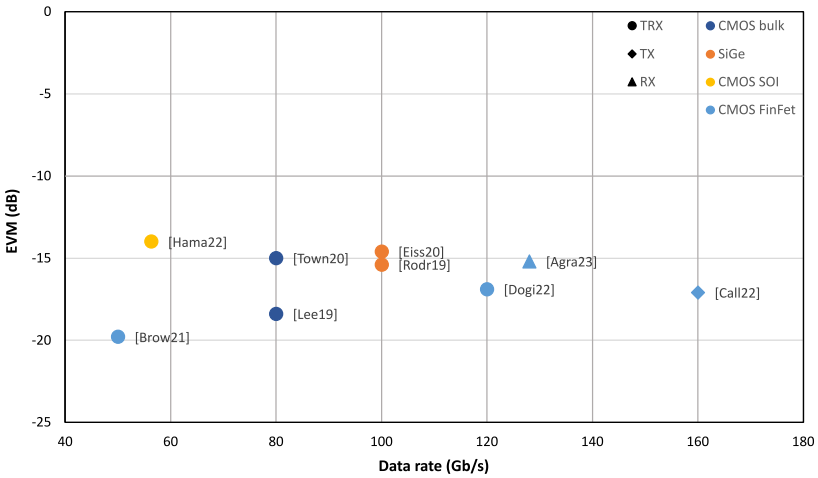


Figure 2.28 EVM as a function of data rate for transceivers over 100GHz

The EVM of a transceiver comes from the quality of the Tx and the Rx. The design of the LNA is crucial in the Rx as it must operate at the carrier frequency, present a low NF, and have good linearity. On the Tx side, the PA must simultaneously present a high output power, a high gain, good linearity with good efficiency, and over a large frequency bandwidth. All these metrics are directly limiting the quality of the emitted signal and consequently, determine the maximum achievable data rate.

2.5 Power Amplifiers over 100GHz

This section focuses on the published PAs over 100GHz in silicon technologies to highlight the challenges that need to be addressed. Figure 2.29 presents the demonstrations of PAs published up to the end of 2019 and represents the state of the art at the beginning of this thesis. The main metrics of PAs are presented in Appendix A.

The saturated power of the PAs is represented as a function of their central frequency of operation. There is a clear decrease in P_{sat} over frequency, between 110GHz to 140GHz, multiple works reported P_{sat} around 20dBm in SiGe technologies and around 15dBm in CMOS bulk nodes. Then around 200GHz and above, one work reports a P_{sat} of 4dBm and the other CMOS nodes power levels below 0dBm. SiGe technologies show a clear advantage over CMOS because of their higher supply voltage and f_{max}, which were reported up to 650GHz recently.

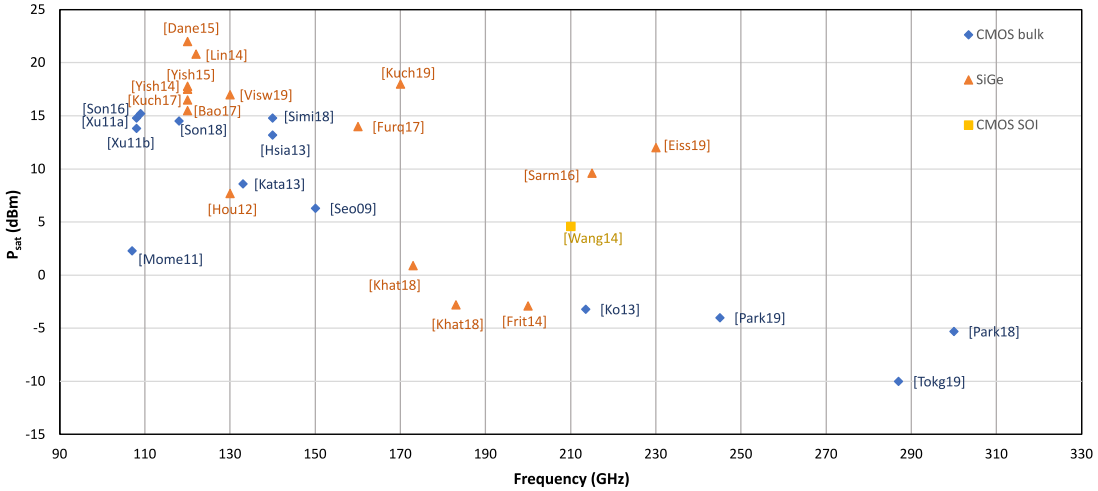


Figure 2.29 Representation of published PA before 2020's P_{sat} as a function of frequency

In the same way as the transceivers, the rate of publication of new PAs over 100GHz greatly increased from 2020. Figure 2.30 depicts the actual state of the art of published PAs, a table with all the cited PAs is present in Appendix B. The gap between SiGe and CMOS nodes reduces below 210GHz regarding the P_{sat} but keeps expanding at higher frequencies. The last improvements in dedicated SiGe devices pushed the boundaries of possible at even higher frequencies. For the CMOS publications, the development and improvement of design techniques allowed to improve the performance.

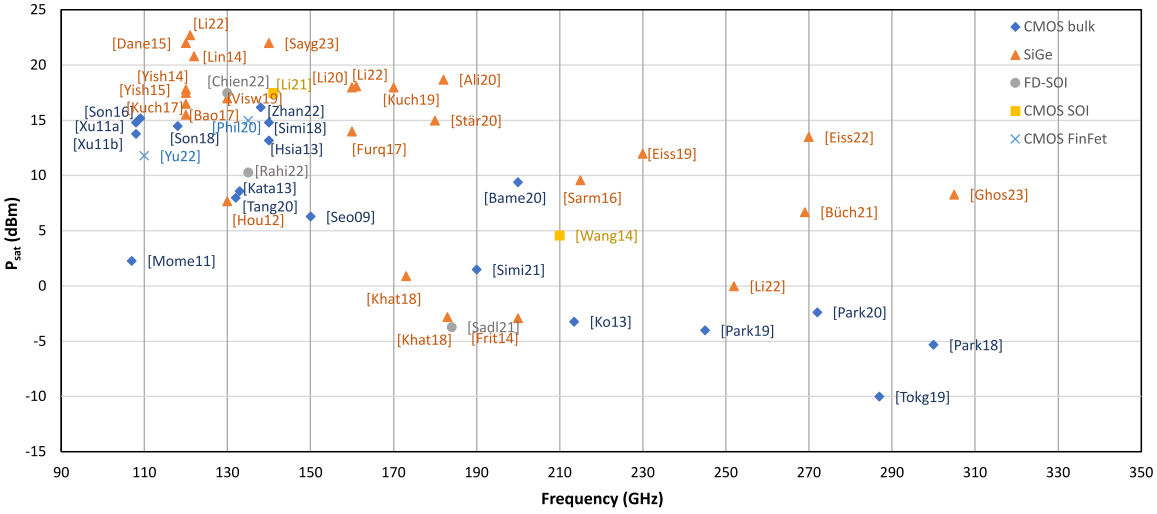


Figure 2.30 Representation of all published PA's P_{sat} as a function of frequency

Figure 2.31 presents the PA from [Eiss22], which has a P_{sat} of 13.5 dBm over a large bandwidth between 220GHz and 320GHz. The power combining of four parallel differential amplifier stages allows to achieve this high output power for this frequency. The use of Cascode stages with a 3V supply also contributes to the performance.

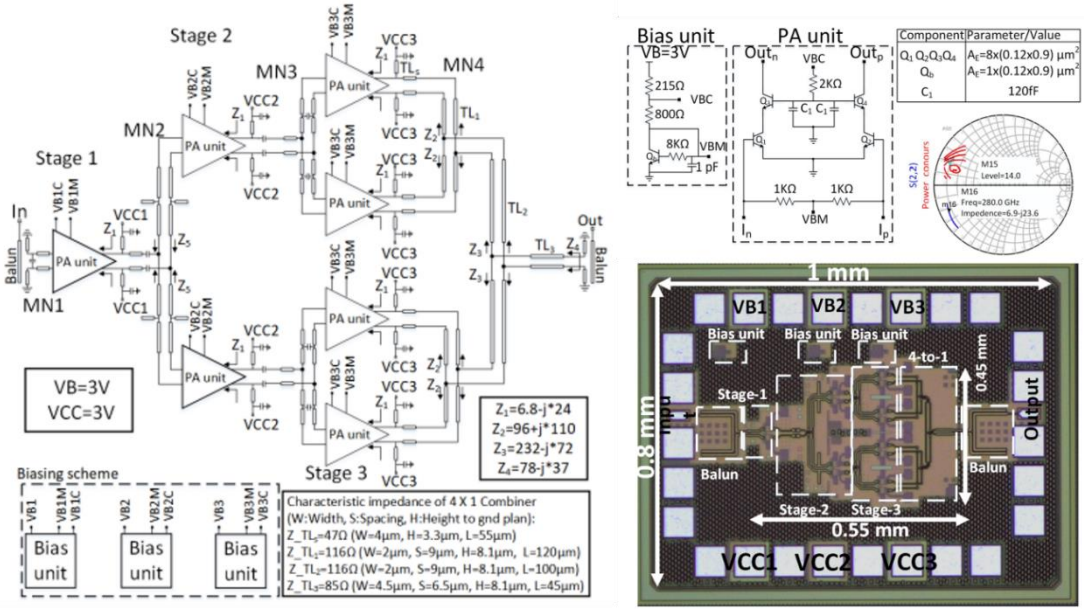


Figure 2.31 4-way combined PA top-level schematic and die photograph [Eiss22]

Another important metric is the power efficiency of the power amplifier. The PA represents the largest power consumption among the RF front-end circuits. Therefore, an efficient PA determines the power efficiency of a transceiver with a low power consumption per transmitted bit of data. The output power possibility of components decreases rapidly with frequency as larger devices present a high capacitive impedance component that needs to be resonated with inductive impedances. Parallel power combining and device stacking are key to obtaining high power levels. However, the passive component efficiency tends to decrease at higher frequencies which ends in poor power efficiency of PAs. Consequently, power combination is implemented at the cost of power efficiency.

Figure 2.32 reports the PAE of the published PA over 100GHz. The drop in efficiency is consequent with frequency.

It is a bottleneck that design techniques cannot alleviate. The maximal reported PAE is below 15% at 140GHz and below 10% above 160GHz. It decreases to a couple of percent above 200GHz. The PAE is the image of the efficiency of the last passive components from the circuits and the total power consumption.

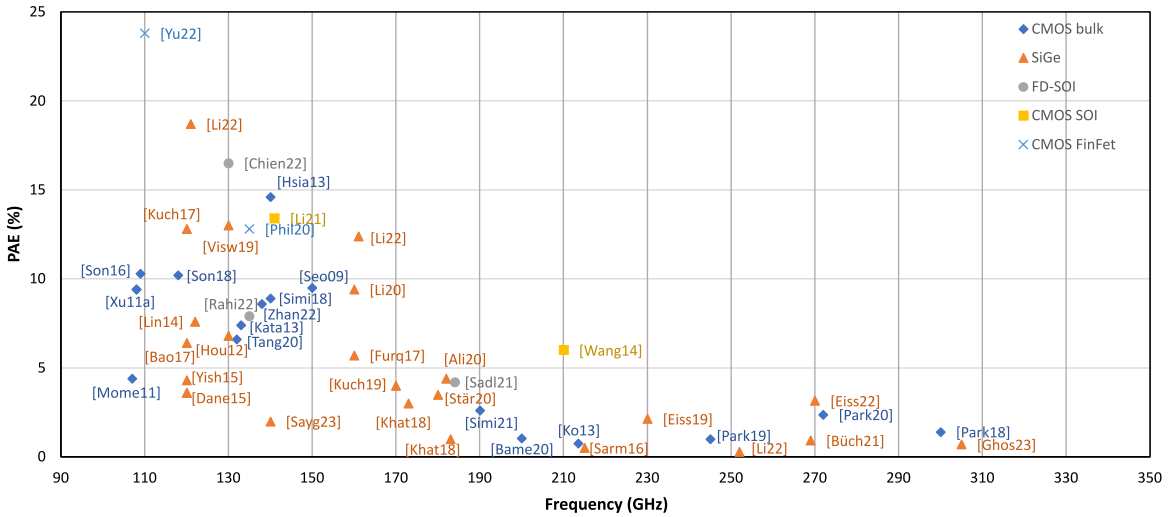


Figure 2.32 Representation of all published PA's PAE as a function of frequency

To transmit high data rates, the transmitter and consequently, the PA must operate over large bandwidths. The bandwidth of a PA is limited by its matching networks as power combiners and splitters also operate impedance transformations. The operation over large bandwidths is difficult for large power combinations. In addition, each additional inter-stage impedance matching narrows the PA bandwidth. It is very challenging to operate in a broadband manner with high P_{sat} and gain. SiGe technologies present far higher bandwidths than CMOS publications and this is due to their higher gain budget. With a higher $f_{max,}$ the interstage matching can be designed for broadband operation at a cost of less gain. For CMOS circuits, the gain per device is limited and each matching network must be efficient, which makes the broadband operation difficult.

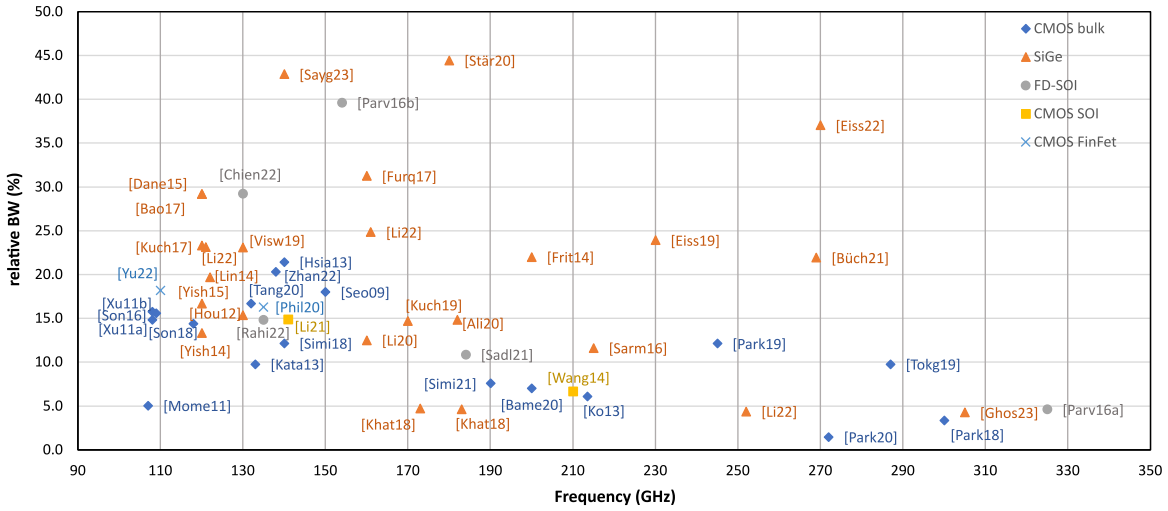


Figure 2.33 Representation of all published PA's relative bandwidth as a function of frequency

To achieve sufficient gain a certain number of amplification stages need to be cascaded in the PA. This quantity is determined by the gain per stage that can be achieved as the efficiency and bandwidth decrease when the number of stages increases.

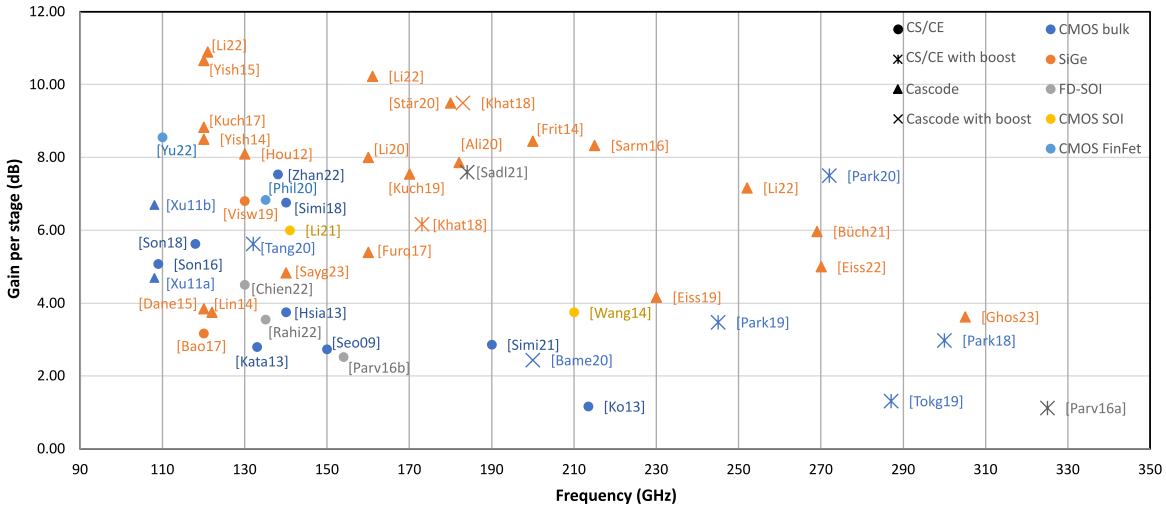


Figure 2.34 Representation of all published PA's gain per stage as a function of frequency

Figure 2.34 exhibits the natural trend of the devices regarding the available gain. The visible decrease with frequency is linear in all technologies. Common Source (CS) and Common Emitters (CE), in CMOS and SiGe, respectively use one transistor, and the decreasing trend with the frequency of the gain per stage is visible. The SiGe publications largely use Cascode configurations that use two transistors, with a final higher gain per stage. The advantage of the higher f_{max} is clear in this graphic. Cascode configurations are less represented in CMOS nodes. Among the classic CS/CE and Cascode topologies, circuits in CMOS nodes essentially started to implement gain-boosting techniques to alleviate the lower device performances. Therefore, all CMOS circuits above 205GHz employ gain boosting.

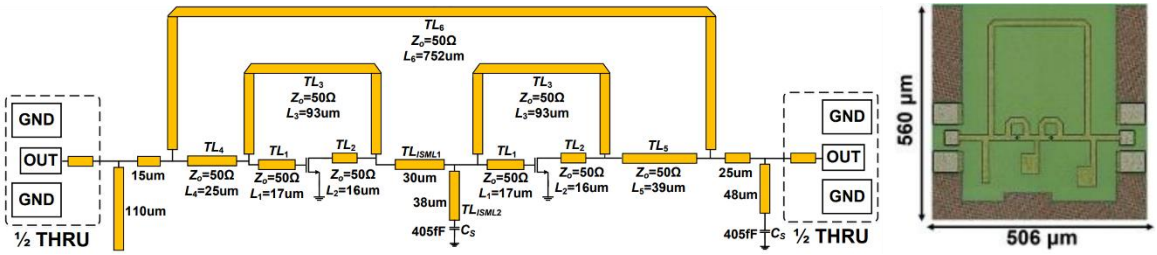


Figure 2.35 Schematic and chip microphotograph of 280GHz amplifier [Park20]

The gain boosting principle is to create positive feedback around an amplifier to increase the available gain while assuring the stability of the amplifier. The principle is usually applied at a single frequency and the increase in gain is a very narrow band. The 280GHz amplifier reported in [Park20] depicted in Figure 2.35, presents a power gain of 15dB with 7.5dB per stage. The P_{sat} is only -2.36dBm but the real limitation is the bandwidth, as the cited PA has a 3-dB bandwidth of 4GHz which represents only 1.5% of relative bandwidth. The additional particularity of this PA is the double use of gain boosting. One first time around each CS stage then around two consecutive boosted stages. It explains the very high gain for the frequency. However, the design is not robust to modeling inaccuracies or production variation due to the nature of the design and the narrow bandwidth. Other publications above 200GHz in CMOS, [Park19] and [Tokg19] employ gain boosting only around single stages and have lower gain per stage but a higher bandwidth.

The gain boosting is necessary for CMOS operations above 180GHz as the devices are not performant enough. The bandwidth of such techniques is limited and the P_{sat} is as well. The PA presented in [Bame20] aims to address this challenge. The amplifier is presented in Figure 2.36,

it consists of two parallel chains of 8 boosted cascode stages. The power splitter and combiner allow doubling the output power. Each Cascode stage has an internal matching, positive feedback, and a λ/4 transmission line to supply the circuit.

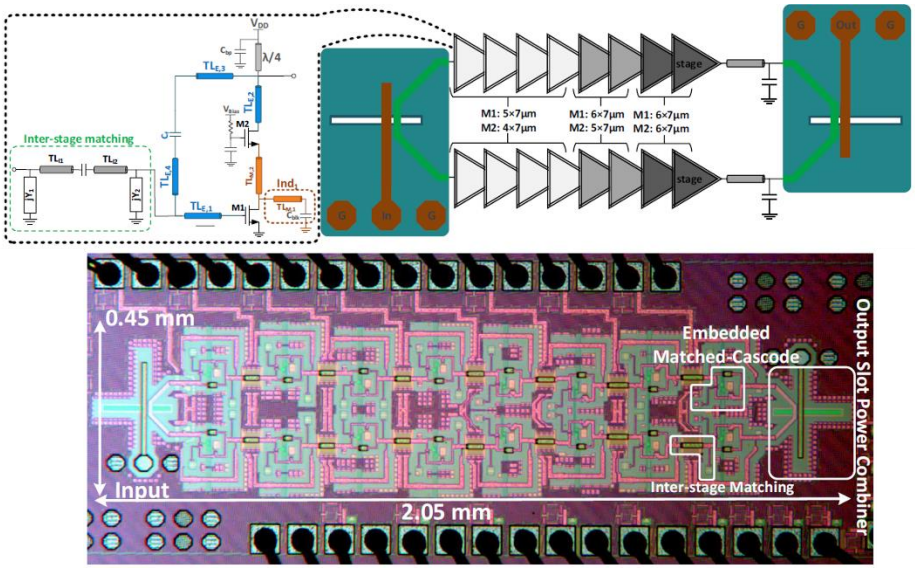


Figure 2.36 Block diagram of 200GHz PA and chip microphotograph [Bame20]

The PA achieves a P_{sat} of 9.4dBm which is 5dB higher than the second best reported P_{sat} in CMOS technologies around 200GHz. It has a gain of 19.5dB which represents 2.4dB of gain per boosted cascode stage. The bandwidth is 14GHz or 7% relative bandwidth. Even with boosted cascode topologies which are optimal for gain performance, the use of relatively large transistors to produce a high output power level requires lossy matching networks, hence only 2.4dB per stage. The circuit maximal PAE is about 1% and is due to the large number of amplification stages and the poor passive component efficiency.

2.6 Conclusion on actual limits in transmission circuits

This chapter presented the application opportunities of the sub-THz and THz spectrum. The demand for faster data rates on wireless communications was discussed and the solution to use the sub-THz band was presented. In light of the technical challenges that must be overcome to design communication transceivers, the state of the art of the recently published demonstrator was analyzed. The key challenge in the transmission part of the wireless link is a performant Power Amplifier that dictates the final emitter performance. The following of this manuscript focuses on the key metrics that have been highlighted in the light of the state of the art. The Gain per stage of the amplifier is a consequence of the transistor performances, gain boosting techniques must be employed to dispose of enough gain to limit the number of stages, keep a large operation bandwidth, and generate a high output power level.

- The next chapter 3 focuses on the active and passive devices study to get the best possible performance from an integrated CMOS process.
- Chapter 4 focuses on the gain-boosting technic theory and proposes a design methodology. The design of an 184GHz amplifier is demonstrated and measurements are presented.
- Chapter 5 aims to study and propose design techniques to extend the gain boosting over large bandwidths, design large bandwidth and efficient impedance matching networks, and

increase the P_{sat} of the PA. Multiple 200GHz amplifier designs are presented along with the fabricated circuits' measurements.

2.7 References

[Rapp19] T. S. Rappaport et al., "Wireless Communications and Applications Above 100 GHz: Opportunities and Challenges for 6G and Beyond," in IEEE Access, 2019

[Crow05] C. Cheng, "Neutralization and Unilateralization," in IRE Transactions on Circuit Theory, June 1955

[Hadi11] R. Al Hadi et al., "A broadband 0.6 to 1 THz CMOS imaging detector with an integrated lens," 2011 IEEE MTT-S International Microwave Symposium, Baltimore, MD, USA, 2011

[Pfei19] U. Pfeiffer, " Integrated Circuit Design for Terahertz Applications," 6G Wireless Summit, 2019

[Cheo17] H. Cheon et al., "Toward Clinical Cancer Imaging Using Terahertz Spectroscopy", IEEE Journal of selected topics in quantum electronics, 2017

[Wang18a] C. Wang and R. Han, "Rapid and Energy-Efficient Molecular Sensing Using Dual mm-Wave Combs in 65nm CMOS: A 220-to-320GHz Spectrometer with 5.2mW Radiated Power and 14.6-to-19.5dB Noise Figure," in International Solid-State Circuit Conference (ISSCC), 2017

[Wang18b] C. Wang et al., "Chip-Scale Molecular Clock," in IEEE Journal of Solid-State Circuits, 2019

[Hoss21] S. M. Hossein Naghavi, et al., "22.4 A 250GHz Autodyne FMCW Radar in 55nm BiCMOS with Micrometer Range Resolution," 2021 IEEE International Solid-State Circuits Conference (ISSCC), 2021

[Fuji18] M. Fujishima, "Key Technologies for THz Wireless Link by Silicon CMOS Integrated Circuits" Photonics, 2018

[Okad22] K. Okada, "mmWaveRF frontend designs over 100GHz for 6G wireless communication," RFIC Workshop, 2022

[Shah22] S. Shahramian et al., "Practical Approaches to Industrializing Near-THz Communication Systems", RFIC Workshop, 2022

[Xing18] Y. Xing and T. S. Rappaport, "Propagation Measurement System and Approach at 140 GHz-Moving to 6G and Above 100 GHz," IEEE Global Communications Conference (GLOBECOM), 2018

[Petro20] V. Petrov, et al., "IEEE 802.15.3d: First Standardization Efforts for Sub-Terahertz Band Communications toward 6G," in IEEE Communications Magazine, 2020

[IEEE17] IEEE Std 802.15.3d–2017, IEEE Standard for High Data Rate Wireless Multi-Media Networks–Amendment 2: 100 Gb/s Wireless Switched Point-to-Point Physical Layer, 2017

[Shaf06] R A Shafik et al., "On the Extended Relationships Among EVM, BER and SNR as Performance Metrics", ICECE 2006

[Gold05] A Goldsmith, "WIRELESS COMMUNICATIONS", Cambridge University Press, 2008, p. 178

[Dogi22] G C Dogiamis et al., "A 120-Gb/s 100-145-GHz 16-QAM Dual-Band Dielectric Waveguide Interconnect With Package Integrated Diplexers in Intel 16", SSCL, 2022

[Rodr19] P Rodríguez-Vázquez et al., “A 16-QAM 100-Gb/s 1-M Wireless Link With an EVM of 17% at 230 GHz in an SiGe Technology”, MWCL, 2019

[Eiss20] M H Eissa et al., “100 Gbps 0.8-m Wireless Link based on Fully Integrated 240 GHz IQ Transmitter and Receiver”, IMS, 2020

[Lee19] S Lee et al., “An 80-Gb/s 300-GHz-Band Single-Chip CMOS Transceiver”, ISSCC, 2019

[Call22] S Callender et al., “A Fully Integrated 160-Gb/s D-Band Transmitter Achieving 1.1-pJ/b Efficiency in 22-nm FinFET”, ISSCC, 2020

[Agra23] A Agrawal et al., “A 128Gb/s 1.95pJ/b D-Band Receiver with Integrated PLL and ADC in 22nm FinFET”, ISSCC, 2023

[Eiss22] M H Eissa et al., “220–320-GHz J -Band 4-Way Power Amplifier in Advanced 130-nm BiCMOS Technology”, MWCL, 2022

[Park20] D-W Park et al., “A 247 and 272 GHz Two-Stage Regenerative Amplifiers in 65 nm CMOS with 18 and 15 dB Gain Based on Double-Gmax Gain Boosting Technique”, VLSI, 2020

[Park19] D-W Park et al., “A 230–260-GHz wideband and high-gain amplifier in 65-nm CMOS based on dual-peak Gmax-core”, JSSCC, 2019

[Bame20] H Bameri et al., “An Embedded 200 GHz Power Amplifier with 9.4 dBm Saturated Power and 19.5 dB Gain in 65nm CMOS”, RFIC, 2020

Chapter 3 IC design considerations for mmWave/sub-THz bands

Radiofrequency (RF) engineering is the science of interacting with the electromagnetic field surrounding us, to sense or bend it according to our will. This interaction is exploited to transmit and receive information through the field and measure the physical properties of materials or physical phenomena.

Electromagnetic waves can be initiated or sensed and processed with electrical circuits. The analysis of these circuits is done by making a quasistatic approximation of the Maxwell equations and locally representing the physical phenomenon by R, L, C, and M components. This simplification made by the equivalent circuit is valid up to a certain frequency and additional representations with distributed elements based on transmission lines capture the higher frequency effects.

The transistor is the main component that enables the RF functions; however, it becomes less and less efficient as frequency increases up to the point at which it cannot operate anymore.

This chapter presents the components that are used in Integrated Circuit (IC) technologies to implement RF functions.

3.1 Network analysis

The analysis of electrical systems consists of finding relations between currents and voltages at different points of the system. They can then be used to estimate the transfer functions from one point to another. Circuits can be composed of discrete components or matrixes that represent complex networks.

A Two-Port Network is a matrix representation of a linear two-terminal component. It is the most used Linear Component representation as it can be used for one input and one output terminal. More details about general Linear N-Ports components are recalled in Appendix C.

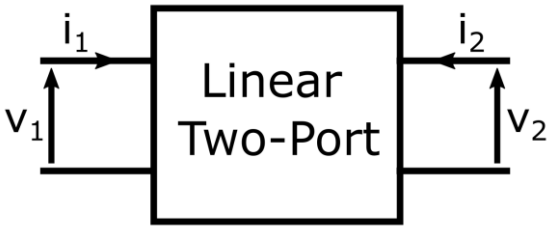


Figure 3.1 Schematic of a linear two-port

The relations between the voltages and currents are given with the impedance parameters Z_{ij}:

$$\begin{cases} V_1 = Z_{11} I_1 + Z_{12} I_2 \\ V_2 = Z_{21} I_1 + Z_{22} I_2 \end{cases} \quad (3.1)$$

We can define impedances at each port when all other ports are ended by a load impedance.

$$Z_1 = \frac{V_1}{I_1} = Z_{11} - \frac{Z_{21}}{Z_{22} + Z_{Load,2}} Z_{12} \quad (3.2)$$

Inversely,

$$Z_2 = \frac{V_2}{I_2} = Z_{22} - \frac{Z_{21}}{Z_{11} + Z_{Load,1}} Z_{12} \quad (3.3)$$

For an admittance representation, the port admittances can be expressed as,

$$Y_1 = \frac{I_1}{V_1} = \frac{1}{Z_1} = Y_{11} - \frac{Y_{21}}{Y_{22} + Y_{Load,2}} Y_{12} \quad (3.4)$$

$$Y_2 = \frac{I_2}{V_2} = \frac{1}{Z_2} = Y_{22} - \frac{Y_{21}}{Y_{11} + Y_{Load,1}} Y_{12} \quad (3.5)$$

Gain definitions

We want to calculate the maximal power gain that can generate a Linear Two-Port under optimal conditions. Thus, the research of these optimal conditions is a prerequisite to obtaining maximal performance from a circuit.

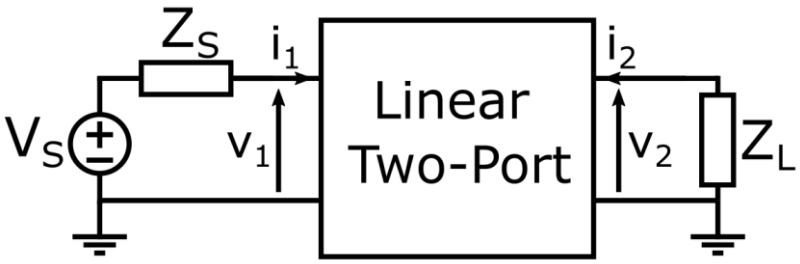


Figure 3.2 Schematic of a linear two port with one port connected to a voltage source and the other connected to a load

From a representation of an LTP by a black box matrix, and knowing the source and load impedances, respectively Z_S and Z_L, it is possible to calculate the power gain of the circuit.
Multiple powers can be defined, either the maximal power that a source can generate, when the port's impedance is matched, or the received power by a load:
-the available power flowing out of the source,

$$P_{av,S} = \frac{|V_S|^2}{4ReZ_S} = \frac{|V_S|^2}{4} Re(Y_S) \quad (3.6)$$

-the power entering the LTP,

$$P_{in} = \frac{|V_1|^2}{4ReZ_1} = \frac{|V_1|^2}{4} Re(Y_1) \quad (3.7)$$

-the available power available for the load, flowing out of the LTP,

$$P_{av,L} = \frac{|V_2|^2}{4ReZ_2} = \frac{|V_2|^2}{4} Re(Y_2) \quad (3.8)$$

-the power received by the load,

$$P_L = \frac{|V_2|^2}{4ReZ_L} = \frac{|V_2|^2}{4} Re(Y_L) \quad (3.9)$$

Given the distinctions between the defined powers, multiple power gains can be defined as well:
-the power gain G_p, which is the gain between the power received by the LTP and the power received by the load,

$$G_p = \frac{P_L}{P_{in}} = f(Z_{ij}, Z_L) \quad (3.10)$$

Thus, it depends on the LTP parameters and the load impedance but is independent of the source impedance Z_S.
-the available gain G_a of the LTP, which is the power gain between the available power at the source $P_{av,s}$ and the output of the LTP $P_{av,L}$.

$$G_a = \frac{P_{av,L}}{P_{av,L}} = f(Z_{ij}, Z_S) \tag{3.11}$$

This time the gain depends on Z_S and is independent of Z_L.

-Finally, the transconductance gain G_t which considers the power received by the load P_L and the power available at the source $P_{av,S}$

$$G_T = \frac{P_L}{P_{av,S}} = f(Z_{ij}, Z_S, Z_L) \tag{3.12}$$

This metric considers both reflections, at the input and output of the LTP.

These gains can be related to the S-parameters of the LTP,

$$G_T = S_{21} \tag{3.13}$$

$$G_p = \frac{S_{21}}{1 - S_{11}} \tag{3.14}$$

$$G_a = \frac{S_{21}}{1 - S_{22}} \tag{3.15}$$

Optimal Matching and Maximum Available Gain

When ports are optimally loaded with conjugate impedances, $Z_S = Z_1^*$ and $Z_L = Z_2^*$, there are no reflections, and the three gains are equal. This maximal gain value under optimal conditions is named the maximally available gain G_{ma}.

Finally, the gains can be expressed in terms of Y parameters,

$$G_T = \frac{4Re\{Y_L\}Re\{Y_S\}|Y_{21}|^2}{|(Y_S + Y_{11})(Y_{22} + Y_L) - Y_{12}Y_{21}|^2} \tag{3.16}$$

$$G_p = \left|\frac{Y_{21}}{Y_{22} + Y_L}\right|^2 \frac{Re\{Y_L\}}{Re\{Y_1\}} \tag{3.17}$$

$$G_a = \left|\frac{Y_{21}}{Y_{11} + Y_S}\right|^2 \frac{Re\{Y_S\}}{Re\{Y_2\}} \tag{3.18}$$

By deriving the expression of G_p or G_a relative to Y_l and Y_S respectively, optimal $Y_{L,opt,}$ and $Y_{S,opt}$ can be derived in terms of Y parameters and thus can be expressed G_{ma}.

This analysis gives:

$$Y_{L,opt} = \frac{|Y_{21}Y_{12}|\left(K + \sqrt{K^2 - 1}\right) + Y_{21}Y_{12}}{2ReY_{11}} - Y_{22} \tag{3.19}$$

$$Y_{S,opt} = \frac{|Y_{21}Y_{12}|\left(K + \sqrt{K^2 - 1}\right) + Y_{21}Y_{12}}{2ReY_{22}} - Y_{11} \tag{3.20}$$

$$G_{ma} = \left|\frac{Y_{21}}{Y_{12}}\right|\left(K - \sqrt{K^2 - 1}\right) \tag{3.21}$$

With,

$$K = \frac{2Re\{Y_{11}\}Re\{Y_{22}\} - Re\{Y_{21}Y_{12}\}}{|Y_{21}Y_{12}|} \tag{3.22}$$

These expressions are defined for $K \geq 1$. When $K < 1$, the square root gives a complex value while it represents a power gain, thus a real quantity.

When these expressions are defined, they can give us instantly the value of G_{ma} and optimal impedances (or admittances) to present to the two ports, only using its Y parameters representation (or any other representation, given their equivalence).

3.1.1 Stability of a linear two-port

The stability of a circuit is crucial as it is an undesired behavior that alters or prevents a circuit from doing the operation that it was designed for. It is characterized by the ability of the circuit to operate without any signal input. While the source of instability can be complex and appear with the non-linear behavior of the circuit, it is necessary to ensure that in the linear approximation, it is stable. The study of linear two-port parameters can indicate if some of the conditions for an oscillation are present or not. A circuit is an oscillator if it generates a signal, hence the current flows out of it. It is characterized by a negative impedance real part:

$$real(Z_{in}) < 0 \tag{3.23}$$

The input impedance of an LTP is a function of the load impedance, hence the real part also depends on the load that is presented to the LTP.

Unconditional stability is therefore characterized by a real part of the impedance being positive for any load impedance. By expressing the input impedance in terms of Z parameters and as a function of the load impedance Z_L, it is possible to find the minimum of this function. This minimal real part impedance is strictly positive if and only if:

$$K = \frac{2Re\{Z_{11}\}Re\{Z_{22}\} - Re\{Z_{21}Z_{12}\}}{|Z_{21}Z_{12}|} > 1 \tag{3.24}$$

Which is the same quantity found in the optimal source and load impedances and G_{ma} expressions. When this quantity is negative, there is at least one impedance that produces a negative real part of the input or output impedance of the linear two port. Hence, it produces a signal without any input, which represents an infinite gain. Therefore, when this condition on K is not met, there is no gain boundary.

To ensure that an LTP cannot generate an oscillation under any circumstances, K must be superior to 1.

The stability of a complex circuit cannot be ensured with only an analysis of its external input and output ports as an internal node can be responsible for instability. K factor must be used on simple circuit parts that cannot be broken into smaller LTPs, then if all elementary blocks are unconditionally stable, the complex circuit that is an assembly is consequently unconditionally stable. The condition on K is necessary but not sufficient for complex structures.

In addition, a circuit that is not unconditionally stable can still operate with source and load impedances that do not produce negative impedances. In this case, simultaneous conjugate matching is not possible, and the environment of the circuit must be verified with attention.

3.2 The scaling of transistors through the development of integrated circuit technologies

Since the invention of the transistor in 1947 by John Bardeen, Walter Brattain and William Shockley, the semiconductor industry kept growing exponentially. As the inventors started to figure out applications for the new device and used it in telecommunication circuits in 1953, other companies all around the world leveraged this invention as well to produce transistor-based radio-circuits. The first mass-produced transistor radio was launched 10 years later by Sony in 1957.

The development of the first MOSFET devices in 1959 on silicon wafers started the race to exponentially increase the number of transistors in integrated circuits. Thanks to the rapid IC process development, a trend appeared regarding the increase in transistor count. In 1965 Gordon Moore predicted a doubling every year of the number of transistors that can fit in the same

silicon area [Moor65]. The prediction came true for the next decades and Moore's prediction became Moore's law and drove the development of the industry roadmap since then (Figure 3.3).

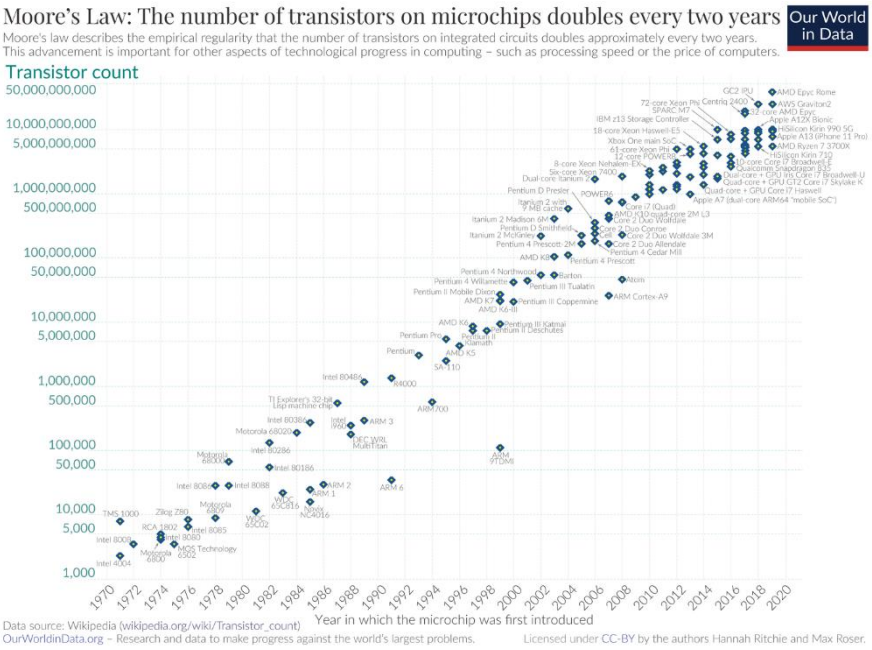


Figure 3.3 Number of transistors per chip over the years [Wiki1]

This scaling of the transistor size allowed for a substantial reduction in the cost of the circuits and an increase in the complexity of the integrated functions. The focus of CMOS technologies was on the integration of digital circuits to make logical operations. As the nodes scaled down, the electrical properties of CMOS transistors improved, and their maximal frequency of operation increased as well.

Some publications from 1970 to 1990 studied the first implementations of RF functions in IC technologies but the interest in RF CMOS grew after 1990. While the RF CMOS research attracted the curiosity of some engineers, the opinion remained divided about the future of RF CMOS. Until then III-V based, and Bipolar technologies dominated the RF industry, and the paradigm wasn't ready to change.

"RF is a solved problem. And using an inferior technology like CMOS to solve it yet again is stupid-squared."
– Unnamed MIT professor, c. 1986.

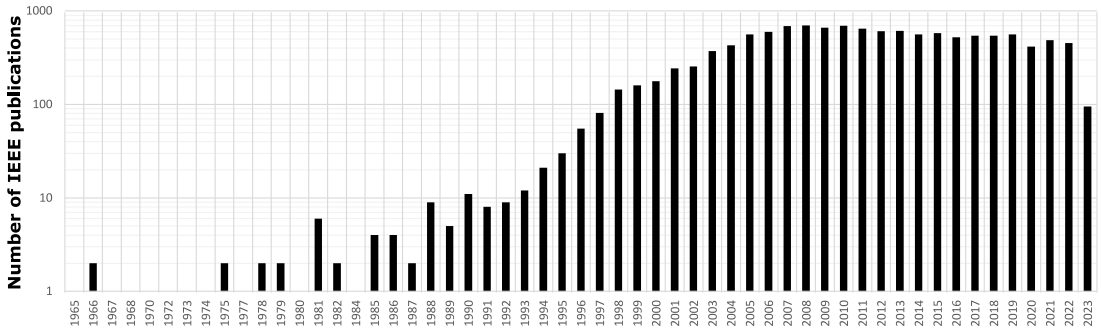


Figure 3.4 Number of results on IEEE Explore for the search "RF"AND("MOS"OR"CMOS") per year of publication.

The interest in the field finally grew (Figure 3.4) and the industry adopted CMOS technology to produce RF products for the mass market. With such versatile technologies, the industry worked

on integrating a complete system on a single chip including the RF functions. This high integration was achieved at the beginning of the 2000s offering low consumption, low cost, Wi-Fi, Bluetooth, or GSM SoCs for mass consumer-connected devices.

To make performant circuits the designer needs performant transistors. The performances that can be achieved by a circuit are intrinsically limited by the component's characteristics. Especially as the frequency of operation increases. The transistor is fundamentally limited by the materials it is made of. In 1965, Edward O. Johnson [John65] presented his analysis of the ultimate performance of a transistor. Independently of the device design details he wanted to find a physical boundary. He concluded that the product of the maximal voltage v_m that can be applied across the transistor and the current gain cut-off frequency f_t is intrinsically linked to the material properties as:

$$v_m f_t = \frac{E_B V_{SAT}}{2\pi} \tag{3.25}$$

Where E_B is the dielectric breakdown voltage and V_{SAT} is the saturated drift velocity of the charges. Figure 3.5 gives the main parameters of materials used to make transistors.

As an example, a silicon transistor with a maximal voltage of 1 V has a limit of 382 GHz. In addition, f_t can be further increased by scaling down the transistor length at the coast of a lower v_m.

Material	Bandgap (eV)	Mobility (cm^2/Vs)	ε_r	E_B (V/cm)	V_{SAT} (cm/s)	K (W/cm-K)
n-SiC (4H)	3.26	300	9.66	2.2×10^6	2×10^7	3.03.8
n-GaN	3.39	1500	9	3×10^6	2.5×10^7	2.2
n-Si	1.12	1400	11.7	3×10^5	0.8×10^7	1.3
n-GaAs	1.4	5000	13.1	3×10^5	0.8×10^7	0.5
n-InP	1.35	4500	12.4	5×10^5	1×10^7	0.68
n-Ge	0.66	3900	15.8	2×10^5	0.6×10^7	0.58
In0.53Ga0.47As	0.78	11000	13.9	2×10^5	0.8×10^7	0.05

Figure 3.5 Properties of the main semiconductor materials used in high-frequency transistors [Voin13]

The Johnson Limit is a physical figure of merit and does not reflect the new complex structure performances. More advanced transistor topologies were developed using multiple materials, such as InP HBTs [Urte17] or High Electron Mobility Transistors (HEMT) [Asif93].

Figure 3.6 depicts the state of the art of technologies regarding the f_t and f_{max}.

InP-based technologies have been developed with the only target to maximize these metrics. The most advanced process achieves an impressive 1.5 THz f_{max}, but they are still experimental technologies. SiGe follows and represents the VLSI technologies with the highest potential regarding RF performances with f_t/f_{max} of 750/500 GHz and are promising for large-scale high-frequency ICs.

However, the scaling and R&D of CMOS transistors pushed the cut-off frequencies to about 400 GHz. Thus, CMOS technologies are competitive at mmWave frequencies, and their versatility makes them a perfect candidate for the integration of entire systems on chip (SOC).

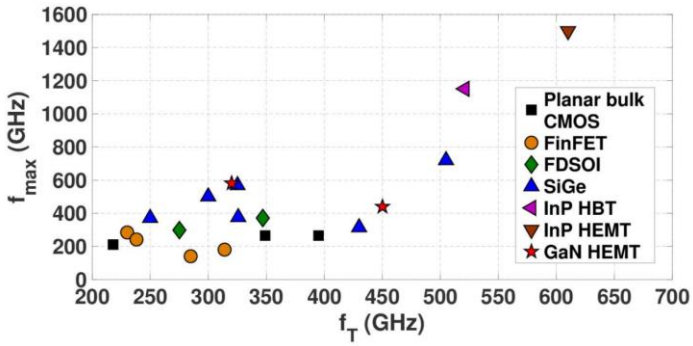


Figure 3.6 f_t and f_{max} performances of different transistor technologies [irds22]

3.3 MOSFET presentation

The Metal Oxide Semiconductor (MOS) Field Effect Transistor (FET) belongs to the large family of unipolar FETs which can be set apart from the Bipolar transistors.

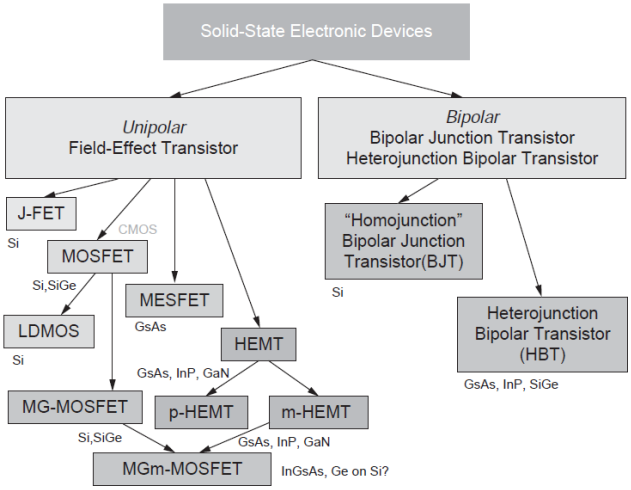


Figure 3.7 Main high-frequency transistor types [Voin13]

The principle of operation of both categories consists of the control of charge transfer between the terminals. In FETs, the control is done with an accumulation of charges on the gate which is isolated from the conduction channel between the drain and source. The charges accumulated on the gate create a conduction channel of opposite charges between the source and drain. Consequently, two types of MOSFET transistors exist based on the charge sign: NMOS where electrons flow through the conduction channel, and PMOS with holes.

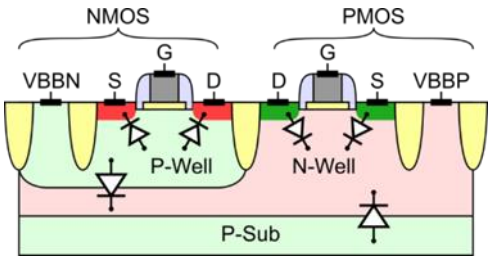


Figure 3.8 Schematic of NMOS and PMOS transistors in CMOS bulk technology [Cath17]

The MOSFET transistor was created in bulk technology as in Figure 3.8 where the intrinsic transistor is fabricated directly into the wafer substrate. Regions under the surface are then dopped to form the transistor. The channel is controlled from the gate above but the region

underneath the transistor lacks control. Multiple evolutions of the MOSFET have seen the day over the years alongside channel length reduction. Figure 3.9 presents the different types of CMOS transistors that have been developed.

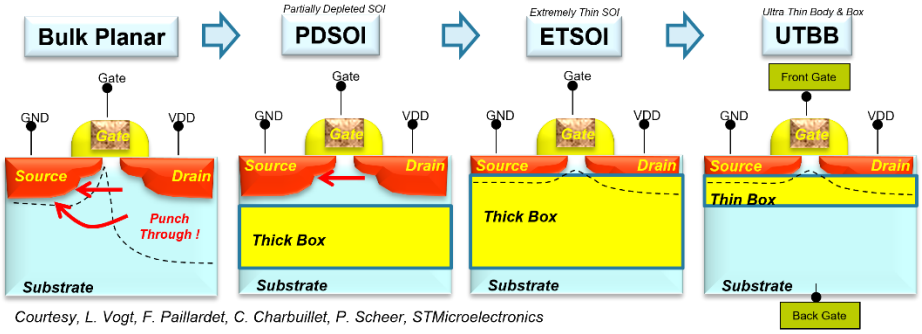


Figure 3.9 Evolution of MOSFET transistor types, from bulk to FDSOI (UTBB)

At one point the further scaling of transistor gate length stopped improving its analog performances [Haen06]. Because of the small dimensions, the proximity of the drain and source, and the thinnest of the Gate oxide, multiple short-channel effects arise:

-High subthreshold conduction: Lowering of V_t brings a higher *On* current but at the coast of an increase of *Off* Current, thus subthreshold leakage becomes a non-negligible part of the circuit DC consumption.

-Gate oxide leakage: because of the thinnest of the oxide, quantum mechanical phenomena of electron tunneling occur between the gate and channel, leading also to an increase in power consumption.

-Drain-induced barrier lowering (DIBL): At small dimensions, the drain voltage starts to contribute to the channel formation leading to a dependency of V_t on V_{DS}.

The different topologies presented in Figure 3.9 had the objective of alleviating or diminishing these short-channel effects.

The most advanced transistor evolution is the Ultra-Thin Body and Box (UTBB), also named Fully Depleted Silicon on Insulator (FD-SOI).

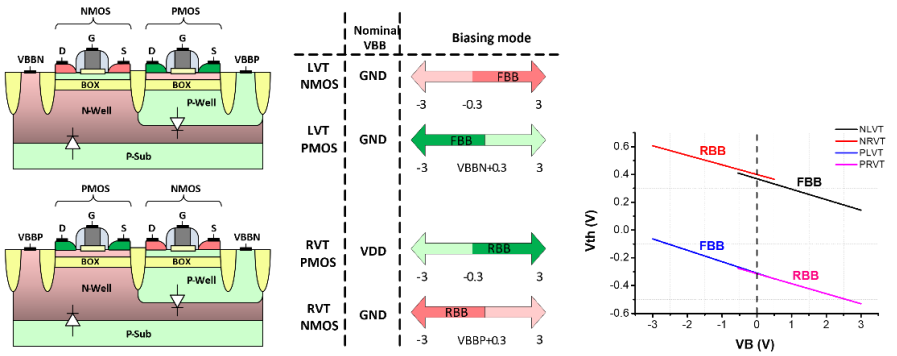
The 28nm FD-SOI technology platform from STMicroelectronics demonstrates 32% and 84% speed boost at 1.0V and 0.6V respectively, without adding process complexity compared to standard bulk technology [Planes12].

Thanks to the undoped 7nm thick channel above an SOI thin box, a superior electrostatic channel control is achieved, and the Short Channel Effects are significantly reduced in comparison to a bulk process. The BOX, which is 25nm thick has a minimal impact on thermal dissipation through the substrate.

The principal feature of the FD-SOI is enabled thanks to the proximity of the body underneath the canal which allows it to affect the channel properties through the potential applied on the body. The effect of a given applied potential on the body of the transistor is the modification of the threshold voltage.

Figure 3.10 (a) presents the two flavors of transistors of the technology, which differ by the Well type under the box. For transistors with complementary Well types (N-Well for NMOS and P-Well for PMOS), V_t can be reduced proportionally to the voltage on the body, hence their name Low-Vt (LVT). Symmetrically, for transistors with the same type of Well types, V_t can be increased with

the body bias, and are named Regular V_t (RVT). Figure 3.10 (b) presents the variations of V_t in the function of the body voltage for the different types of transistors.



(a) (b)

Figure 3.10 Schematic of 28nm FD-SOI transistor flavors (a) and threshold voltage control capability through the backgate voltage (b) [Cath17]

Because of the analogous function of the body voltage on the drain-source current to gate one, in FD-SOI the body is called backgate.

3.4 I-V characteristics

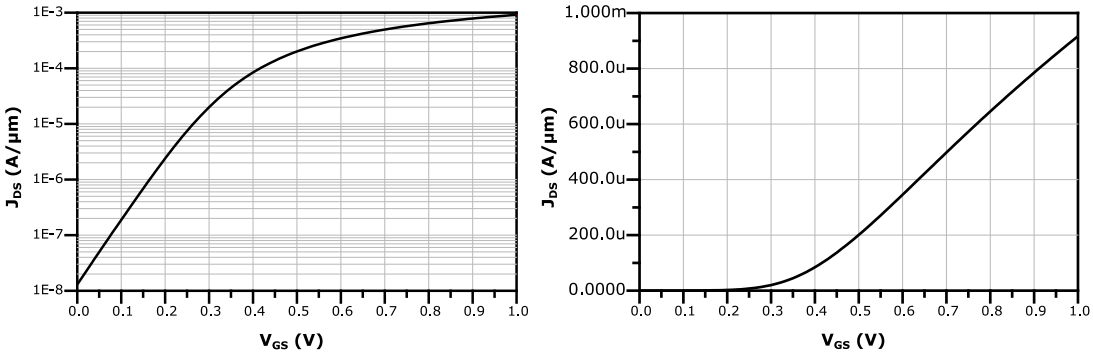
Figure 3.11 presents the simulated $J_{DS}(V_{GS})$ characteristics of a lvtnfet transistor with L=30 nm in 28 nm FD-SOI. Three regions can be identified:

-the subthreshold region (for $V_{GS}<V_t$) on the logarithmic scale graph, which presents exponential variations,

-the square-law region (for $V_{GS}\leq V_t$ and $J_{DS}\leq 0.15$ mA/μm),

where the drain current is a second-order polynomial function of V_{GS},

-the linear-law region (for $V_{GS}>V_T$ and $J_{DS}\geq 0.15$mA/μm), where the drain current is a linear function of V_{GS}.



(a) (b)

Figure 3.11 Simulation of the current density J_{DS} in function of gate-source voltage with $V_{DS}=1V$ and $L=30nm$, for a lvtnfet in 28nm FD-SOI in log (a) and linear (b) scale

The following Figure 3.12 presents the simulated $J_{DS}(V_{DS})$ for different values of gate-source voltages. Two regions can be distinguished:

-$V_{DS}<V_{GS}-V_t$ the linear region where the drain-source current is a second-order polynomial function of V_{DS},

-$V_{DS}>V_{GS}-V_t$ the saturation region where the current density is a linear function of V_{DS}.

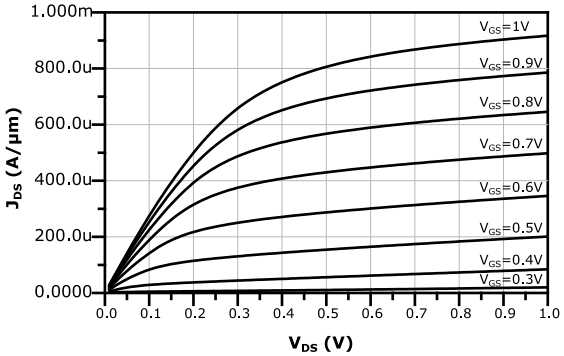


Figure 3.12 Simulation of the current density J_{DS} in function of drain-source voltage with different values of V_{GS} and *L=30nm*, for a lvtnfet in 28nm FD-SOI

Hence, in the linear-law region and saturation, the drain-source current is a linear function of both V_{GS} and V_{DS} and can be expressed as:

$$J_{DS} = \frac{I_{DS}}{W_{tot}} = g_m(V_{GS} - V_t)(1 + \lambda V_{DS}) \tag{3.26}$$

Where g_m is the transistor transconductance per unit of gate total width and λ is the channel length modulation parameter.

3.5 High-frequency transistor linear model

3.5.1 Introduction

The design of mm-wave and sub-THz circuits requires a great understanding of device operation. Usual blocks such as Power Amplifier, Low Noise Amplifiers, and Frequency multipliers… are based on transistors. The design of such transistors requires a meticulous layout to reduce the parasitical elements.

This part studies the key parameters of these devices.

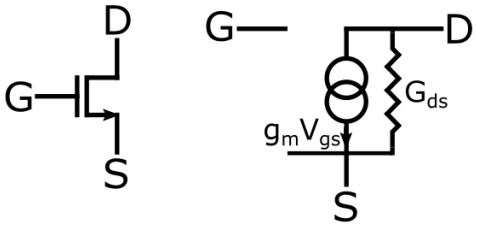


Figure 3.13 Transistor symbol (a) Intrinsic transistor model (b)

To study frequency responses, the transistor is represented by a linear model around a bias point. The linear model then reflects the small signal response of the component.

The intrinsic mechanism inside a MOS transistor is a current source I_{DS} controlled with a voltage difference V_{GS}. The current I_{DS} is also dependent on the voltage V_{DS}, which is represented by the admittance G_{DS}. According to Figure 3.13 and Equation (3.65), the linear-law region and saturation can be represented in one equation:

$$I_{DS} = g_m V_{GS} + G_{DS} V_{DS} \tag{3.27}$$

The physical layout of the transistor leads to electrical coupling between its ports and consequently to the presence of parasitical capacitances C_{gd} and C_{ds}. The intrinsic function of the transistor is enabled through the accumulation of charges inside the conduction canal, thus, C_{gs} is intrinsically present.

While at lower frequencies the parasitical capacitances do not modify the ideal intrinsic function of the transistor, as the frequency increases, more current flows through the capacitances, and the function of the device is modified.
Hence the model considering the capacitances is the following,

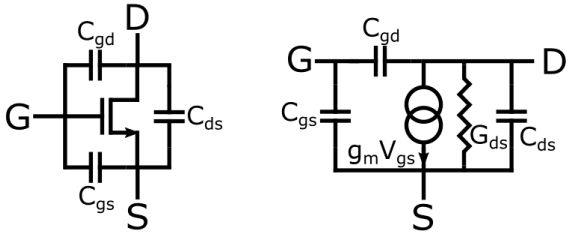


Figure 3.14 Transistor with capacitances (a) Transistor with capacitances model

Finally, access resistances are non-negligible and can be added to the model.
The access is made of thin metal layers of polysilicon for the gate which presents a consequent resistance.

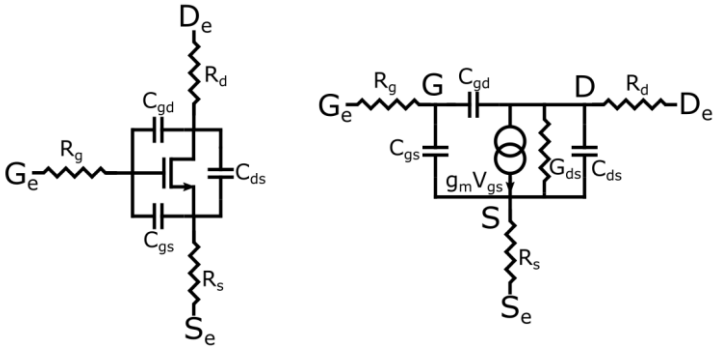


Figure 3.15 Transistor with extrinsic resistances (a) Transistor with extrinsic resistances model (b)

As we have seen in the previous chapter on two-port networks, any two-port can be represented by a matrix linking its ports' voltages and currents.
From the matrix, formulas for maximum available gain, impedances, or stability can be calculated.
Hence, transistors are used as part of amplification blocks, we want to study them in cases where they can amplify. Among the three possible dispositions, Common Source, Common Gate, and Common Drain, only the first two can be used as amplifier blocks.

3.5.2 Common source

Common Source topology considering intrinsic transistor with its capacitances:

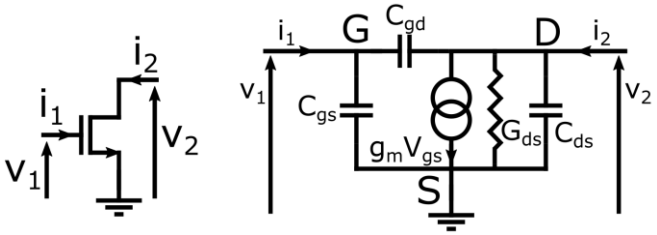


Figure 3.16 Transistor used in Common Source topology schematic (a) model (b)

i Two-port representation

The currents of the two-port are related to the voltages as follows:

$$i_1 = sC_{gs}v_1 + sC_{gd}(v_1 - v_2) \quad (3.28)$$

$$i_2 = (g_{ds} + sC_{ds})v_2 + sC_{gd}(v_2 - v_1) + g_m v_1 \quad (3.29)$$

Consequently, *Y* parameters expressions are:

$$Y_{11} = \frac{i_1}{v_1}\bigg|_{v_2=0} = s\left(C_{gs} + C_{gd}\right) \tag{3.30}$$

$$Y_{12} = \frac{i_1}{v_2}\bigg|_{v_1=0} = -sC_{gd} \tag{3.31}$$

$$Y_{22} = \frac{i_2}{v_2}\bigg|_{v_1=0} = g_{ds} + s\left(C_{ds} + C_{gd}\right) \tag{3.32}$$

$$Y_{21} = \frac{i_2}{v_1}\bigg|_{v_2=0} = g_m - sC_{gd} \tag{3.33}$$

ii Transition frequency (f_t) and maximal operation frequency (f_{max})

Two common figures of merit (FOM) of the transistor are the transition frequency f_t and the maximal operation frequency f_{max}.

f_t is defined as the frequency at which the magnitude of the two port's parameter H_{21} is equal to *1*.

By definition,

$$H_{21} = \frac{I_2}{I_1}\bigg|_{V_2=0} \tag{3.34}$$

It represents the current gain of the common source transistor topology when the output drain node is shorted.

From the matrix transformation formulas, H_{21} can be expressed in terms of *Y* parameters as:

$$H_{21} = \frac{Y_{21}}{Y_{11}} \tag{3.35}$$

Thus,

$$H_{21} = \frac{g_m - sC_{gd}}{s\left(C_{gs} + C_{gd}\right)} \tag{3.36}$$

Finally, by solving the following equation for f_t,

$$|H_{21}| = \left|\frac{g_m - j2\pi f_t C_{gd}}{j2\pi f_t\left(C_{gs} + C_{gd}\right)}\right| = 1 \tag{3.37}$$

it gives the expression of f_t.

$$f_t = \frac{g_m}{2\pi C_{gs}\sqrt{1 + 2\frac{C_{gd}}{C_{gs}}}} \tag{3.38}$$

Intuitively, when the output is shorted, the output current is equal to g_m multiplied by the voltage across the gate-source port. The voltage decreases proportionally to $1/\omega$, thus the decreases of the current gain. A part of the input current flows through C_{gd} which explains the presence of it in equation 3.38.

The f_{max} is defined as the frequency when the Unilateral Mason's Gain *U* is equal to unity. Which represents the maximal power gain of the unilateralized common source topology transistor.

As we have seen *U* is defined as:

$$U = \frac{|Z_{21} - Z_{12}|^2}{4(Re\{Z_{11}\}Re\{Z_{22}\} - Re\{Z_{21}\}Re\{Z_{12}\})} = \frac{|Y_{21} - Y_{12}|^2}{4(Re\{Y_{11}\}Re\{Y_{22}\} - Re\{Y_{21}\}Re\{Y_{12}\})} \tag{3.39}$$

Because Y_{11} and Y_{12} are purely imaginary (equations 3.30 and 3.31), the denominator of U is equal to 0 and the result is infinite.

A unilateralized transistor implies a perfectly compensated C_{gd}, then because it is a maximum power gain, i.e., when the two ports are matched, the C_{gs} is also compensated by the input port impedance. Then the equivalent f_t is infinite and consequently also the power gain.

Without the presence of series resistances, the maximal unilateralized power gain is infinite. Hence,

$$f_{max} = +\infty \tag{3.40}$$

iii Input and Output impedances

From the expressions of the input and output admittances in terms of Y parameters we can write:

$$Y_{in} = Y_{11} - \frac{Y_{21}}{Y_{22} + Y_L} Y_{12} \underset{C_{gd}=0}{\approx} sC_{gs} \tag{3.41}$$

$$Y_{out} = Y_{22} - \frac{Y_{21}}{Y_{11} + Y_{Load,1}} Y_{12} \underset{C_{gd}=0}{\approx} g_{ds} + sC_{ds} \tag{3.42}$$

In these expressions, the second term is relatively small compared to the first one. Hence, the input impedance is only a capacitance while the output is a capacitance in parallel with a relatively large resistance of $1/g_{ds}$.

iv Maximum available gain and stability

When applying the equation for stability factor to the Y parameters, we obtain:

$$K = \frac{1}{\sqrt{1 + \left(\frac{g_m}{\omega C_{gd}}\right)^2}} < 1 \tag{3.43}$$

The stability factor is always smaller than 1 but always positive. Simultaneous two ports conjugate matching is not possible and there are source and load impedances which produce negative impedance real part.

Hence, the Maximum Available Gain (G_{ma}) which represents the two-port power gain for matched terminals is not defined.

However, as an indication of the available power gain, the Maximum Stable Gain can be calculated:

$$G_{msg} = \left|\frac{Y_{21}}{Y_{12}}\right| = \sqrt{1 + \left(\frac{g_m}{\omega C_{gd}}\right)^2} = \left(\frac{\omega C_{gd}}{g_m}\right) \sqrt{1 + \left(\frac{\omega C_{gd}}{g_m}\right)^2} \tag{3.44}$$

G_{msg} is proportional to g_m/C_{gd} and decreases at a rate of $1/\omega$.

The frequency $g_m/(2\pi C_{gd})$ is very high and above the f_t and f_{max} of the device, thus in the conditions where the transistor is used the square root term can be removed.

$$G_{msg} \approx \left(\frac{\omega C_{gd}}{g_m}\right) \tag{3.45}$$

3.5.3 Common source topology in the presence of extrinsic resistances

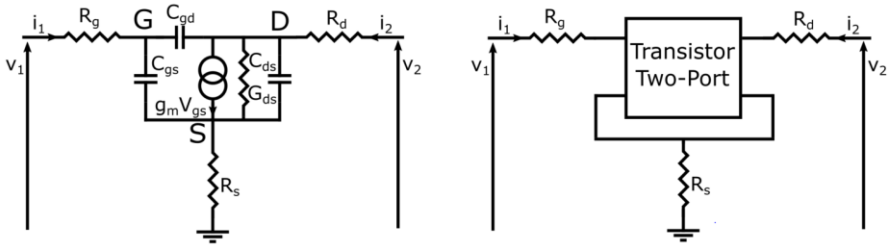


Figure 3.17 Transistor used in Common Source topology model with components (a) with intrinsic two-port (b)

i Two-port representation

The calculation of the transistor with its extrinsic resistances two port parameters can be simply done from the previously calculated *Y* parameters as follows:

$$Z_e = \begin{bmatrix} R_g + R_s + Z_{11} & R_s + Z_{12} \\ R_s + Z_{21} & R_d + R_s + Z_{22} \end{bmatrix} \tag{3.46}$$

Where Z_{ij} parameters are the result of *Y* to *Z* matrix transformation,

$$Z = \frac{1}{\Delta Y}\begin{bmatrix} Y_{22} & -Y_{12} \\ -Y_{21} & Y_{11} \end{bmatrix} \tag{3.47}$$

Thus,

$$Z_e = \begin{bmatrix} R_g + R_s + \frac{Y_{22}}{\Delta Y} & R_s - \frac{Y_{12}}{\Delta Y} \\ R_s - \frac{Y_{21}}{\Delta Y} & R_d + R_s + \frac{Y_{11}}{\Delta Y} \end{bmatrix} = \begin{bmatrix} R_g + R_s + \frac{g_{ds} + s(C_{ds} + C_{gd})}{\Delta Y} & R_s - \frac{-sC_{gd}}{\Delta Y} \\ R_s - \frac{g_m - sC_{gd}}{\Delta Y} & R_d + R_s + \frac{s(C_{gs} + C_{gd})}{\Delta Y} \end{bmatrix} \tag{3.48}$$

The determinant of *Y* is equal to,

$$\begin{aligned} \Delta Y &= s(C_{gs} + C_{gd})[g_{ds} + s(C_{ds} + C_{gd})] + sC_{gd}(g_m - sC_{gd}) \\ &= s^2 a + sb = sb\left(1 + s\frac{a}{b}\right) \end{aligned} \tag{3.49}$$

With,

$$a = (C_{gs} + C_{gd})(C_{ds} + C_{gd}) - C_{gd}^2 \tag{3.50}$$

$$b = C_{gd} g_m + (C_{gs} + C_{gd}) g_{ds} \tag{3.51}$$

When $\omega \ll b/\mathrm{a}$, the s^2 term is negligible and the determinant is only proportional to *s*.

$$\frac{b}{a} \approx 10^{12}\ \mathrm{Hz}$$

For frequencies below the THz, the following simplification can be done:

$$\Delta Y \approx sb = s(C_{gd} g_m + (C_{gs} + C_{gd}) g_{ds}) \tag{3.52}$$

ii Transition frequency (f_t) and maximal operation frequency (f_{max})

The transformation relation between *Z* parameters and H_{21} is:

$$H_{21} = -\frac{Z_{21}}{Z_{22}} \tag{3.53}$$

By replacing the Z_{ij} parameters it becomes,

$$H_{21} = \frac{Y_{21} - R_s \Delta Y}{Y_{11} + (R_d + R_s)\Delta Y} = \frac{g_m - sC_{gd} - sR_s b}{s(C_{gs} + C_{gd}) + s(R_d + R_s)b} \tag{3.54}$$

$$|H_{21}| = \frac{g_m}{\omega} \frac{\sqrt{1 + \left(\omega \frac{C_{gd} + R_s b}{g_m}\right)^2}}{\left(C_{gs} + C_{gd}\right) + (R_d + R_s)b} \tag{3.55}$$

It follows that when $|H_{21}|$ is equal to unity, the f_t expression is:

$$f_t = \frac{g_m}{2\pi\left(C_{gs} + R_d b\right)\sqrt{1 + 2\frac{C_{gd} + R_s b}{C_{gs} + R_d b}}}$$

$$= \frac{g_m}{2\pi\left(C_{gs} + R_d\left(C_{gd} g_m + \left(C_{gs} + C_{gd}\right) g_{ds}\right)\right)\sqrt{1 + 2\frac{C_{gd} + R_s\left(C_{gd} g_m + \left(C_{gs} + C_{gd}\right) g_{ds}\right)}{C_{gs} + R_d\left(C_{gd} g_m + \left(C_{gs} + C_{gd}\right) g_{ds}\right)}}} \tag{3.56}$$

It shows that the f_t is inversely proportional to the drain and source access resistances. Finally, we observe that the gate access resistance does not enter into account.

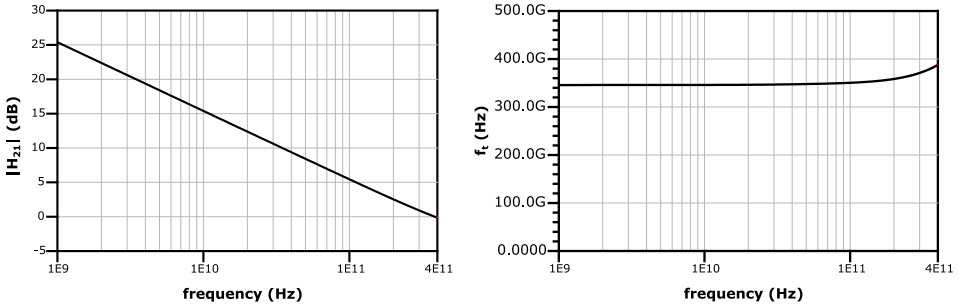


Figure 3.18 Simulations of $|H_{21}|$ (a) and f_t extrapolation (b) over frequency of a lvtnfet transistor in 28 nm FD-SOI with *L=30 nm*, V_{DS}*=1 V*, and J_{DS}*=0.3 mA/µm*

By injecting expressions from (3.48) into the definition of the Mason's Gain (3.39) and following development, *U* can be expressed as:

$$U = \frac{g_m^2}{4\omega^2\left(\left(R_g R_d + \left(R_g + R_d\right)R_s\right)b^2 + b\left[\left(R_g + R_s\right)\left(C_{gs} + C_{gd}\right) + \left(R_d + R_s\right)\left(C_{ds} + C_{gd}\right) - 2R_s C_{gd}\right]\right)} \tag{3.57}$$

f_{max} expression can then be expressed as:

$$f_{max} = \frac{g_m}{4\pi\sqrt{\left(R_g R_d + \left(R_g + R_d\right)R_s\right)b^2 + b\left[\left(R_g + R_s\right)\left(C_{gs} + C_{gd}\right) + \left(R_d + R_s\right)\left(C_{ds} + C_{gd}\right) - 2R_s C_{gd}\right]}} \tag{3.58}$$

It can be approximated as:

$$f_{max} \approx \frac{g_m}{4\pi C_{gs}\sqrt{\left(\frac{C_{gd}}{C_{gs}} g_m + g_{ds}\right)\left(R_g + R_s\right)}} \tag{3.59}$$

U decreases proportionally to $1/\omega^2$ (Figure 3.19 (a)), hence f_{max} can be extrapolated for a large range of frequencies with:

$$f_{max,ex} = f\sqrt{U} \tag{3.60}$$

As we can see in Figure 3.19, because Mason's Gain function decreases at a steady rate, the f_{max} value can be extrapolated. The best accuracy is achieved for frequencies above 100 GHz where second-order terms intervene. Below the f_{max} is overestimated at around 20 GHz which is less than a 10% difference.

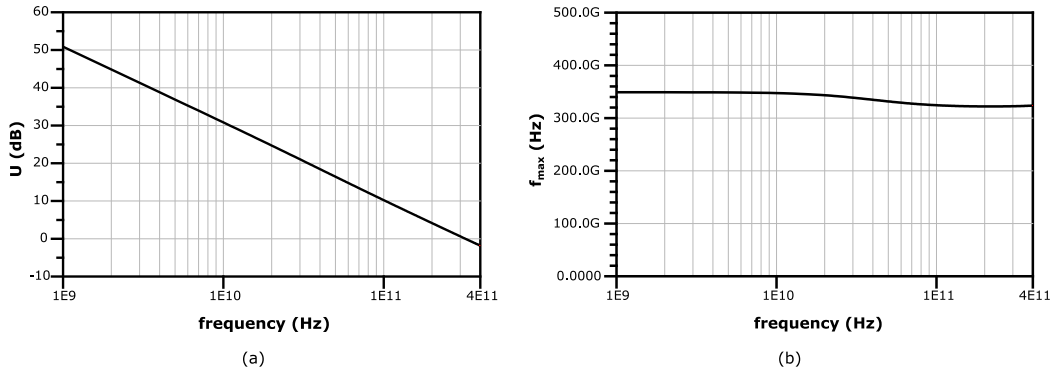


Figure 3.19 Simulations of *U* (a) and f_{max} extrapolation (b) over frequency of a lvtnfet transistor in 28nm FD-SOI with *L=30 nm*, V_{DS}*=1 V* and J_{DS}*=0.3 mA/µm*

ii Input and Output impedances

By applying the definition of the input impedance to the Z_e parameters, for a purely resistive load charge R_{Load}, the expression of Z_{in} is:

$$Z_{in} = R_g + R_s + \frac{C_{ds}+C_{gd}}{b} - \frac{\left(R_s + \frac{C_{gd}}{b}\right)^2}{R_d + R_s + R_L + \frac{C_{gs}+C_{gd}}{b}} - \frac{j}{\omega b}\left(g_{ds} + g_m \frac{R_s + \frac{C_{gd}}{b}}{R_d + R_s + R_{Load} + \frac{C_{gs}+C_{gd}}{b}}\right) \quad (3.61)$$

The real part of Z_{in} is independent of frequency which is noteworthy and can ease the input impedance matching. The imaginary part is negative and decreases proportionally to $1/\omega$. Hence the input impedance is equivalent to a resistance in series with a capacitance.

Similarly, in the presence of a resistive source impedance R_{Source}, the expression of Z_{out} is:

$$Z_{out} = R_d + R_s + \frac{C_{gs}+C_{gd}}{b} + \left(R_s + \frac{C_{gd}}{b}\right)\frac{g_m}{g_{ds}} \frac{1 - \frac{\omega^2}{g_m g_{ds}}\left(R_s b + C_{gd}\right)\left(\left(R_{Source} + R_s + R_g\right)b + C_{ds} + C_{gd}\right) - j\omega\left[\left(R_{Source} + 2R_s + R_g\right)b + C_{ds} + 2C_{gd}\right]}{1 + \frac{\omega^2}{g_{ds}^2}\left(\left(R_g + R_s + R_{Source}\right)b + C_{ds} + C_{gd}\right)^2} \quad (3.62)$$

On the contrary, the output impedance variations over frequency are different. The real part is frequency-dependent and starts to decrease. For the imaginary part, it is not linear with frequency as well.

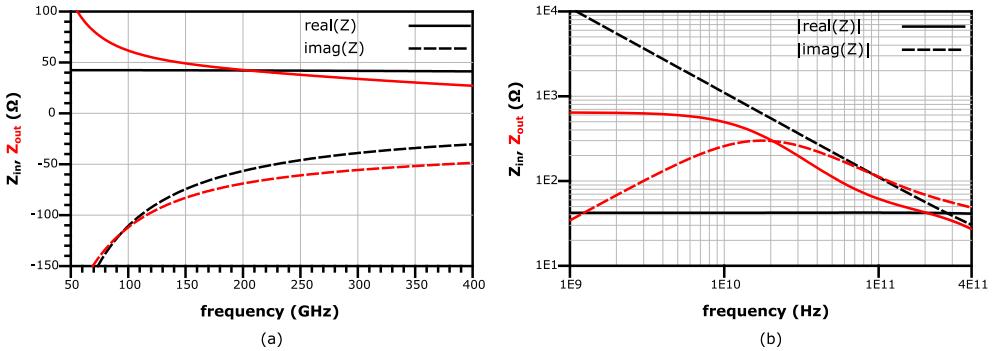


Figure 3.20 Simulations of input and output impedances over frequency in linear (a) and log (b) scale of a CS transistor with *L=30 nm* and *W=16 µm* at J_{DS}*=0.3 mA/µm*

Figure 3.20 presents the simulations of Z_{in} and Z_{out} impedances of a transistor with source and load port impedances of 50 Ω. Their variations over frequency follow the predictions from expressions up to 400 GHz.

iii Maximum available gain and stability

Stability factor:

$$K = \frac{\omega}{\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^2}} \frac{bN}{g_m\left(R_s b + C_{gd}\right)} = \frac{\omega}{\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^2}} A \tag{3.63}$$

with,

$$\omega_0 = \frac{g_m}{R_s b + C_{gd}} \tag{3.64}$$

$$N = 2\left[R_s\left(C_{gs}+C_{ds}+C_{gd}\right) + R_g\left(C_{gs}+C_{gd}\right) + R_d\left(C_{ds}+C_{gd}\right) + b\left(R_g R_d + R_s\left(R_g+R_d\right)\right)\right] \tag{3.65}$$

The stability factor increases linearly with frequency, it is proportional to the access resistances and inversely proportional to the contributors of *real(Z_{12})*: R_s and C_{gd}.

For a unilateral transistor, i.e., when these two components are null, the stability factor is infinite.

ω_0 is a very high frequency, higher than the f_t or f_{max}. Hence the denominator of *K* can be simplified.

It is inversely proportional to the transistor transconductance g_m, consequently, a higher gain through an increase of g_m implies a lower stability factor.

The frequency at which *K* is equal to 1 is:

$$\omega_s = \frac{1}{\sqrt{A^2 - \frac{1}{\omega_0^2}}} \approx \frac{1}{A} = \frac{g_m\left(R_s b + C_{gd}\right)}{bN} \tag{3.66}$$

To increase a common source topology stability factor (by reducing ω_s), it is necessary to increase the access resistances or decrease its transconductance.

Additional capacitances added to the nodes of the extrinsic transistor do not impact the stability factor as the capacitances in the previous equations are inside the intrinsic transistor model, thus after the access resistances.

Maximum Stable Gain and Maximum Available Gain:

$$G_{msg} = \left|\frac{Z_{21}}{Z_{12}}\right| = \frac{\omega_0}{\omega}\sqrt{1+\left(\frac{\omega}{\omega_0}\right)^2} \tag{3.67}$$

Unlike the unilateral gain, the maximum stable gain decreases proportionally to $1/\omega$, however, it doesn't have a physical meaning. It gives an estimation of the device capability as the Maximum Available Gain is proportional to the G_{msg} when *K* becomes larger than unity.

The expression of G_{msg} in the presence of extrinsic access resistances is the same as without them. The difference arises in the expression of ω_0, where R_S intervenes.

$$G_{ma} = \left|\frac{Z_{21}}{Z_{12}}\right|\left(K - \sqrt{K^2-1}\right) = \left|\frac{Z_{21}}{Z_{12}}\right| \frac{1}{K+\sqrt{K^2-1}} \tag{3.68}$$

$$G_{ma} = \frac{\omega_0}{\omega} \frac{1+\left(\frac{\omega}{\omega_0}\right)^2}{\omega A + \sqrt{\left(\frac{\omega}{\omega_s}\right)^2 - 1}} \tag{3.69}$$

G_{ma} is defined when $\omega > \omega_s$, then it decreases at a rate of $1/\omega^2$ as the unilateral gain.

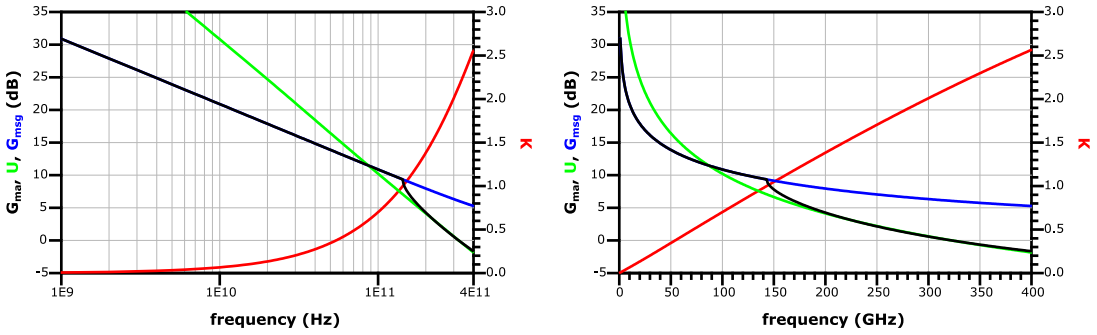


Figure 3.21 Simulations of G_{ma}, G_{msg}, and K over frequency for a lvtnfet transistor in 28nm FD-SOI with $L=30\ nm$, $Vds=1\ V$, and $j=0.3\ mA/\mu m$ in log (a) and linear (b) scale

Figure 3.21 presents the simulation of the different power gains and the stability factor of a transistor in a common source topology. The variations follow the predictions of the previous equations.

3.6 f_t, f_{max} dependency on current density and geometry

The transistor's linear characteristics depend on the drain-source current density. The transconductance and the gate-source capacitance which are intrinsic components of the conduction channel vary largely with J_{DS}. On the contrary, C_{gd}, C_{ds}, and R_g are physical parasitic elements and are almost independent of the current density.

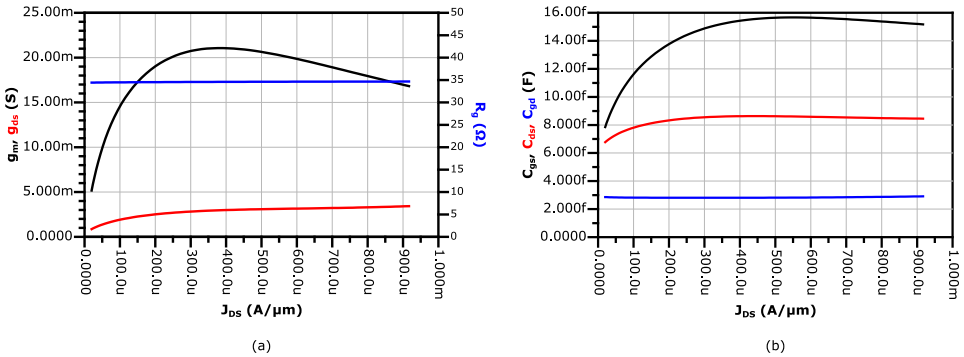


Figure 3.22 Simulations of transistor parameters in function of current density for a lvtnfet transistor in 28nm FD-SOI with $L=30\ nm$, $W_{tot}=16\ \mu m$, $Vds=1\ V$

Figure 3.22 presents the extracted linear component values at 150 GHz of the transistor as a function of J_{DS}.

Because the power gain performances of the transistor depend on the linear components (equations 3.56 and 3.58), consequently the f_{max} and f_t also depend on the current density.

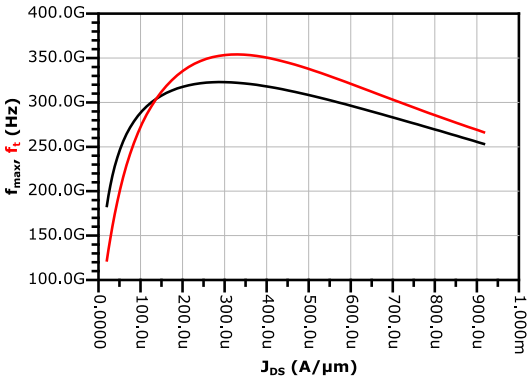


Figure 3.23 Simulations of transistor f_t/f_{max} as a function of current density for a lvtnfet transistor in 28nm FD-SOI with *L=30 nm*, W_{tot}*=16 µm*, *Vds=1 V*

Figure 3.23 presents the f_t and f_{max} of the transistor extracted at 150GHz as a function of J_{DS}. The f_t and f_{max} present maximal values for $J_{ds,opt,ft}$*≈0.35 mA/µm* and $J_{ds,opt,fmax}$*≈0.3 mA/µm* respectively. However, the f_t doesn't vary significantly in the *0.1mA/µm* interval around the peak and *0.2 mA/µm* interval for the f_{max}. Hence the transistor can be biased in this interval with almost no performance loss.

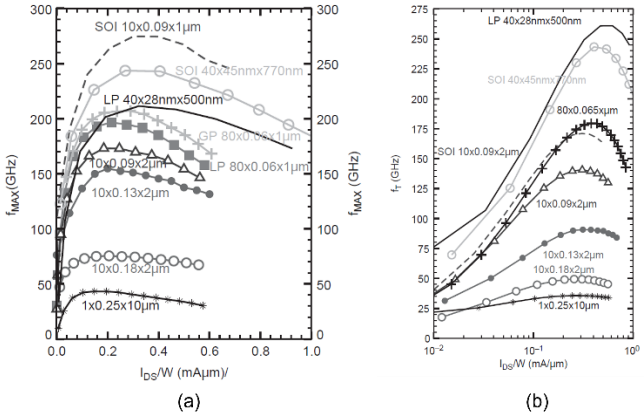


Figure 3.24 Measured f_t/f_{max} of N-MOS transistors in different processes in function of current density [Voin13]

Figure 3.24 presents the measured f_{max} and f_t performances of multiple MOSFET technologies as a function of the current density. It was observed that the optimal current densities for speed and noise are intrinsic to MOSFET and invariant with node scaling [Dick06]. Consequently, *0.3 mA/µm* is the well-known optimal bias point for MOSFET transistors to optimize the f_t and f_{max}.

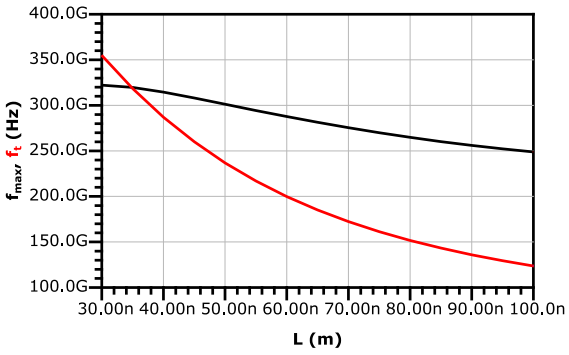


Figure 3.25 Simulated and extracted maximal f_{max} and f_t at 150 GHz in function of gate length *L* for *W=16 µm*

Figure 3.25 presents the simulated maximal (regarding current density) f_{max} and f_t as a function of the gate length. The g_m is proportional to L; hence the f_t decreases linearly with this dimension. The f_{max} represents the tradeoff between the transconductance and the access resistances. The drop is slower for the $f_{max,}$ and 40 nm presents only slightly lower performance in comparison to 30 nm.

The minimal available gate length of the technology has been chosen in this work, as this allows for maximizing the high-frequency performances of the device which keeps aligned with the main objective of this work.

The geometry of the transistor has important consequences on the device's performance. From its equation (3.58) f_{max} is inversely proportional to the access resistances. Among the access resistances, R_g is at least ten times higher than the R_s and R_d and is consequently the main resistive limiting factor in high-frequency performances. Figures 3.26 (a)-(d) depict the schematic views of transistor gate fingers with one or two gate contacts and one or multiple fingers. Figure 3.26 (e) presents one gate finger and its equivalent electrical circuit. The silicide layer on polysilicon is used to reduce the contact resistivity of the gate but the resistivity is still 200 times higher than copper thin metal layer *M1*. Because the drain and source fingers are connected to *M1* right above the channel explains the difference between their access resistances and the gate one.

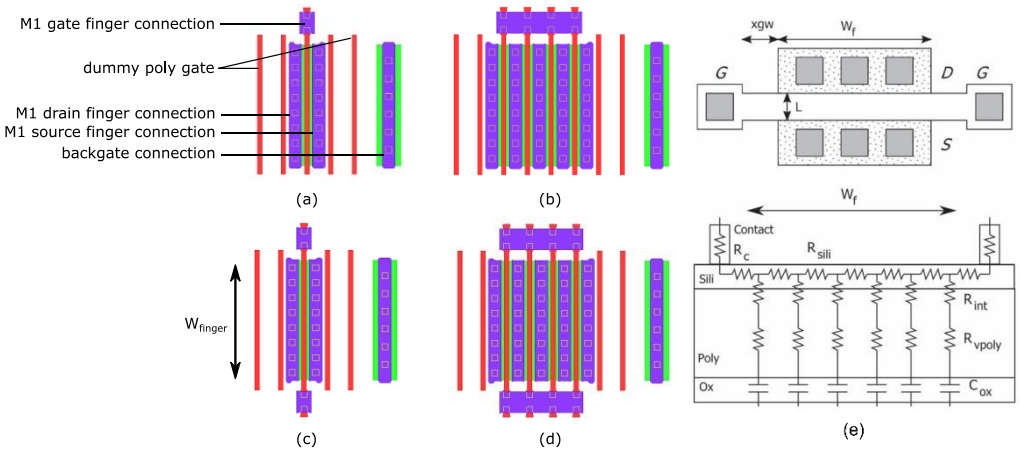


Figure 3.26 Schematics of lvtnfet rf transistor layouts with one gate access and one (a) or 4 gate fingers (b); two gate accesses and one (c) or 4 gate fingers (d). Top view and cross-section of a single-gate-finger transistor illustrating gate resistance components [Dorm13] (e)

Hence, the effective length of the gate polysilicon access determines the performance of the transistor. For two gate accesses on both sides of the gate finger, the effective polysilicon length is divided by two. Figure 3.27 presents the simulations of f_t and f_{max} for different transistor geometries. Between one and two gate accesses, f_t is not significantly impacted, and double access lowers its value because of additional capacitive coupling between the source, drain, and the additional gate area. In addition, from f_t expression (3.56) the absence of R_g dependence explains the simulation results. Conversely, as expected, f_{max} is largely increased by 20% from one gate access to two. Because multiple gate fingers are connected through *M1* metal where additional resistance is negligible in comparison to the poly finger resistance, the addition of fingers does not alter greatly f_t and f_{max}.

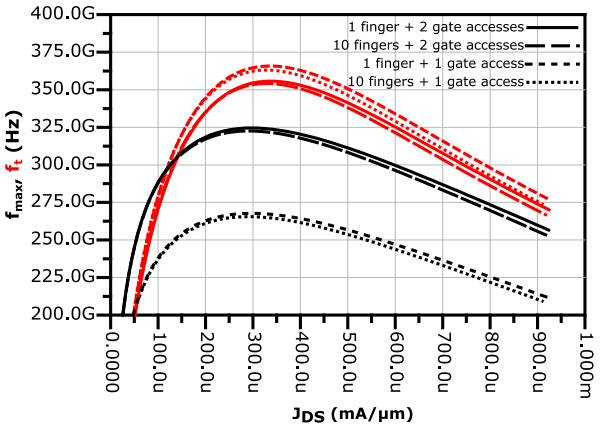


Figure 3.27 Simulations of f_t and f_{max} as a function of current density for one or two gate accesses and one or ten gate fingers with *L=30 nm* and W_{finger}*=800 nm*

3.7 Inductors

Inductors are fundamental in the design of RF circuits and their quality directly determines the performance of a design. Their physical understanding is important to identify IC inductor's limitations and loss mechanisms and therefore design the best possible inductors.

Inductors are components that store energy in a magnetic field when an electric current flows through it. It is characterized by inductance which is the capacity of an element to store magnetic energy. Hence the magnetic flux Φ_B generated by a current *I* is related to it by:

$$\Phi_B = LI \tag{3.70}$$

It is a representation of the tendency of an electrical conductor to oppose the change in the electrical current flowing through it. Faraday's law of induction indicates that a voltage *V* is induced on a circuit by any change in magnetic flux and is given by:

$$V = -\frac{d\Phi_B}{dt} \tag{3.71}$$

Consequently, the change in a circuit's current *I* is related to the inductance *L* and the resulting voltage across the circuit by:

$$V = -L\frac{dI}{dt} \tag{3.72}$$

According to (3.72) the impedance of an inductance is purely imaginary and is equal to $Z_L = \mathrm{j}\omega L$. It increases linearly with frequency as a faster variation in time requires a higher voltage. Therefore, to induce a certain alternative current in a conductor an energy proportional to its inductance must be spent. This energy is released when the current source stops, the inductors then discharge by continuing to deliver a current until the energy depletes. The inductor is made of a non-ideal conductor and thus has a non-infinite conductivity, leading to a series of resistance with the inductance. Consequently, a part of the energy flowing into the inductor is dissipated through the heat in the resistance. The quality factor of the inductor is defined as the ratio between the stored energy and the losses. Hence, the quality factor can be expressed as:

$$\mathrm{Q_{ind}} = \frac{\mathrm{L\omega}}{\mathrm{R}} \tag{3.73}$$

The quality factor can also be expressed as the ratio between the imaginary part of the impedance and the real part.

$$Q = \frac{\mathrm{imag(Z)}}{\mathrm{real(Z)}} \tag{3.74}$$

Inductors are made with conductors whose geometry has been optimized to have a high inductance with low resistance and thus a high-quality factor. A more general case is the mutual inductance between two conductors. As in Figure 3.28, a conductor C_1 with a current I_1 generates a magnetic field around it. A part of the field crosses the surface made by a closed loop of a conductor C_2.

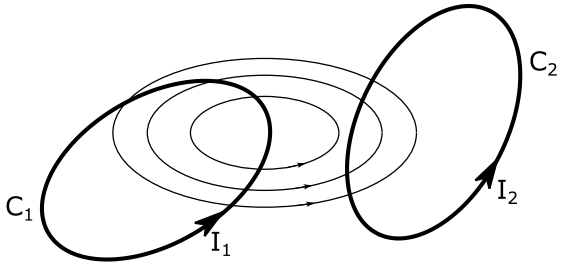


Figure 3.28 Schematic of magnetic fields coupling between two conductor loops

This part of the flux shared by both conductors can be expressed as:

$$\Phi_{21} = \int_{S2} \boldsymbol{B_1} \cdot \boldsymbol{dS} \tag{3.75}$$

From equation (3.75), the shared flux can be expressed in function of the conductors' geometries and the I_1 current:

$$\Phi_{21} = \frac{I_1 \mu_0}{4\pi} \oint_{C2} \oint_{C1} \frac{dl_1 . dl_2}{d_{12}} \tag{3.76}$$

Where d_{12} is the distance between the infinitesimal vectors dl_1 and dl_2.
The definition of inductance (3.70) can be extended to the mutual inductance.

$$\Phi_{21} = M_{21} I_1 \tag{3.77}$$

with

$$M_{21} = \frac{\mu_0}{4\pi} \oint_{C2} \oint_{C1} \frac{dl_1 . dl_2}{d_{12}} \tag{3.78}$$

M_{21} represents the shared magnetic energy storage between the two conductors. Consequently, all conductors that share a part of their magnetic field with another conductor share a part of their stored magnetic energy through mutual inductance. The mutual inductance is symmetric from (3.78), hence.

$$M_{21} = M_{12} \tag{3.79}$$

Faraday's law in the case of mutual inductance is then expressed as

$$\varepsilon_2 = -M_{21} \frac{dI_1}{dt} \tag{3.80}$$

Finally, (3.80) indicates that a change of current in conductor 1 induces a voltage difference in conductor 2.

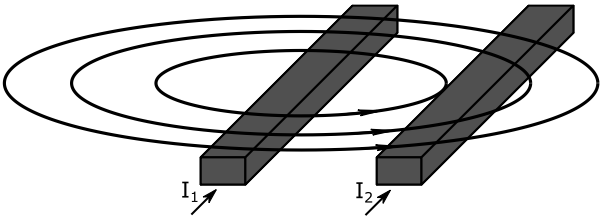


Figure 3.29 Two rectangular conductors with magnetic field coupling

The problem of mutual and self-inductance between closed conductors can be divided into partial inductances that represent the magnetic coupling of parts of conductors. The total inductance is

then the sum of the two-by-two coupling between the conductors' parts. The mutual inductance of two conductors as in Figure 3.29 can be calculated with a generalized expression from (3.117).

$$\mathrm{M}_{21} = \frac{\mu_0}{4\pi I_1 I_2} \int_{V1} \int_{V2} \frac{\mathrm{dj}_1 . \mathrm{dj}_2}{\mathrm{d}_{12}} \tag{3.81}$$

Where V_1 and V_2 are the volumes of conductors 1 and 2, dj_1 and dj_2 are the current densities vectors in the infinitesimal volumes of the integral. For uniform current distributions in the conductors, it can be simplified to:

$$\mathrm{M}_{21} = \frac{\mu_0}{4\pi} \int_{V1} \int_{V2} \frac{\mathrm{dv}_1 . \mathrm{dv}_2}{\mathrm{d}_{12}} \tag{3.82}$$

In addition, self-inductance is the mutual inductance of the conductor to itself, hence, from (3.82),

$$L = \frac{\mu_0}{4\pi} \int_V \int_V \frac{\mathrm{dv}_1 . \mathrm{dv}_2}{\mathrm{d}_{12}} \tag{3.83}$$

For uniform current distributions, the mutual and self-inductance calculation only depends on the geometries and position in space. Numerous exact formulas exist for integrals from (3.82) and (3.83) for given conductors geometries [Rueh72][Grov73][Piat12] and can be used to quickly evaluate given inductors characteristics [Kamo93]. Despite the speed of this approach, EM solvers offer the evaluation of any circuit shape and capture all magnetic and electric field interactions and thus are more generalist.

In the presence of a second conductor as in Figure 3.29, the voltage across the first conductor can be expressed as:

$$V_1 = (j\omega L_1 + R_1) I_1 + j\omega M_{12} I_2 \tag{3.84}$$

The magnetic coupling interaction can have multiple effects on a circuit, the three main cases for two coupled conductors are:

-the two conductors are flowed by equal currents, (3.84) becomes then

$$\mathrm{V}_1 = (j\omega(\mathrm{L}_1 + M_{12}) + \mathrm{R}_1)\mathrm{I}_1 \tag{3.85}$$

the effective inductance of the conductor is artificially increased by M_{12}, as well as for conductor 2.

-similarly, when the signals are out of phase the voltage across the conductor is

$$\mathrm{V}_1 = (j\omega(\mathrm{L}_1 - M_{12}) + \mathrm{R}_1)\mathrm{I}_1 \tag{3.86}$$

Therefore, the effective inductance is then decreased by the mutual inductance

-only one conductor is flowed by a current, hence the voltage difference across the second conductor is null which gives the following relations

$$V_1 = (j\omega L_1 + R_1) I_1 + j\omega M_{12} I_2 \tag{3.87}$$

$$0 = (j\omega L_2 + R_2) I_2 + j\omega M_{21} I_1 \tag{3.88}$$

By substituting 3.88 into 3.87, V_1 can be expressed as:

$$V_1 = \mathrm{j}\omega \left(L_1 - (\omega M_{12})^2 \frac{L_2}{R_2^2 + (\omega L_2)^2} \right) I_1 + \left(R_1 + (\omega M_{12})^2 \frac{R_2}{R_2^2 + (\omega L_2)^2} \right) I_1$$

$$= j\omega L_1 \left(1 - \frac{k^2}{1 + \frac{1}{Q_2^2}} \right) I_1 + R_1 \left(1 + \frac{Q_1}{Q_2} \frac{k^2}{1 + \frac{1}{Q_2^2}} \right) I_1 \tag{3.89}$$

Where,

$$k = \frac{M_{12}}{\sqrt{L_1 L_2}} \tag{3.90}$$

k is known as the magnetic coupling factor; it is the ratio between the shared magnetic energy storage M_{12} and the geometric mean inductance between L_1 and L_2. It indicates how much of the magnetic energy is shared, 1, being all, and 0, none. Therefore, *k* is higher as the two conductors get closer to each other.

Equation (3.89) shows that the presence of a conductor that doesn't conduct any signal or isn't used in the main circuit has an impact on the main conductor. The effective inductance is reduced proportionally to the magnetic coupling, and conversely, the resistivity increases proportionally to the coupling. Consequently, the quality factor reduces with the magnetic coupling. Hence, the presence of an additional conductor deteriorates the performance of the main conductor.

However, a particular case can be deduced from (3.89). When the quality factor of the parasitical conductor is close to 0, its impact on the main conductor vanishes. From (3.88), I_1 generates a potential difference in the second conductor that induces a current I_2. If R_2 is infinitely high, I_2 tends to 0, and no current is induced in the second conductor.

This case is observed with the silicon substrate of integrated technologies. Because of the small dimensions, the conductors are always close to the substrate which is a highly resistive conductor compared to common copper or aluminum metal layers, but still induces losses through magnetically induced currents. The use of a very high resistivity substrate, designed for this purpose, allows for the reduction of the losses in integrated passive circuits and hence increases their quality factors, due to the mechanism described above. With standard silicon substrate, when designing passive elements, the substrate should be kept non-dopped, as it reduces the conductivity and hence the losses.

The following Figure 3.30 sums up the results from the three cases.

Second conductor signal type	Effective inductance in Conductor 1	Effective resistance in Conductor 1	Effective quality factor in Conductor 1
In phase	$\mathrm{L}_1 + M_{12}$	R_1	$\frac{\omega(\mathrm{L}_1 + M_{12})}{\mathrm{R}_1}$
Out of phase	$\mathrm{L}_1 - M_{12}$	R_1	$\frac{\omega(\mathrm{L}_1 - M_{12})}{\mathrm{R}_1}$
None	$L_1\left(1 - \frac{k^2}{1+\frac{1}{Q_2^2}}\right)$	$R_1\left(1 + \frac{Q_1}{Q_2}\frac{k^2}{1+\frac{1}{Q_2^2}}\right)$	$Q_1\frac{1+(1-k^2)Q_2^2}{1+Q_2^2+Q_1Q_2k^2}$

Figure 3.30 Summary of the effect of magnetic coupling between two conductors on the effective inductance, resistance, and quality factor

The current distribution was first considered uniformly distributed inside the conductor. However, the skin effect tends to force the current to flow less in the middle of the conductor than close to the surface and the phenomenon increases with frequency. Figure 3.31 depicts a rectangular conductor divided into sub-conductors.

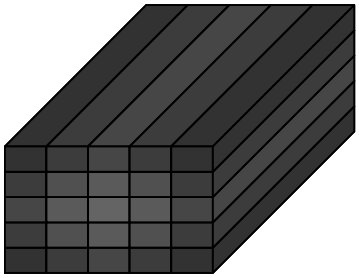


Figure 3.31 Schematic of a conductor divided into sub-conductors where the number of conductor's neighbors is proportional to its brightness.

As recently presented the presence of near conductors with a flowing in-phase current increases the effective inductance and hence the impedance. Thus, sub-conductors in the center of the conductor have more close neighbors than those on the exterior. Consequently, the center parts are more magnetically coupled, and their effective inductance is higher than those on the sides. Therefore, the current tends to distribute more to the paths with lower impedance, hence, the exterior. The impedance difference increases with frequency and the current is more and more near the surface. To illustrate this phenomenon, a 2µm high and 5µm wide copper conductor is divided into 20x50 sub-conductors to calculate and represent the relative current density. Figure 3.32 presents the current distribution of 5GHz, 60GHz, and 150GHz. At 5GHz, the current is more concentrated on the right and left side of the conductor but a part of the current flows through the center.

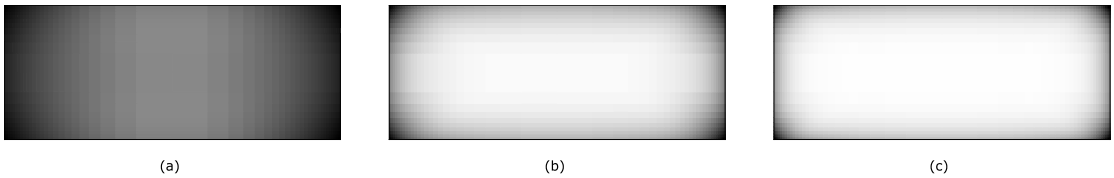


Figure 3.32 Current distribution in a 2µm thick and 5µm wide rectangular copper conductor at 5GHz (a) 60GHz (b) and 150GHz (c)

At 60GHz, the current is mainly concentrated at the corners and the current density is almost null in the center. This phenomenon is increasing at 150GHz. The consequence of this current distribution is an increase in conductor resistance and a decrease in inductance. Figure 3.33 presents the calculated resistance and inductance of a conductor with a 5x2µm section and 100µm length.

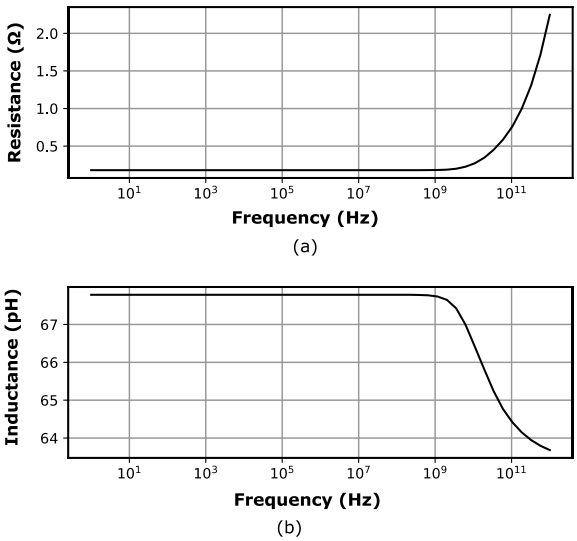


Figure 3.33 Calculated resistance and inductance of a 5x2µm and 100µm long conductor over frequency.

The resistance increases from a certain frequency which depends on the conductor's dimensions and properties and then increases with frequency. The Inductance decreases down to a limit value, of about 10% in the example. From Maxwell equations analysis, the skin depth is the distance below the surface of the conductor at which the current density is equal to 1/e times the surface current density. It shows that the effective conductive volume diminishes with frequency, hence the increase of resistance. The conductive volume is therefore proportional to the skin depth, which expression is:

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}} \tag{3.91}$$

Where ρ is the conductor resistivity, *f* is the frequency of operation, and µ the permeability of the material. It demonstrates that the effective conduction volume is inversely proportional to $\sqrt{f}$ and consequently, the resistance is proportional to $\sqrt{f}$.

Finally, from the definition of the inductor's quality factor, a consequence of the resistance's increase with frequency is that the quality factor doesn't increase proportionally to frequency but to the square root of it.

$$\mathrm{Q_{ind}} \propto \sqrt{f} \tag{3.92}$$

The inductors are designed in integrated silicon technologies, using top metallization levels that offer moderate to high thickness. Figure 3.34 depicts an example of an octagonal inductor over the silicon substrate.

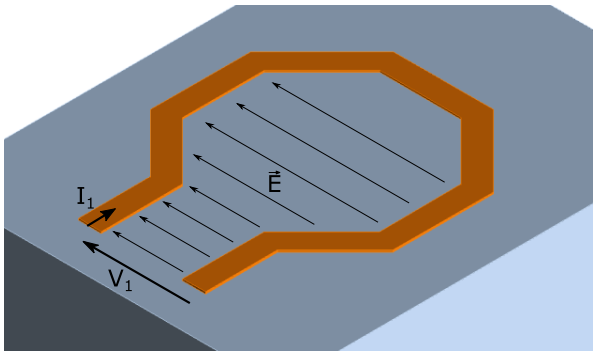
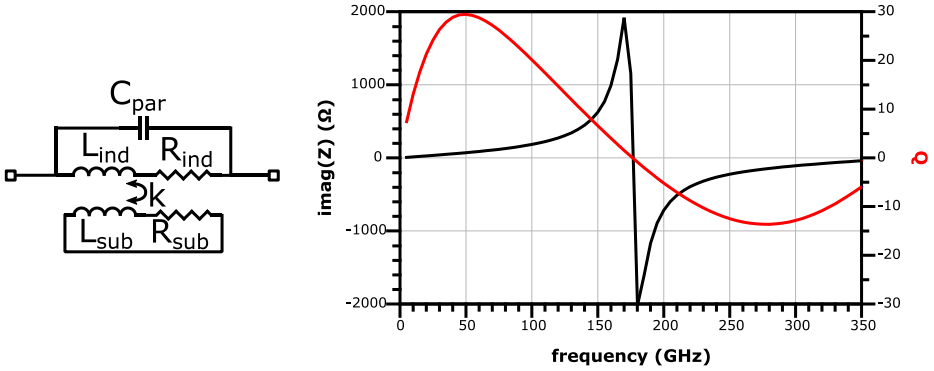


Figure 3.34 3D view of an octagonal inductor over the substrate

The inductor is designed to achieve the required inductance value with the highest quality factor. It is a two-port component and is used for its magnetic storage capability. However, when a voltage is applied between its ports, an electrical field is generated and a capacitive coupling between its ports exists in addition to the inductance. As the frequency increases, the impedance of the parasitical capacitance decreases while the inductance increases. From a certain frequency, the capacitance becomes dominant, and the inductor starts to act as a capacitor. This transition frequency is the self-resonance frequency of the inductor, it is expressed as:

$$f_{self} = \frac{1}{2\pi\sqrt{LC}} \tag{3.93}$$

Figure 3.35 (a) presents an equivalent schematic of the inductor with the parasitical capacitance.



(a) (b)

Figure 3.35 Inductor lumped component model (a) simulated reactance and quality factor of inductor from figure 3.32 over frequency (b)

Figure 3.35 (b) presents the simulated quality factor of the inductor and the reactance of the component. Below 175GHz, the reactance is inductive, while after the self-resonant frequency, the reactance becomes capacitive. The quality factor increases with frequency until a maximum; in this case at 50GHz. Then the quality factor decreases until the self-resonant frequency, where it is equal to 0. Further, the reactance changes sign as well as the quality factor and passes a certain frequency, due to higher order effects the reactance and quality factor converge again to 0.

Each inductor has a frequency response with a pic of the quality factor followed by a self-resonant frequency as presented in Figure 3.35 (b). At a given inductance value, the frequency response of the inductor can be designed to arrive at the frequency of operation. To obtain this result, the width of the conductor can be used to place the self-resonant frequency at the right frequency. However, the parasitical capacitance cannot be reduced under a minimal value that is a consequence of the used metal stack. When the inductor is relatively far from the substrate, the electrical fields close in the dielectric or the air above. When the inductor gets closer to the substrate, a greater part of the electrical field closes through the substrate that has a relative permittivity of 11.9 which is higher than the average used dielectric (~3.6) and air (~1). Consequently, the proximity of the inductor to the substrate increases the parasitical capacitance and reduces then the f_{self} and thus the maximal quality factor of the inductor. Another consequence of the inductor resonance is the variation of its effective inductance. The effective inductance increases to infinity at f_{self} and its value changes rapidly around this frequency. Consequently, for an operation near the f_{self}, the effective value of the inductor can change greatly for a small f_{self} shift which can alter the circuit in which the component is used. The other consequence of the f_{self}

is the limitation of possible inductance values at high frequency. For a given frequency of operation, the upper limit of achievable inductance is set by the f_{self}.

3.8 Capacitors

Capacitors are complementary components to the inductors and are responsible for electrical energy storage. The capacitance of a component characterizes the number of charges stored for a fixed potential:

$$q = CV \tag{3.94}$$

The capacitance between two objects can be defined similarly, as the relation between the charge difference between two objects and their potential difference.

$$q_{12} = C_{12}V_{12} \tag{3.95}$$

A capacitor is a two-element component that possesses a given capacitance. The current that flows through this component represents the charge variation between the terminals, hence

$$I = C\frac{dV}{dt} \tag{3.96}$$

The simplest case is the parallel plate capacitor as in Figure 3.36 (a), in which capacitance can be expressed as:

$$C = \frac{\varepsilon S}{d} \tag{3.97}$$

Where *S* is the plate's surface, *d* is the distance between the plates, and ε the permittivity of the material that separates the conductors. The capacitance is proportional to the surface of the two conductors and their proximity, as a stronger electrical field is generated at a given potential difference when the conductors are close. The material between the terminals also determines the strength of the electrical field.

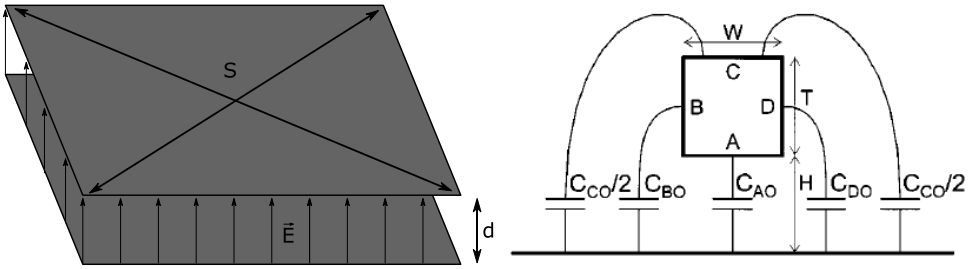


Figure 3.36 Parallel plate capacitor schematic (a), a schematic of a rectangular conductor over a ground plane and the electrical field [Stel00] (b)

The parallel plate capacitor example ignores the fields on the edge of the plates that are called fringing fields (Figure 3.36 (b)). Those fields' contribution to the capacitance becomes important as the capacitor's plate thickness increases. The calculation of these fringing fields is difficult and field solvers are convenient for arbitrary real structures. However, similarly to the inductance, empirical [Saku 83] and exact [Stel00] formulas exist in the literature for given geometries and can be used to reconstruct more complex capacitors and to understand the coupling mechanisms in a capacitor.

The capacitance opposes the variations of the voltage as the charges must be exchanged. When an alternative voltage is applied to the capacitor, from equation (3.96), the higher the frequency, the faster the charges must be exchanged, hence the increase in current. The complex admittance of the capacitor is equal to $Y_C = \mathrm{j}\omega C$.

The capacitor terminals are made from conductors that present non-infinite conductance. Thus, the capacitor presents a resistance in series with the capacitance.
The quality factor of the capacitor is equal to:

$$\mathrm{Q_{cap}} = \frac{-\mathrm{imag}(\mathrm{Z}_C)}{\mathrm{real}(\mathrm{Z}_C)} = \frac{1}{\omega C R} \tag{3.98}$$

Conversely to the inductance, the quality factor of a capacitor decreases with frequency.

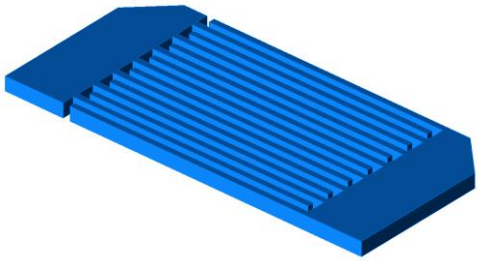


Figure 3.37 3D view of an interdigitated capacitor

Figure 3.37 depicts the 3D view of an interdigitated capacitor made with one metal layer. At mmWave and subTHz frequencies, the required capacitance values are small, and their footprints are limited. The main challenge is the modeling of these components. Simple geometries such as the one presented in the figure above are convenient to be simulated with field solvers.
Passive integrated circuit capacitors exist in two forms:

-Metal oxide metal (MOM) capacitors are like the capacitor depicted in Figure 3.37 and are made of interdigitated metal fingers. Multiple thin metal layers are commonly used and are interconnected with numerous metal vias. Such complex geometry is difficult to simulate with field solvers because of the structure complexity. The advantage of this type of capacitor is the availability as it uses standard technological metal layers and does not require additional process steps. The dielectric used to fill the capacitor is also the typical dielectric that has a relative permittivity of around 4, which is not optimal to increase the capacitance density.

-Metal insulator metal (MIM) capacitors, are made on the principle of parallel plate capacitors. A specific high permittivity dielectric is used between close metal plates to create a capacitor. It has the advantage of a simple geometry, which simplifies the characterization. In addition, the achieved quality factor is higher than the MOM capacitors and the capacitance density is high as well. However, this capacitor requires specific process steps.

Each capacitor terminal possesses a given series inductance in addition to the resistance that limits the operations of the capacitor. The impedance of an ideal capacitor decreases with frequency, however, due to the inductance, from a given frequency the inductance becomes dominant in comparison to the capacitance. This frequency is the self-resonant frequency of the capacitor and from this point, the presented impedance is not capacitive anymore. Figure 3.38 (a) presents the lumped elements capacitor model with parasitical resistance and inductance.

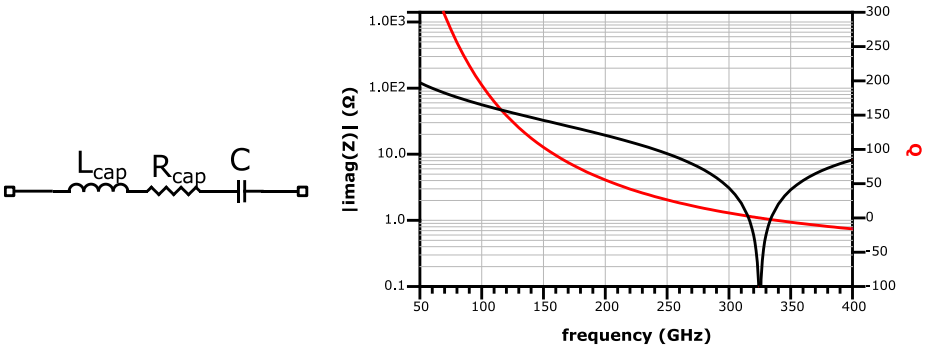


Figure 3.38 Lumped elements capacitor model (a) simulated reactance and quality factor of the capacitor from Figure 3.35

The self-resonant frequency can be expressed as:

$$f_{self} = \frac{1}{2\pi\sqrt{LC}} \tag{3.99}$$

Figure 3.38 (b) presents the reactance of the capacitor over frequency and the quality factor of the component presented in Figure 3.37. The self-resonance frequency is at 325GHz, after which the quality factor changes its sign, indicating the change of sign of the reactance. The capacitor inductance is then dominant. Similarly, to the inductor, the maximal achievable value of capacitance for a given frequency is limited, as high capacitance implies a lower f_{self}. The available technological metal stacks and the available dielectrics impose a limit on the achievable capacitor performance.

Multiple loss mechanisms exist in the capacitor. The access resistance is non-zero and the skin effect enhances the resistive losses as the frequency increases. Another non-ideality is the non-ideal oxide that constitutes the capacitors and fills the space between metal layers. Displacement currents in the dielectric induce conduction in the dielectric and thus additional losses. The conduction is frequency-dependent and is represented in an effective permittivity that has an imaginary part:

$$\varepsilon_{eff} = \varepsilon' - j\varepsilon'' = \varepsilon' - j\frac{\sigma}{\omega} \tag{3.100}$$

Where ε' is the value used in the calculation of the capacitance and ε'' represents the conduction coefficient. The equation of the admittance of a parallel plate capacitor can be applied with the complex permittivity.

$$Y_C = \mathrm{j}\omega C_{eff} = \mathrm{j}\omega\frac{\varepsilon_{eff}S}{d} = \mathrm{j}\omega\frac{\varepsilon' S}{d} + \sigma\frac{S}{d} = \mathrm{j}\omega\mathrm{C} + \mathrm{G} \tag{3.101}$$

In IC technologies, the conduction losses in the dielectrics represent a minor part of the total losses. However, the silicon substrate has a similar behavior. It has a resistivity of 15Ωcm and a permittivity of 11.9. Hence, the electrical field induces a conduction current in the substrate.
When signals are applied on capacitors and inductor's ports, the voltages are applied relatively to a reference, a local ground. Consequently, electrical coupling to this reference exists and can be represented with capacitances. A part of the coupling is done through the substrate and hence induces conduction currents that can be represented with conductance. The electrical coupling to the reference can be added to the lumped equivalent model of the capacitors and inductors. This is a π representation of the components and is presented in Figure 3.39.

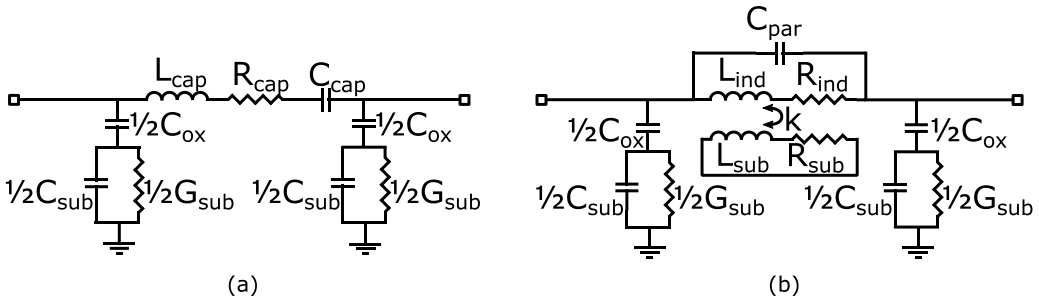


Figure 3.39 Equivalent π lumped model of a capacitor (a) and an inductor (b)

At low frequencies, the impedance of the substrate is mostly resistive, while at very high frequencies, the impedance is only capacitive. The transition of this *R//C* impedance is known as the cut-off frequency of the substrate and is equal to:

$$f_c = \frac{1}{2\pi RC} = \frac{1}{2\pi\rho\varepsilon_0\varepsilon_r} \approx 10.3GHz \qquad (3.102)$$

Where

$\rho = 15\ \Omega.cm = 0.15\ \Omega.m$ is the silicon substrate resistivity

$\varepsilon_0 = 8{,}854\times 10{-}12\ F/m$ is the free space permittivity

$\varepsilon_r = 11.68$ is the silicon relative permittivity

Below this frequency the conduction losses in the substrate are important and a patterned ground shield is commonly used under the components to prevent the electrical field from penetrating the substrate at the cost of an increased ground capacitance. However, at mmWave and sub-THz frequencies the substrate coupling is only capacitive, and the ground shields are not necessary.

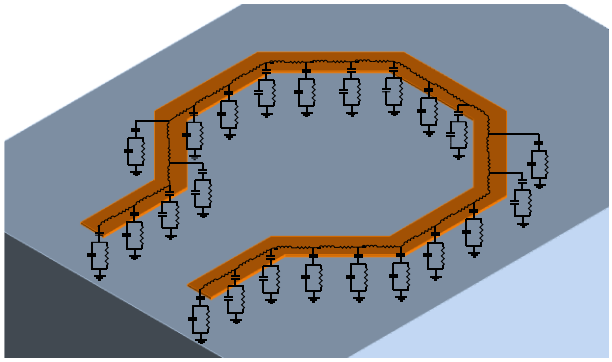


Figure 3.40 3D view of an inductor with distributed elements

As presented in Figure 3.40 the magnetic and electrical couplings are distributed along the component. The *π* representation is a simple lumped model that is correct at very low frequencies and does not reflect the component behavior passed a certain frequency. The model can be divided into two, three, or more *π* networks up to infinitesimal *π* networks that represent the distributed behavior of the electrical and magnetic couplings.

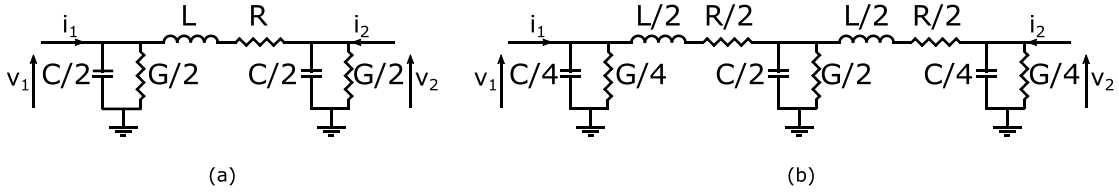


Figure 3.41 π-model (a) and double-π-model (b) of an inductor

The analysis of the accuracy of a one *π* compared to a 2*π* model is presented in Appendix D.

One can define the frequency $f_{1\%}$ at which the error of the single-π approximation to a two-π network is equal to 1%.

$$f_{1\%} = \frac{1}{10\pi}\sqrt{\frac{2}{LC}} \approx \frac{1}{22\sqrt{LC}} \tag{3.103}$$

Equation (3.103) indicates that a π model is correct with less than 1% error up to 1/22 of the *LC* resonance frequency.

Consequently, the model of a component must be divided into multiple π networks to ensure that the model is correct at the highest frequency of operation. Equation (3.103) can be used recursively. A 2π model is correct up to $2f_{1\%}$, a 4π model, up to $4f_{1\%}$ etc.

Finally, for a maximal operation frequency $f_{op,max}$, to present less than 1% error, the model must be divided into at least n_{min} π networks, defined as:

$$n_{min} = \left\lfloor \frac{f_{op,max}}{f_{1\%}} \right\rfloor + 1 \tag{3.104}$$

The lumped element representation of components like capacitors or inductors is useful when the distributed behavior is not important and when the simple π or double π is sufficient. For higher frequencies, the lumped analysis becomes difficult because of the large number of required networks to capture the high-frequency effects.

Another approach is to consider the infinitesimal distributed behavior and study how a signal propagates through the conductor.

3.9 Transmission Lines

The transmission line approach considers the voltage and current variation along the length of the conductor. It is made of at least two conductors, at least one signal conductor, and one reference conductor. Figure 3.42 (a) presents the 3D view of a conductor over a ground plane beneath it. The voltage reference follows the signal along the line.

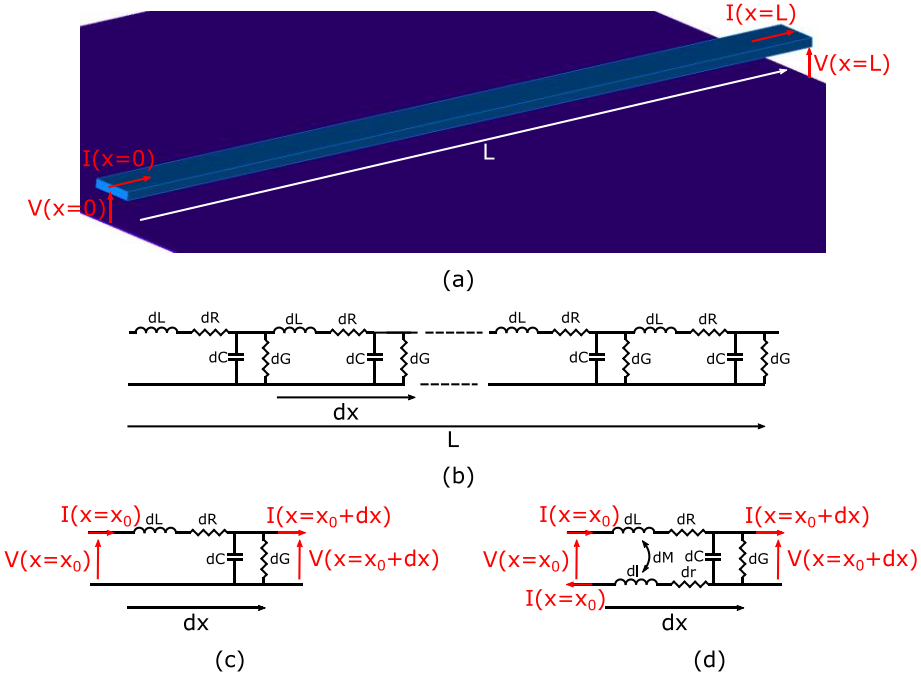


Figure 3.42 3D view of a two-conductor transmission line (a), distributed electric model of a transmission line (b), section of a transmission line electric model (c), section of a transmission line electric model with ground conductor impedance (d)

Figure 3.42 (b) presents the distributed electric model of the line, with infinitesimal lumped components that represent an infinitesimal line section of length *dx*. One section of it is presented in Figure 3.42 (c). The voltages and currents across the section terminations are related by the following equations:

$$I(x = x_0 + dx) = I(x = x_0) - V(x = x_0 + dx)(j\omega dL + dR) \quad (3.105)$$

$$V(x = x_0 + dx) = V(x = x_0) - I(x = x_0 + dx)(j\omega dC + dG) \quad (3.106)$$

Figure 3.42 (d) presents an infinitesimal section with a non-ideal ground conductor and the magnetic coupling between the signal and ground conductors.

$$I(x = x_0 + dx) = I(x = x_0) - (j\omega(dL + dl - 2dM) + dR + dr)V(x = x_0 + dx) \quad (3.107)$$

$$V(x = x_0 + dx) = V(x = x_0) - (j\omega dC + dG)I(x = x_0 + dx) \quad (3.108)$$

In the presence of a non-ideal ground conductor, the effective inductance and resistance of the signal conductor depends on the properties of the ground as follows:

$$dL_{\text{eff}} = \mathrm{d}L + \mathrm{d}l - 2\mathrm{d}M \quad (3.109)$$

$$d\mathrm{R}_{\text{eff}} = dR + dr \quad (3.110)$$

It implies that the ground conductor and its proximity to the signal conductor impact the transmission line characteristics and must be designed with caution. However, the 3.42 (c) model can be used using effective inductance and resistance.

The infinitesimal components *dL*, *dC*, *dR*, and *dG*, can be respectively expressed as the inductance *L*, capacitance *C*, resistance *R*, and conductance *G* per unit length multiplied by the infinitesimal length of a section dx. Hence, Equations (3.105) and (3.106) can be transformed by differentiating into the following equations:

$$\frac{\partial I(x)}{\partial x} = -(j\omega L + R)V(x) \quad (3.111)$$

$$\frac{\partial V(x)}{\partial x} = -(j\omega C + G)I(x) \quad (3.112)$$

Then by injecting (3.111) and (3.112) into each other, the result is the complex telegraphist's equations of the transmission line:

$$\frac{\partial^2 I(x)}{\partial x^2} = I(x)(j\omega C + G)(j\omega L + R) = \gamma^2 I(x) \quad (3.113)$$

$$\frac{\partial^2 V(x)}{\partial x^2} = V(x)(j\omega C + G)(j\omega L + R) = \gamma^2 V(x) \quad (3.114)$$

Where γ is the propagation constant:

$$\gamma = \alpha + \mathrm{j}\beta = \sqrt{(j\omega C + G)(j\omega L + R)} \quad (3.115)$$

The real part *α* and imaginary part *β* represent respectively the attenuation and phase constants of the transmission line.

Solutions to equations (3.111) and (3.112) are:

$$V(x) = V^+ e^{-\gamma x} + V^- e^{+\gamma x} \quad (3.116)$$

$$I(x) = I^+ e^{-\gamma x} + I^- e^{+\gamma x} \quad (3.117)$$

Where V^+ and I^+ are the voltage and current of the waves propagating forwards and V^- and I^- are the waves propagating backward as both are solutions of the differential equations. They are defined by the boundary condition at *x=0*.

The characteristic impedance Z_c of the transmission line is defined as:

$$Z_c = \frac{V^+}{I^+} = -\frac{V^-}{I^-} \quad (3.118)$$

Using (3.118) in (3.111) and (3.112) allows us to solve the equations for Z_c:

$$\mathrm{Z_c} = \sqrt{\frac{j\omega L + R}{j\omega C + G}} \tag{3.119}$$

In the case of a low-loss transmission line, i.e., when $\omega L >> R$ and $\omega C >> G$, Z_0, and γ can be simplified as:

$$\mathrm{Z_c} \approx \sqrt{\frac{L}{C}} \tag{3.120}$$

and

$$\gamma = \sqrt{(j\omega C)(j\omega L)}\sqrt{\left(1+\frac{G}{j\omega C}\right)\left(1+\frac{R}{j\omega L}\right)}$$
$$\approx \sqrt{\mathrm{LC}}\left(j\omega + \frac{1}{2}\left(\frac{R}{L}+\frac{G}{C}\right)\right) \tag{3.121}$$

Hence, the two propagation constants can be approximated as:

$$\alpha \approx \frac{1}{2}\sqrt{\mathrm{LC}}\left(\frac{R}{L}+\frac{G}{C}\right) \tag{3.122}$$

$$\beta \approx \omega\sqrt{\mathrm{LC}} \tag{3.123}$$

For sinusoidal signals, the signal power can be calculated at each point of the line with:

$$P(x) = |V(x)I^*(x)| = P_0 e^{-2\alpha x} \tag{3.124}$$

Where P_0 is the signal power at $x=0$. Hence, the signal power is attenuated by a factor $e^{-2\alpha x}$ over a length x in the transmission line.

The electrical length is equal to βx and to normalize the losses of the line to the electrical length, the quality factor of the line is defined as:

$$Q_{line} = \frac{\beta}{2\alpha} \approx \frac{1}{\frac{R}{\omega L}+\frac{G}{\omega C}} = \frac{1}{\frac{1}{Q_L}+\frac{1}{Q_C}} \tag{3.125}$$

The quality factor of the transmission line is equal to the quality factor of the parallel L and C of the line.

The phase velocity v_p of the transmission line is related to the phase constant as follows:

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{3.126}$$

For a signal of frequency f, the wavelength in the transmission line is:

$$\lambda = \frac{v_p}{f} = \frac{c}{f\sqrt{\varepsilon_{eff}}} = \frac{2\pi}{\beta} = \frac{1}{f\sqrt{LC}} \tag{3.127}$$

Therefore, knowing the wavelength in free space, given that the materials are non-ferromagnetic (i.e., the relative permeability $\mu_r=1$), the effective permittivity of the transmission line is equal to:

$$\varepsilon_{\mathrm{eff}} = \left(\frac{c}{v_p}\right)^2 = c^2 LC \tag{3.128}$$

The equations (3.116) and (3.117) that relate the voltages and currents of the line can be represented in a matrix form:

$$\begin{bmatrix} V(x) \\ I(x) \end{bmatrix} = \begin{bmatrix} e^{-\gamma x} & e^{\gamma x} \\ \frac{1}{Z_c} e^{-\gamma x} & -\frac{1}{Z_c} e^{\gamma x} \end{bmatrix} \begin{bmatrix} V^+ \\ V^- \end{bmatrix} \quad (3.129)$$

The TL can be represented with a two-port network using the cascade ABCD matrix. From (3.129), the ABCD matrix of a TL is:

$$ABCD = \begin{bmatrix} cosh(\gamma x) & Z_c sinh(\gamma x) \\ \frac{sinh(\gamma x)}{Z_c} & cosh(\gamma x) \end{bmatrix} \quad (3.130)$$

When the transmission line of length *l* is terminated with a load Z_L the impedance seen at the other end of the line can be calculated using (3.130).

$$Z_{in} = Z_c \frac{Z_L + Z_c tanh(\gamma l)}{Z_c + Z_L tanh(\gamma l)} \quad (3.131)$$

For a low-loss line, Z_{in} is approximately equal to:

$$Z_{in} \approx Z_c \frac{Z_L + jZ_c tan(\beta l)}{Z_c + jZ_L tan(\beta l)} \quad (3.132)$$

There are special cases for the TL impedance:

$Z_L=Z_c$

$$Z_{in} = Z_c = Z_L \quad (3.133)$$

When the characteristic impedance of the line equals the load impedance, the impedance seen at the end of the line is equal to $Z_C=Z_L$, independently of the line's length.

$\beta l = \pi$

$$Z_{in} = Z_L \quad (3.134)$$

The load impedance is unmodified by the TL. It represents a TL of length *λ/2*. Consequently, any length multiple of *λ/2* is transparent and the impedance seen at the end of the line is the load impedance Z_L.

$\beta l = \pi/2$

$$Z_{in} = \frac{Z_0^2}{Z_L} \quad (3.135)$$

The impedance seen at the end of the TL is inversely proportional to the load impedance. It represents a line of length *λ/4*. Thus, when a TL is terminated with a short, i.e., *$Z_L=0$*, an infinite impedance will be seen at the other end of the line. It transforms a short into an open circuit and vice versa. It is commonly used to bias active circuits.

$Z_L=0$

$$Z_{in} = jZ_c tan(\beta l) \quad (3.136)$$

A shorted transmission line acts as a shorted inductance when the line is short, then the impedance becomes successively capacitive and inductive with a period of *λ/4*. The impedance is periodic and changes sign with frequency (or line length) conversely to a lumped inductance or capacitance. A TL has a particular impedance frequency response.

$Z_L=+\infty$

$$Z_{in} = -j \frac{Z_c}{tan(\beta l)} \quad (3.137)$$

When the line is open at one end, the impedance seen from the other end is first capacitive and then has a similar behavior to a shorted line, alternating capacitive and inductive behavior.

To illustrate the special cases of TL impedance, Figure 3.43 presents the calculated impedances of a TL with a characteristic impedance of 50Ω using (3.132) for different loads as a function of the line length.

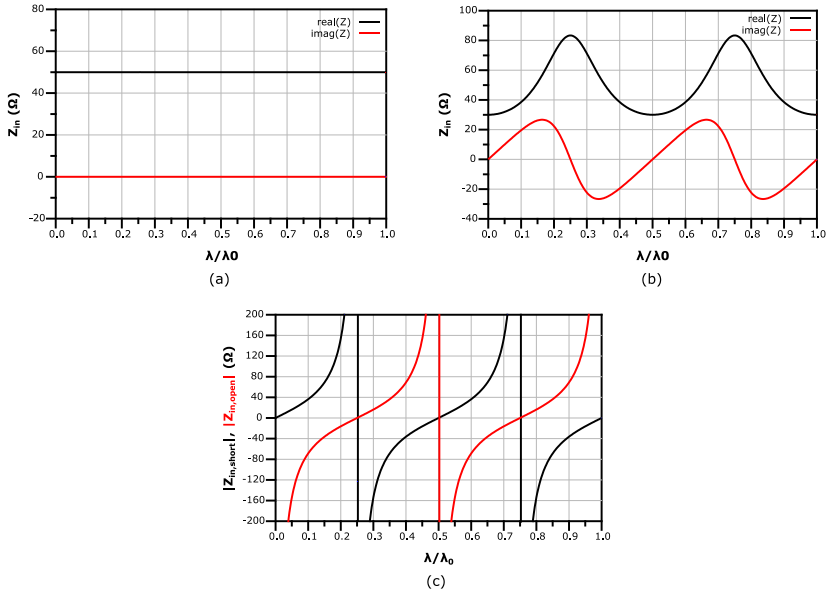


Figure 3.43 Calculated input impedance of a 50Ω TL as a function of electrical length for different load impedance: 50Ω (a), 30Ω (b), open and short (c)

The transmission lines can be implemented in different geometries. The IC technologies offer back-end-of-line (BEOL) metal layer stacks, which are used to realize RF passive components. The designs are constrained by these available metals. Figure 3.44 presents a cross-section view of the 28nm FD-SOI process BEOL composed of 8 copper metal layers and one aluminum layer. 6 Thin metal layers are available for short interconnections, usually used for active device accesses. Two thick metal layers are available for larger interconnections and RF component design. The additional aluminum layer is used for pad design and RF purposes as well. The height of the stack and the vertical distance between metals is imposed and the only available degrees of freedom are in the horizontal plane.

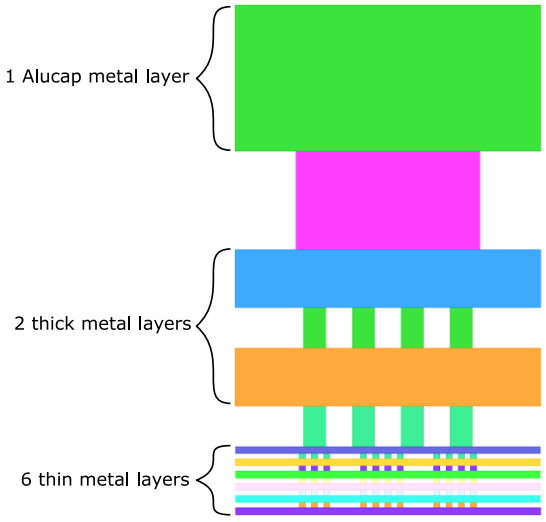


Figure 3.44 Cross-section view of 28nm FD-SOI process BEOL 8 metal layers stack

The usually used transmission lines in IC technologies are presented in Figure 3.45.

- Microstrip TL presented in 3.45 (a) is composed of a rectangular conductor for the signal and a plane beneath it for the reference ground. It is a simple and convenient geometry for design as the top metals conduct the signals and the bottom is used as the common ground plane. However, the distance between the conductor and the ground plane is fixed, hence, the only available parameter is the width of the line. To design a microstrip transmission line with a specific characteristic impedance, only one solution exists, which doesn't allow any optimization regarding the losses of the line. In addition, the achievable characteristic impedances are bounded by the technology design rules that impose a maximal and minimal conductor width.
- The coplanar waveguide (CPW) transmission line is presented in Figure 3.45 (b) and is composed of a signal conductor surrounded by two conductors for the reference ground. It allows a large range of characteristic impedances as two parameters are available, the distance between the signal and ground conductors and the width of the signal conductor. Hence, for each Z_c multiple solutions exist and better performances can be achieved in some cases compared to a Microstrip. However, two distinct ground planes are disconnected, and the discontinuity can alter the behavior of other components, such as other TLs that are perpendicular to the CPW or two active circuits on each side of the CPW that share the same reference.
- The grounded coplanar waveguide (CPWG) transmission line is presented in Figure 3.45 (c) and is a hybrid between the Microstrip and CPW TLs. It has the disadvantage of the Microstrip relative to the achievable values of Z_c. However, the electrical fields are closed all around the conductor in comparison to Microstrip where parasitical electrical couplings with other circuit components can appear. In addition, the stringent IC design rules impose a certain minimal metal density on all layers that Microstrip cannot fulfill. Hence, the CPWG is usually used in IC circuits for isolation and density compliance reasons. The side ground walls can be pushed far enough that the capacitive coupling to them is negligible compared to the ground beneath the line, allowing for a behavior close to the Microstrip.

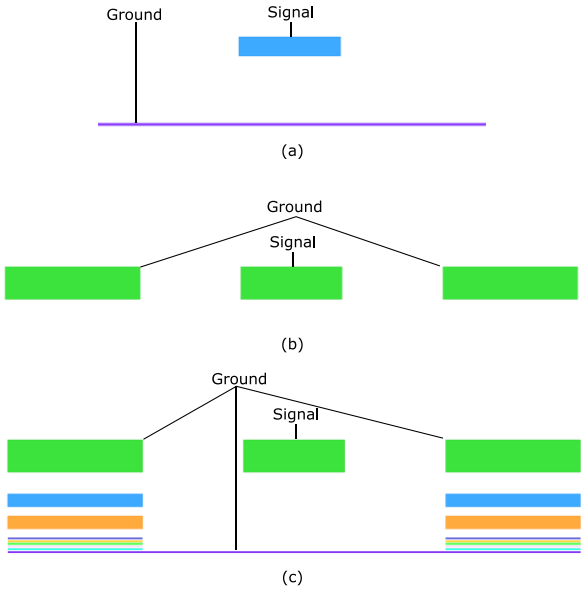


Figure 3.45 Cross section of different transmission line types that can be implemented in IC technologies: Microstrip (a), CPW (b), and CPWG (c)

The parameters of a transmission line can be extracted from the S-parameters of the line as follows:

$$Z_c = Z_0 \sqrt{\frac{(1+S_{11})(1+S_{22}) - S_{21}S_{12}}{(1-S_{11})(1-S_{22}) - S_{21}S_{12}}} \tag{3.138}$$

Where Z_0 is the ports' reference impedance,

$$\gamma = \frac{1}{l} acosh\left(\frac{(1+S_{11})(1-S_{22}) - S_{21}S_{12}}{S_{21}+S_{12}}\right) \tag{3.139}$$

Where l is the length in meters of the simulation or measured TL.

The L, R, C, and G per unit length parameters can be extracted from Z_c and γ with the following equations:

$$L = \frac{Imag(Z_c\gamma)}{\omega} \tag{3.140}$$

$$R = Real(Z_c\gamma) \tag{3.141}$$

$$C = \frac{Imag(\gamma/Z_c)}{\omega} \tag{3.142}$$

$$G = Real(\gamma/Z_c) \tag{3.143}$$

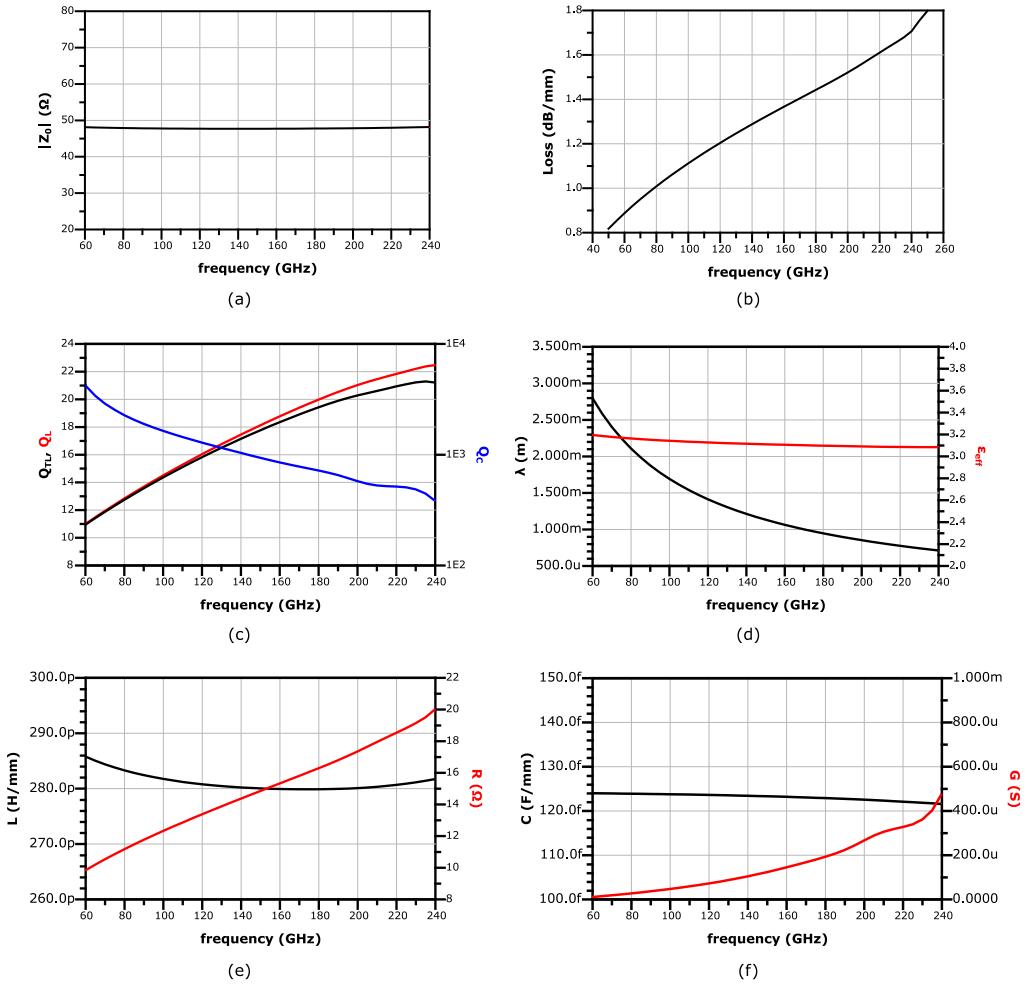


Figure 3.46 Simulated parameters of the CPWG TL presented in Figure 3.46(c)

Figure 3.54 presents the simulated parameters of the 200µm long CPWG TL presented in Figure 3.45(c). For a characteristic impedance of 47Ω, it has *1.5dB/mm* losses at 200GHz. The quality

factor of the TL is limited by the quality factor of the inductance. The effective permittivity in the TL is about 3.1 at 200GHz.
The wavelength at 200GHz is around 850µm which corresponds to 1.3dB of losses. Therefore, a *λ/4* TL that is commonly used for bias feed or power couplers and splitters, has 0.35dB of losses.

3.10 Conclusion

This chapter presented the network analysis formalism to represent active and passive components. Analysis of any two-port circuit consists of evaluating the maximum available gain of the two-port and the port impedances. The stability can also be evaluated from the two-port parameters. The transistor linear model has been presented and their main metrics have been introduced. The frequency limitation and its relation to the transistor geometry have been discussed. Then the passive components that constitute the RF circuits, inductors, capacitors, and transmission lines have been defined and their main properties have been described. The principle of designing such active and passive components was presented in this chapter. All the necessary building blocks for RF circuits have been presented.
While at frequencies low relative to the f_{max}, the conjugate matching of transistors is sufficient to design amplification stages.
Because of the limited performance of transistors at frequencies above 100GHz, it is necessary to implement gain-boosting techniques to increase the maximum available gain.

3.11 References

[Moor65] G. E. Moore, “Cramming more components onto integrated circuits”, Electronics, vol. 38, no. 8, pp. 114–117, April 1965. pages 16, 19

[Wiki1] Wikipedia, “Moore’s law— Wikipedia, The Free Encyclopedia”

[irds22] International Roadmap for Devices and Systems, 2022, irds.ieee.org

[Voin13] S. Voinigescu, “High-Frequency Integrated Circuits”, 2013.

[Cath17] A. Cathelin, “Fully Depleted Silicon on Insulator Devices CMOS: The 28-nm Node Is the Perfect Technology for Analog, RF, mmW, and Mixed-Signal System-on-Chip Integration,” IEEE Solid-State Circuits Magazine, vol. 9, no. 4, pp. 18–26, 2017.

[Haen06] W. Haensch et al., "Silicon CMOS devices beyond scaling," in IBM Journal of Research and Development, vol. 50, no. 4.5, pp. 339-361, July 2006

[Plane12] N. Planes et al., "28nm FDSOI technology platform for high-speed low-voltage digital applications," 2012 Symposium on VLSI Technology (VLSIT), Honolulu, HI, USA, 2012

[John65] E. Johnson, “Physical limitations on frequency and power parameters of transistors”, IRE International Convention Record, vol. 13, pp. 27–34, March 1965.

[Urte17] M. Urteaga, Z. Griffith, M. Seo, J. Hacker, and M. J. W. Rodwell, "InP HBT Technologies for THz Integrated Circuits," in Proceedings of the IEEE, June 2017

[Asif93] M. Asif Khan, A. Bhattarai, J. N. Kuznia, and D. T. Olson, "High electron mobility transistor based on a GaN-AlxGa1−xN heterojunction", Appl. Phys. (1993)

[Dick06] T. O. Dickson et al., “The invariance of the characteristic current densities in nanoscale MOSFETs and its impact on algorithmic design methodologies and design porting of Si(Ge) (Bi)CMOS high-speed building blocks,” IEEE JSSC, August 2006.

[Dorm13] B. Dormieu et al., "Revisited RF Compact Model of Gate Resistance Suitable for High-K/Metal Gate Technology," in IEEE Transactions on Electron Devices, Jan. 2013

[Rueh72] Ruehli A. E.: Inductance Calculations in a Complex Integrated Circuit Environment. IBM J. Res. Develop. Sept. 1972.

[Grov73] Grover F. W. Inductance Calculations. Dover Publications, Inc., New York, 1973.

[Piat12] Z Piatek et al., «Self inductance of long conductor of rectangular cross section", Electrical Review, 2012

[Kamo93] M KAMON et al, "FASTHENRY: A multipole-accelerated 3-D inductance extraction program", Proceedings of the 30th international design automation conference, 1993

[Saku83] T Sakurai and K Tamaru, "Simple Formulas for Two- and Three-Dimensional Capacitances", IEEE Transaction on Electron Devices, 1983

[Stel00] F Stellari et al., "New Formulas of Interconnect Capacitances Based on Results of Conformal Mapping Method", IEEE Transaction on Electron Devices, 2000

Chapter 4 Gain Boosting technics

The power gain of transistors decreases linearly with frequency and the frequency of operation is intrinsically limited by the device f_{max}. As we have seen the maximal operation frequencies of silicon components increased substantially but the sub mmWave and sub-THz frequencies above 100 GHz are at 1/2 to 2/3 of the f_{max}. The purpose of this chapter is to study and propose design technics to maximize the component performances close to its f_{max}.
First, the maximum achievable gain of an active two-port will be studied and the conditions of the two-port to achieve this gain will be presented. Then the commonly used gain boosting technic, the common source neutralization will be introduced, and a theoretical study will highlight its limitations. Next, the gain plane formalism will be presented to study the design of positive feedback embeddings to enhance the maximum available gain of active two ports. An analytical gain-boosting design methodology will be presented. Finally, the design and the measurements of a one-stage gain-boosted amplifier at 184GHz in 28nm FD-SOI will be presented.

4.1 Two port Gain boosting theory

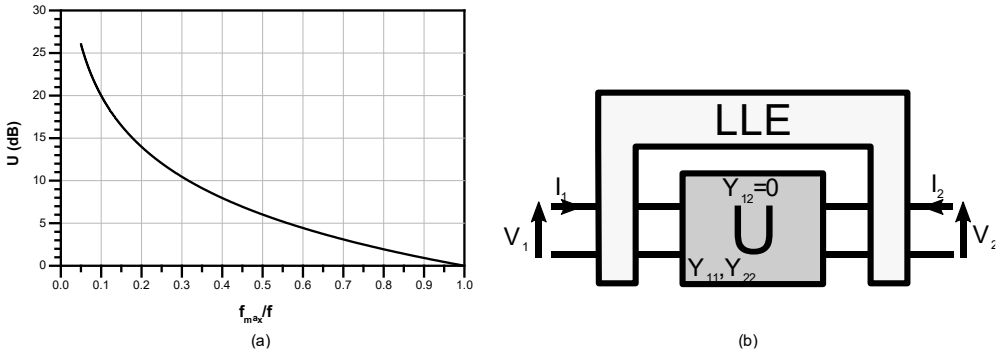


Figure 4.1 Transistor's Mason Gain as a function of f_{max}/f (a) general representation of an ATP with its intrinsic unilateral Gain and internal embedding (b)

The Mason's Gain (U) is a good indication of a transistor's performance because it considers the intrinsic losses when the feedback is null or compensated with lossless components. It gives the potential intrinsic power gain.
The transistor's Mason's Gain decreases at a known rate, the available power gain can be then deduced from the f_{max} and the operation frequency. Hence figure 4.1 (a) depicts values of U as a function of f_{max}/f. At 1/3, 1/2, and 2/3 of f_{max}, the power gain is respectively 9.5dB, 6dB, and 3.5 dB. It highlights the limitations of operations above $1/3\ f_{max}$, as it represents the gain of the intrinsic device to which must be added the interconnection and impedance matching network losses.
Any active two-port can be represented as in Figure 4.1 (b) by an internal unilateral active two-port (ATP) embedded in a linear embedding (LE) that represents the internal feedback.
Generally, at a given frequency, it is theoretically possible to increase or reduce the feedback of an active two-port (ATP) by using linear lossless embeddings (LLE). Consequently, we will demonstrate that all ATPs with the same value of U can achieve the same range of G_{ma} by changing the amount of feedback through LLEs.

4.1.1 Maximum Achievable Gain G_{max}

The fundamental equation for the study of Gain Boosting was presented first in [Sing64a] in the context of the study of ATPs:

$$\sqrt{\frac{G_{ma}}{U}} = \frac{|A - G_{ma}|}{|A - 1|} \tag{4.1}$$

with

$$\mathrm{A} = \frac{\mathrm{Y}_{21}}{\mathrm{Y}_{12}} = \frac{\mathrm{Z}_{21}}{\mathrm{Z}_{12}} = \frac{\mathrm{S}_{21}}{\mathrm{S}_{12}}$$

It shows that the maximum available gain G_{ma} is only a function of A for a given value of U. In addition, when the component is unilateral, i.e., $Y_{12}=Z_{12}=S_{12}=0$, A is infinite, and consequently $G_{ma}=U$.

From the stability factor K expression, it is also equal to infinity when the ATP is unilateral.

From the definition of G_{ma}, when $Y_{12}=Z_{12}=0$, it gives:

$$G_{ma} = \frac{|Y_{21}|^2}{4ReY_{11}ReY_{22}} = U \tag{4.2}$$

Thus, all unilateral components with the same value of U verify (4.2).

Because U is invariable under LLEs, such embeddings can increase the ATP G_{ma}. The gain is independent of $Z_{ii,}$ and Y_{ii} parameters and only depends on the ratio A, which indicates the reciprocity of the ATP. Infinity is unilaterality, and 1 is the gain symmetry.

On another side, G_{ma} only exists when $K \geq 1$ and can be expressed as:

$$G_{ma} = |A|\left(K - \sqrt{K^2 - 1}\right) \tag{4.3}$$

The right term of (4.3) is a decreasing function of K, which is equal to 1 when $K=1$.

Hence, to find the maximal value of G_{ma} at a given U, (4.3) imposes that K must be equal to 1. When $K=1$, (4.3) becomes,

$$G_{ma} = |A| \tag{4.4}$$

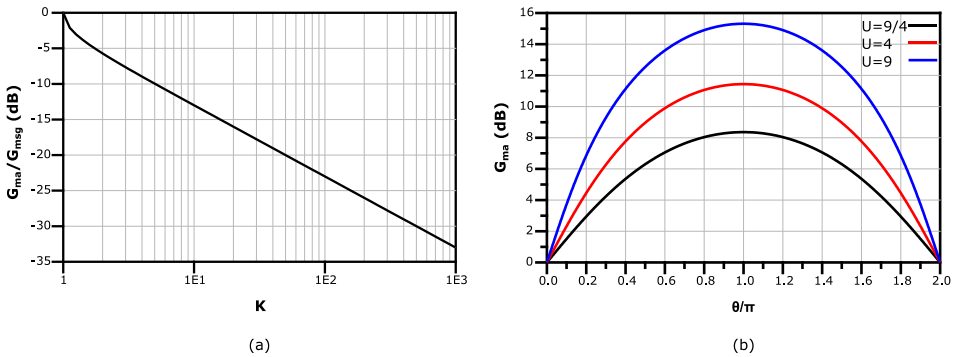


Figure 4.2 Contribution of the right factor in (4.3) as a function of K (a) G_{ma} when K=1 for different values of θ

Figure 4.2(a) shows the contribution of the right term of (4.3) to G_{ma}, the term rapidly decreases as K increases at a rate of 10dB per decade.

Next, to find the maximal achievable gain of an ATP for a given *U*, one must find the values of *A* that verify *K=1* and finally find the optimal A_{opt} to maximize G_{ma}.
Injecting (4.4) into (4.1) allows us to find all the A values that verify *K=1* for constant *U*.
It is then solved for the variable $\mathrm{U/A} = \mathrm{r}e^{j\theta}$

$$r = U(\cos(\theta)\,(1-U)+U) - U\sqrt{U-1}\sqrt{\cos^2(\theta)\,(U-1) - 2U\cos(\theta) + U + 1} \qquad (4.5)$$

$|A|$ is maximal when r is minimal thus from (4.5)

$$\theta_{opt} = \pi \qquad (4.6)$$

$$r_{\min} = U(2U-1) - 2U\sqrt{U(U-1)} \qquad (4.7)$$

The maximum achievable gain, G_{max} is defined as the maximal value that can take G_{ma} for a given *U* and can be expressed from (4.4) and (4.6) as

$$G_{max} = (2U-1) + 2\sqrt{U(U-1)} \qquad (4.8)$$

In conclusion, for an ATP with a unilateral gain *U*, the maximal achievable gain is achieved when K=1 and $\angle A = \pi$.

Figure 4.2(b) presents the values of G_{ma} when *K=1* as a function of the normalized angle of *A* for different values of *U*: *9/4*, *4*, *9* that correspond respectively to an operation at *1/3*, *1/2*, and *2/3* of f_{max}. It shows that *K=1* is not the only condition to achieve the highest possible gain, but the phase of *A* is as important, and the performance decreases quickly as the phase moves from the optimal.

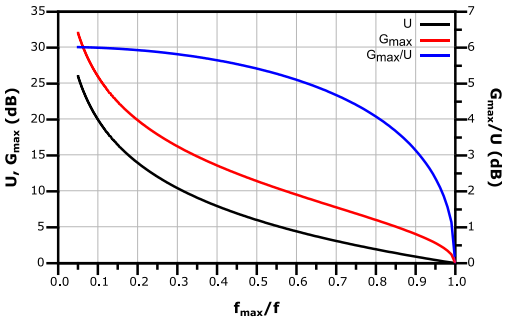


Figure 4.3 U, G_{max}, and the difference between both gains as a function of f_{max}/f

Figure 4.3 presents the values of *U*, G_{max}, and the ratio between the two gains in the function of f_{max}/f. It shows that G_{max} offers +6dB of gain in comparison to U but this gap decreases close to f_{max}. It demonstrates that an increase in the feedback of an ATP compared to a unilateral one can increase the gain of up to 6dB. At *1/2* and *2/3* of f_{max}, G_{max} is respectively equal to 11.4dB and 8.4dB which is enough to consider designing amplification circuits at these frequencies.
In conclusion, G_{max} is achieved when, $|A| = G_{max}$ and $\angle A = \pi$. Plus, the use of LLE to increase the G_{ma} of an ATP towards G_{max} enables the design at frequencies as far as 2/3 of the f_{max}.

4.2 Inductive and capacitive Common Source Neutralization

From the beginning of transistor-based RF circuit design, internal transistor feedback has been studied and techniques have been developed to neutralize it. As we have seen in the previous chapter, the intrinsic feedback elements, principally C_{gd} and to a lesser extent R_S, decrease the available power gain and deteriorate the stability. Following the work of Mason on ATPs, multiple neutralization techniques have been published [Ster55],[Chen55].

As the common source topology is among the simplest and most effective amplification topologies, especially at very high mmWave frequencies, this part focuses on the neutralization of common source amplification stages. The principle is presented in Figure 4.4 (a), it consists of presenting an element with an opposite impedance value to the C_{gd} impedance. Thus, all current going through C_{gd} is compensated with a current equal and of the opposite sign through Z_{neutr}. However, the gate access resistance R_g is an important parameter when working in the mmWave frequency range, and the achieved result is depicted in Figure 4.4(b), leading to an imperfect C_{gd} compensation. Finally, figure 4.4(c) presents the two-port matrix representation of a neutralized transistor and is used to calculate the resulting two-port parameters.

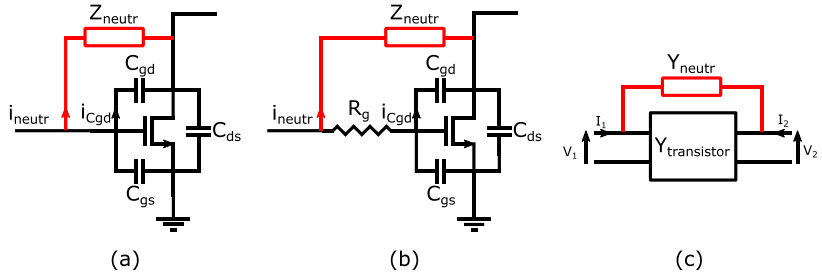


Figure 4.4 Representation of the principle of CS neutralization (a) and representation of real implementation (b). Y-Matrix two-port representation of the neutralized CS

The calculation of the equivalent matrix of a parallel embedding as in Figure 4.4(c) is straightforward with Y-matrixes, as the result is the sum of the matrixes.

$$Y_{transistor} = \begin{bmatrix} \dfrac{j\omega(C_{gd}+C_{gs})}{1+j\omega R_g(C_{gd}+C_{gs})} & -\dfrac{j\omega C_{gd}}{1+j\omega R_g(C_{gd}+C_{gs})} \\ \dfrac{g_m - j\omega C_{gd}}{1+j\omega R_g(C_{gd}+C_{gs})} & g_{ds}+j\omega(C_{ds}+C_{gd}) + \dfrac{j\omega R_g C_{gd}(g_m - j\omega C_{gd})}{1+j\omega R_g(C_{gd}+C_{gs})} \end{bmatrix} \quad (4.9)$$

$$Y_{neutr} = jy_{neutr}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (4.10)$$

Then the equivalent matrix is,

$$Y_{equ} = Y_{transistor} + Y_{neutr} \quad (4.11)$$

From the equivalent matrix, the stability factor of the new ATP as a function of the amount of neutralization $k=y_{neutr}/(\omega C_{gd})$ is:

$$\mathrm{K_{equ}} = \frac{\omega}{\omega_0}\frac{1-2k+(2-k)\frac{\omega_0}{\omega_1}+k^2\left(1+\left(\frac{\omega}{\omega_1}\right)^2\right)+2g_{\mathrm{ds}}R_g\left(1+\frac{C_{gs}}{C_{gd}}\right)^2}{\sqrt{(1-k)^2+\left[\frac{\omega}{\omega_0}(1-k)^2+k\frac{\omega}{\omega_1}\left(k\frac{\omega^2}{\omega_1\omega_0}-1\right)\right]^2}} \quad (4.12)$$

Which simplifies when $k=1$ to

$$K_{equ} = \frac{\frac{\omega^2}{\omega_1\omega_0}+1+2\frac{g_{ds}}{g_m}\left(1+\frac{C_{gs}}{C_{gd}}\right)}{\left(\frac{\omega^2}{\omega_1\omega_0}-1\right)} \approx \frac{\frac{\omega^2}{\omega_1\omega_0}+1}{\left(\frac{\omega^2}{\omega_1\omega_0}-1\right)} \quad (4.13)$$

With,

$$\omega_0 = \frac{g_m}{C_{gd}}, \omega_1 = \frac{1}{R_g(C_{gd}+C_{gs})} \quad (4.14)$$

Equation (4.13) shows that when the C_{gd} is perfectly compensated, the stability factor K becomes superior to 1 but the achieved value is not infinite and is as expected limited by the R_g. g_m/g_{ds} is

the DC intrinsic component gain which is expected to be large and makes the last numerator term in (4.13) negligible.

It shows that the achievable *K* value with neutralized C_{gd} is constant over frequency.

A can also be expressed in function of the amount of neutralization *k*:

$$A_{equ} = 1 - \frac{\omega_0}{\omega}\frac{1}{k\frac{\omega}{\omega_1} + j(1-k)} \tag{4.15}$$

Which simplifies when *k=1*:

$$A_{equ} = 1 - \frac{\omega_1\omega_0}{\omega^2} \tag{4.16}$$

Hence, the phase of *A* is equal to 180° when *k=1*. In addition, *K* is not equal to 1 in these circumstances and consequently, all conditions are not fulfilled to increase G_{ma} towards G_{max}. Neutralization does not allow to boost of the gain to G_{max}.

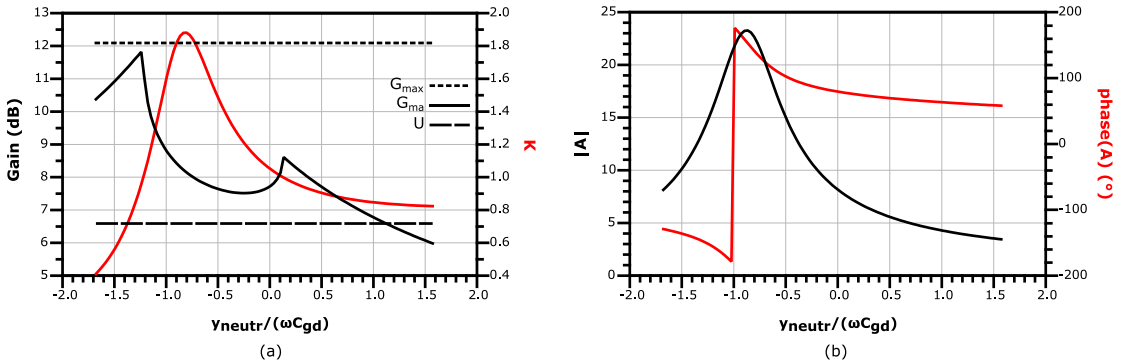


Figure 4.5 Simulation of neutralized CS's parameters at 150GHz as a function of neutralization impedance: G_{ma} and *K* (a) magnitude and phase of *A* (b)

Figure 4.5 depicts the G_{ma} and *K* as a function of the amount of neutralization. The values achieved by G_{ma} when *K* is maximal and equal to 1 are bounded respectively by *U* and G_{max}. However, these values are not attained. Because of R_g, the transistor is not completely neutralized which leads to a non-infinite *K* and a G_{ma} higher than *U*. The simulations confirm the observations made from the previous equations.

The implementation of CS neutralization requires the design of an element with an impedance of the opposite sign to C_{gd}. When considering a differential schematic, the common implementation is presented in Figure 4.6 (a) and consists of a differential CS pair where neutralization capacitances are connected on one side to a gate and the other side to the drain of the opposite branch. The equivalent circuit is depicted in Figure 4.6 (b) and allows us to calculate the equivalent circuit matrix representation. From (4.10) the Y_{neutr} component is represented by two equations relating its voltages to currents as follows:

$$\begin{cases} I_1 = Y_{neutr,11}V_1 + Y_{neutr,12}V_2 \\ I_2 = Y_{neutr,21}V_1 + Y_{neutr,22}V_2 \end{cases} \tag{4.17}$$

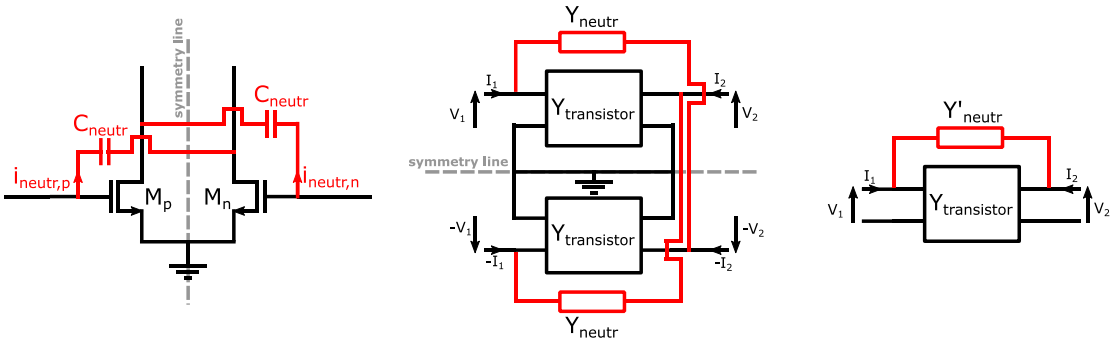


Figure 4.6 Differential capacitively neutralized CS pair schematic (a), matrix representation (b), and half equivalent matrix representation (c)

Because of the circuit's symmetry and the differential operation, it is possible to express the matrix representation of Y'_{neutr} as in Figure 4.6(c) which represents the equivalent single-ended circuit from 4.6(b). Considering the top half circuit part, the current and voltage of the second port have a negative sign. Changing the sign of these values in (4.16) provides the equivalent matrix parameters as follows:

$$Y'_{neutr} = \begin{bmatrix} Y_{neutr,11} & -Y_{neutr,12} \\ -Y_{neutr,21} & Y_{neutr,22} \end{bmatrix} = jy_{neutr} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = j\omega C_{neutr} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad (4.18)$$

Comparing equation (4.18) to the single-ended neutralization equation (4.10) shows that the neutralization capacitance operates as a negative capacitance in the G_{ma} expression. However, regarding the admittances, Y_{11} and Y_{22} of the new ATP, the C_{gd} is not canceled, but the C_{neutr} adds up to the port's capacitance. Capacitive neutralization presents benefits regarding stability and gain but at the cost of increasing equivalent capacitances at input and output. This leads to an increase in the difficulty of matching ATP's impedances. At high frequencies, the size limit of transistors and thus the output power is due to the impedance's capacitances, and consequently, neutralization limits the usable transistor size.

The design of differential amplifiers has the benefit of providing virtual ground and enabling capacitive neutralization by the facility to obtain equivalent negative impedances, but it is not always possible or convenient to generate differential signals at high mmWave frequencies.

Hence, the remaining solution for single-ended schemes is inductive neutralization. While capacitive neutralization presents an impedance of the same nature as C_{gd}'s and thus follows the C_{gd} impedance over a large frequency range, the inductive method works in a narrow band. C_{gd} resonated only at one frequency of interest because of the nature of inductance impedances. In addition, inductive neutralization necessitates dedicating a silicon area for the inductance that can be relatively large below 100GHz and a series large capacitance as a DC block. While required capacitances are small and can be implemented using transistors. However, at subTHz frequencies, the required inductances are small, and this design choice becomes viable. For better modeling accuracy, transmission lines can be used to implement the neutralization impedance.

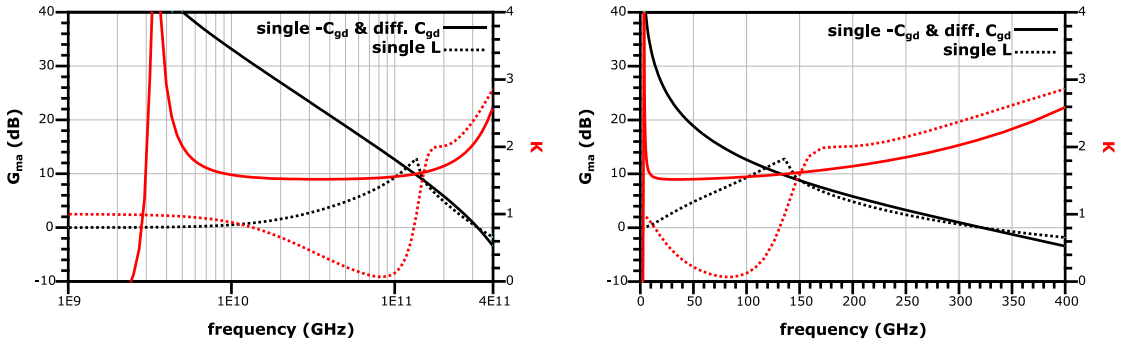


Figure 4.7 Simulations of a CS topology based on a 28nmFD-SOI transistor (*L=30nm*, *W=16µm* biased at J_{opt}*=0.3mA/µm*) with different neutralization methods.

Figure 4.7 presents the simulations of a CS topology using 28nmFD-SOI transistors (*L=30nm*, *W=16µm* biased at J_{opt}*=0.3mA/µm*). It compares the differential pair capacitive neutralization, the single inductive neutralization, and the theoretical single-ended neutralization with a negative capacitance. The neutralization has been done at 150GHz with impedances that compensate C_{gd} as previously. The figures compare the G_{ma} and K over frequency in log (a) and linear (b) frequency scales. As expected, the capacitive neutralization achieves the same result as the negative capacitance over a large range of frequencies as predicted by (4.13). Below a certain low frequency, the neutralization is not effective, and the K becomes negative. To operate at these lower frequencies, this resonance can be reduced by affecting the pulsations ω_0 and ω_1 (Equation (4.14)). The inductive neutralization achieves the same performance at 150GHz regarding G_{ma} and K, but it differs at other frequencies. Below 150GHz the gain decreases quickly and the stability factor falls close to 0 in a large frequency band. The gain improvement is also quite narrowband compared to the capacitive neutralization.

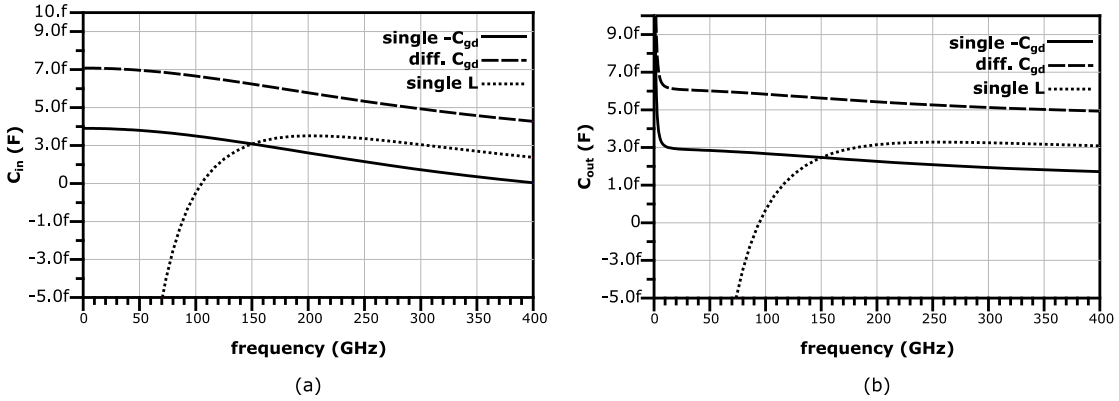


Figure 4.8 Simulated input (a) and output (b) parallel capacitance over frequency for different neutralization methods

Figure 4.8 presents the equivalent parallel capacitance of the port's impedances for the different neutralization methods. According to the theory, capacitance is compensated with negative capacitance and inductive neutralization while differential capacitance doubles the C_{gd} contribution.

From the theoretical study of the C_{gd} compensation (figure 4.5), when C_{gd} is overcompensated, the stability factor decreases and G_{ma} admits a maximum when $K=1$. Hence, over-neutralization consists of increasing the neutralization amount ($k>1$) to increase G_{ma}. Figure 4.9 presents the simulation of G_{ma} and K over frequency for an over-neutralization towards maximum gain at

150GHz. Similarly, theoretical negative capacitance, cross-coupled capacitance, and inductance techniques are compared.

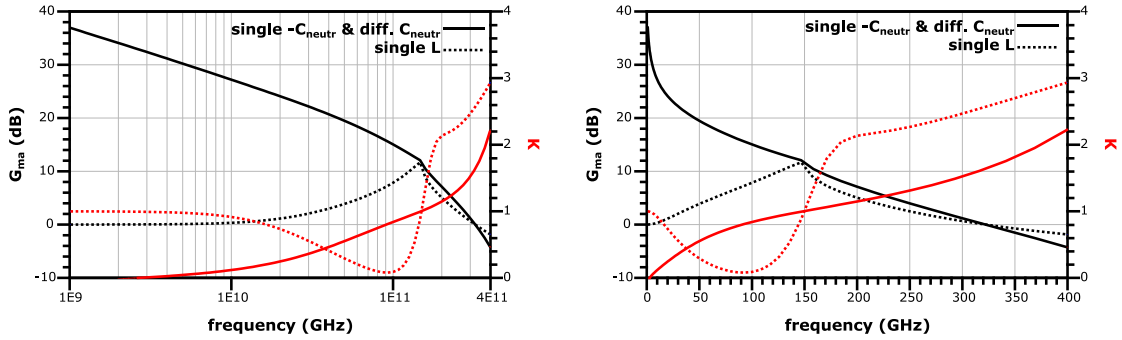


Figure 4.9 Simulations of a CS topology based on a 28nmFD-SOI transistor (L=30nm, W=16µm biased at J_{opt}) with different over-neutralization methods

At 150GHz, *K=1*, the G_{ma} is increased and presents the same value with all techniques. However, the stability factor decreases also for the capacitive neutralization this time, falling under 1 before 150GHz and approaching 0 at low frequencies. Hence, in the case of over-neutralization, both techniques cause potential instability below the frequency of operation.

The common source neutralized topology is commonly used in the design community at mmWave frequencies [Chan10] and even at sub-THz bands. The paper [Simi21] presents a 190GHz power amplifier with over-neutralization to enhance the available gain. It studies the impact of the non-infinite quality factor of the neutralization capacitance and proposes to use this additional series resistance in the design process. The schematic of the PA is presented in Figure 4.9.

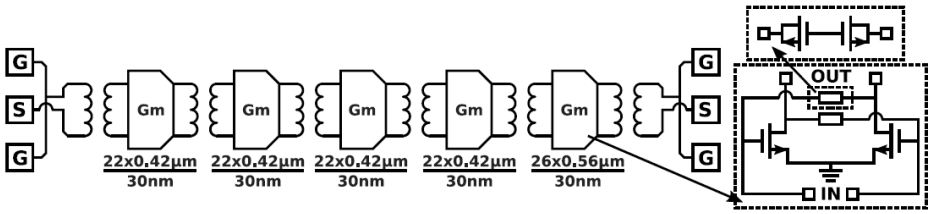


Figure 4.10 190GHz over-neutralized PA in 28nm bulk CMOS [Simi21]

The amplifier achieves a total gain of *14.5dB* with 5 stages or *2.9dB* per stage. It has a bandwidth of 14.5GHz, an *OIP1* and P_{sat} of respectively *-2.4dBm* and *1.5dBm*.

The paper [Wang14] presents a PA in 32nm bulk integrated into a transceiver operating at 210 GHz. The schematic of the amplifier is presented in Figure 4.11. It presents a gain of *15dB* with 3 stages or *5dB* per stage. It has a bandwidth of 14GHz, an *OIP1* and *Psat* respectively of *2.7dBm* and *4.6dBm*.

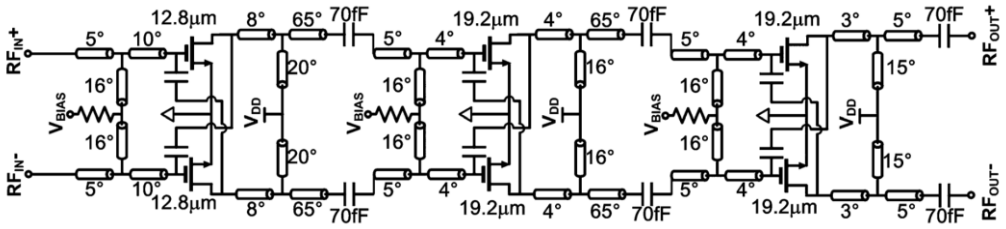


Figure 4.11 210GHz over-neutralized PA in 32nm bulk CMOS [Wang14]

Inductive neutralization has also been implemented in the literature, the paper [Mome11] reports a three-stage single-ended amplifier at 104GHz. The amplifier exploits inductive feedback to produce over-neutralization and increase the available gain to operate near f_{max}.

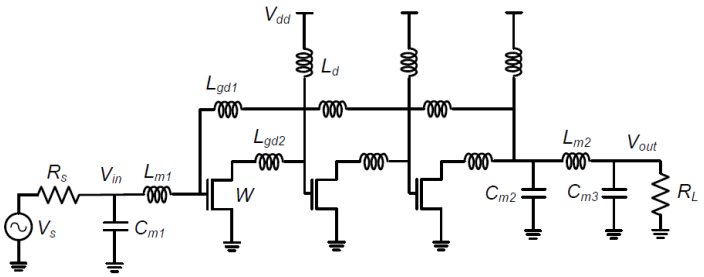


Figure 4.12 107GHz amplifier with inductive feedback in 130nm bulk CMOS [Mome11]

This implementation does not use DC blocks in series with feedback inductances to bias separately drains and gates. It presents a gain of *12.5dB* with 3 stages or *4.2dB* gain per stage with a bandwidth of 5.4GHz. It has an *OIP1* and P_{sat} respectively of *-1dBm* and *2.3dBm*.
One can observe from the published works that gain per stage is limited and is the key challenge to address. Many stages require numerous matching networks which reduces the overall bandwidth and efficiency. In addition, the saturation power is limited because of the use of small transistors.

4.3 Gain planes

In this section, we generalize the approach that has been taken to the case of CS topology can be generalized to any ATP. The objective is to use a formalism that allows designing an LLE for any ATP to increase its gain to G_{max}. Results from section 4.1 can be applied to any ATP that is characterized by a unilateral gain *U*. The graphical representation of ATPs in gain planes with gain and stability circles has been introduced already in [Sing64b] and has been later investigated for optimization of silicon technology transistors for sub-THz applications [Amak14], [Amak16], [Bame17]. It is a powerful tool to graphically estimate LLE elements' effect on an ATP regarding gain and stability. At a given intrinsic unilateral gain *U*, the values of *A* that produce *K=1* and thus represent the points at the boundary of unconditional stability, are given by equation 4.5 and form an ellipsis. An example of such an ellipsis is depicted in Figure 4.13.

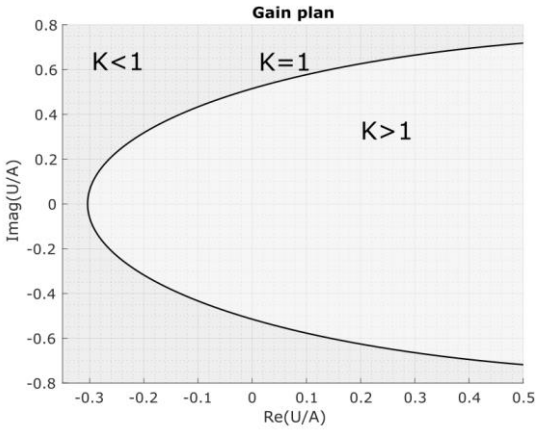


Figure 4.13 Stability circle in the Gain plane

The graphical formalism uses the inverted normalized gain chart, the *(Real(U/A), Imag(U/A))* space because it offers the unconditionally stable region inside of the chart. The other points are

outside of this region and expand to infinity. Because the design objective is to increase the gain while maintaining stability, these are the points of interest, and this representation is more convenient. On the contrary, the representation in the *(Real(A/U), Imag(A/U))* plane inverts the stable and unstable regions and the point of interest resides outside of the visible chart.
Any ATP with a given *U* can be placed in this plane from the *A* value and thus its stability can be graphically estimated. Additional circles of constant *K* can be drawn but in the following considerations, only the stability boundary is studied.
The gain circles represent all the points in the space that have the same value of G_{ma}. The equation (4.1) can be transformed with the quantity $k=U/G_{ma}$ which is the inverse normalized G_{ma} regarding *U*. It gives the following equation:

$$y^2 + \left(x - \frac{Uk(U-1)}{(U^2-k)}\right)^2 = \frac{kU^2[U^2(1-k)+k(k-U)]}{(U^2-k)^2} \qquad (4.19)$$

With $\frac{U}{A} = \mathrm{x} + \mathrm{jy}$

It represents a circle of the center $\left(\frac{Uk(U-1)}{(U^2-k)}, 0\right)$ and radius $\frac{U\sqrt{k\sqrt{U^2(1-k)+k(k-U)}}}{(U^2-k)}$.

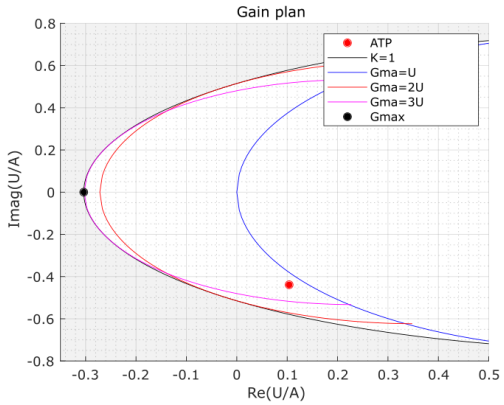


Figure 4.14 Constant Gain circles in the Gain plane

Figure 4.14 represents the Gain Plane with the stability and gain circles for G_{ma} equal to one, two- and three times *U* respectively. In addition, from the previous results, the point representing the G_{max} can be drawn from conditions (4.6) and (4.7). As an example, an ATP is represented. Just by graphical estimation, the chart shows that the ATP is unconditionally stable and has a G_{ma} between *U* and *2U*. Because such a graphical representation is constructed thanks to a Z (or Y, or any equivalent matrix) matrix, it represents the performances of the ATP for a unique set of frequency, size, and bias of the circuit that is represented by the ATP. Consequently, a new gain plane must be drawn for each set of these parameters.

Graphical design of an LLE

The graphical representation in the Gain Plane of the stability and gain circles enables us to graphically estimate the variation in performances of an ATP when lossless passive elements are added to it by placing the point of the equivalent ATP in the plan.
When considering an ATP, it has two inputs and a reference (two references which are shorted in the formalism of matrix representation of a two-port), lumped elements with two ports can be connected to the ATP in 3 ways:

- In series

- In parallel,
- In degeneration (as CS degeneration impedance),

Because one only considers lossless embeddings, their impedances are purely imaginary, positive, or negative.

A combination of these elements can constitute an LLE for an ATP, but we can study independently their effect and successfully add them.

Figure 4.15 presents an example of an embedding.

An active two-port *A* is embedded step by step and the effect of each embedding element is represented in the gain plane:

-*A* is first degenerated with an element e_1, which gives an equivalent ATP A_1,

-Then with an additional element e_2 added to A_1, it gives A_2,

-and so on.

We can estimate step by step the effect of a complex LLE on an ATP. A representation in the gain plane of each step allows us to graphically evaluate the gain and stability of each step ATP. In other words, it is graphically possible to design an LLE for a desired gain point.

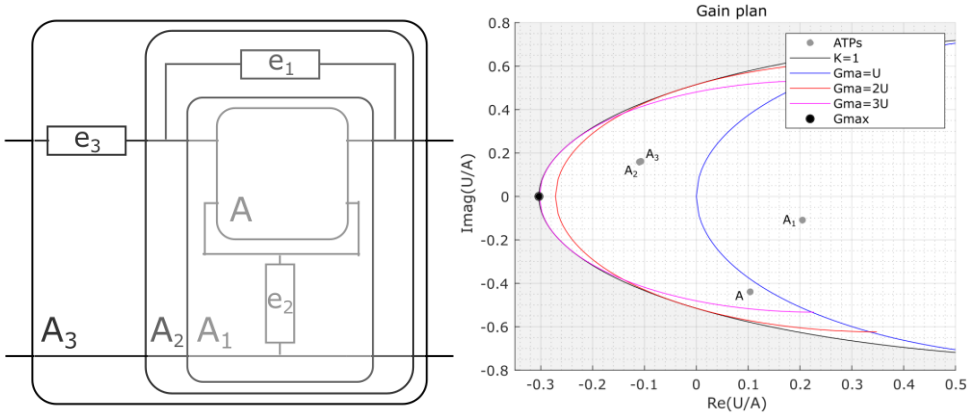


Figure 4.15 LLE on an ATP and movement on the Gain plane

One may want to analyze the effect of each kind of embedding on ATPs through graphical representation.

Series embedding

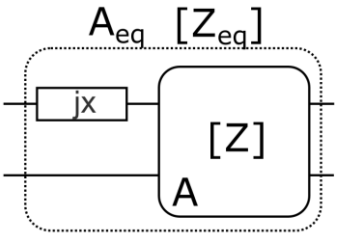


Figure 4.16 Schematic of an ATP embedded with a series element

The ATP representation can be done using any matrix of two ports (Z, Y, H, ABCD, S).

For each embedding one will use the most practical representation to express the equivalent ATP.

For the series embedding, the impedance representation *[Z]* is the most practical.

$$[Z_{eq}] = \begin{bmatrix} jx + Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \quad (4.20)$$

Hence, $A_{eq} = \frac{Z_{21}}{Z_{12}}$ remains unchanged for series embedding. The graphical representation in the Gain Plane also stays the same, so the Gain cannot be modified with series embedding alone.

Degeneration embedding

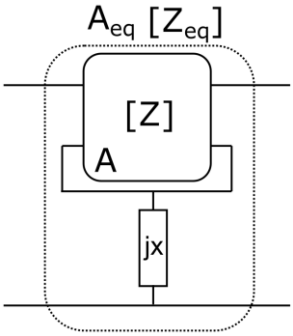


Figure 4.17 Schematic of an ATP embedded with a degeneration element

Similarly, to the series embedding, one will use the impedance representation *[Z]* also for the degeneration embedding calculation.

$$[Z_{eq}] = \begin{bmatrix} jx + Z_{11} & jx + Z_{12} \\ jx + Z_{21} & jx + Z_{22} \end{bmatrix} \quad (4.21)$$

This time A_{eq} and consequently the gain of the equivalent ATP is modified. The change can be graphically estimated and depends on the starting point for Z_{eq} and the sign of the added element *x*.

Parallel embedding

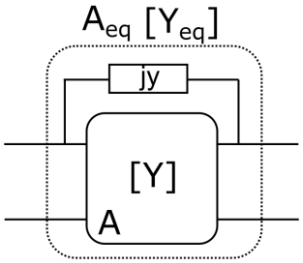


Figure 4.18 Schematic of an ATP embedded with a parallel element

Parallel embedding consists of the addition of an element between the two positive terminals of the ATP. To calculate the equivalent ATP matrix, the Admittance representation is the most convenient this time because the passive element of admittance *jy* is connected in parallel to the ATP. Hence the resulting Y_{eq} admittance matrix of the equivalent ATP is:

$$[Y_{eq}] = \begin{bmatrix} jy + Y_{11} & -jy + Y_{12} \\ -jy + Y_{21} & jy + Y_{22} \end{bmatrix} \quad (4.22)$$

4.4 Design methodology

The choice and the sizing of the embedding elements allowing to increase the G_{ma} towards G_{max} is not simple. The first method of approaching this problem is to graphically move on the gain plane by adding the embedding elements and finding a combination that offers the best performances. The paper [Bame17] presents this methodology to design a 260GHz gain-boosted amplifier. It achieves a total gain of *9.2dB* with 4 CS stages and a unilateral gain of the devices of *2.6dB* per transistor at the frequency of operation.

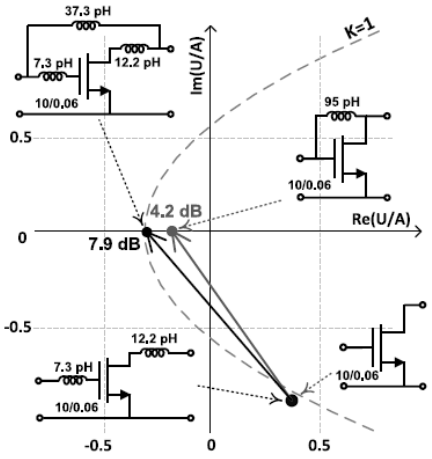


Figure 4.19 Presented graphical design methodology [Bame17]

Another approach is to use an optimization algorithm to find combinations of passive elements that offer the wanted gain boosting characteristics. As presented in [Kath18b], the models of transmission lines and capacitances can be extracted from EM simulations to find an optimum that considers these non-idealities. It achieves a total gain of *18.5dB* at 173GHz with 3 CE stages and a unilateral gain of the devices of *4dB* per transistor at the frequency of operation.

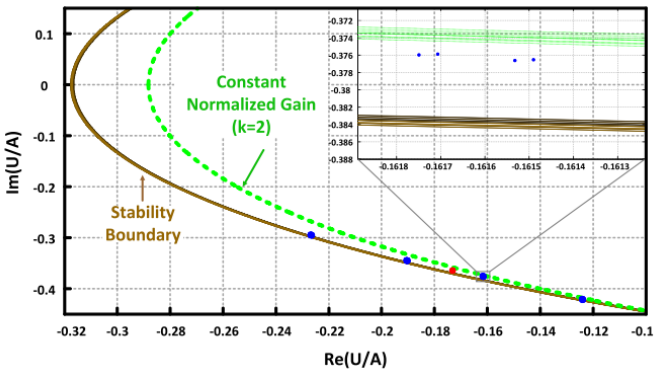


Figure 4.20 Gain boosting design method using an optimization algorithm [Kath18]

The methods are time-consuming and limit the analysis of optimal embeddings in the function of other parameters such as frequency, bias, or transistor size.

The objective of this work is to propose an analytical solution to the problem of finding LLE that increases G_{ma} to G_{max}.

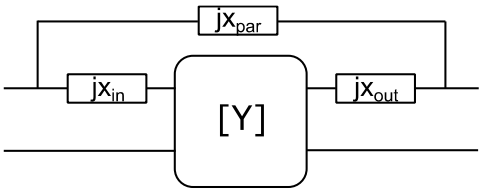


Figure 4.21 Three elements embedding around an ATP

As one has seen, two conditions must be met to achieve G_{max}, one on the amplitude of A and one on the phase. Hence, as illustrated by the CS neutralization case study, only one element is not sufficient and a second has to be used. In addition, parallel embedding is chosen for its direct effect on the amount of feedback. The second is a series embedding to control the phase. The two embedding elements offer one set of solutions; thus, a second series embedding is used to

keep a degree of freedom in the solution choice. Consequently, three elements embedding as depicted in Figure 4.21 were selected.

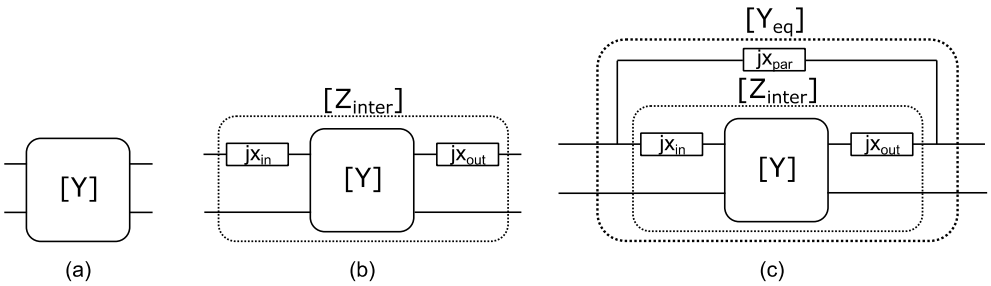


Figure 4.22 Step-by-step equivalent ATP matrix calculation from starting ATP (a), intermediate matrix with series elements (b), total equivalent matrix with three elements LLE (c)

To find the parameters of the equivalent ATP from Figure (4.21), the two-step approach presented in Figure 4.22 is applied:

-Start with an initial *Y* matrix

-Transform into *Z* matrix then use 4.20 for each series element x_{in} and x_{out}

-Transform back to the *Y* matrix and add the parallel element x_{par}.

The Intermediate Z_{inter} matrix is:

$$[Z_{inter}] = \begin{bmatrix} jx_{in} + \dfrac{Y_{22}}{\Delta Y} & -\dfrac{Y_{12}}{\Delta Y} \\ -\dfrac{Y_{21}}{\Delta Y} & jx_{out} + \dfrac{Y_{11}}{\Delta Y} \end{bmatrix} \tag{4.23}$$

It follows that the Y_{inter} matrix is:

$$\begin{aligned}[Y_{inter}] &= \frac{\Delta Y}{(Y_{22} + j\Delta Y x_{in})(Y_{11} + j\Delta Y x_{out}) - Y_{12}Y_{21}} \begin{bmatrix} Y_{11} + j\Delta Y x_{out} & Y_{12} \\ Y_{21} & Y_{22} + j\Delta Y x_{in} \end{bmatrix} \\ &= \frac{1}{F} \begin{bmatrix} Y_{11} + j\Delta Y x_{out} & Y_{12} \\ Y_{21} & Y_{22} + j\Delta Y x_{in} \end{bmatrix} \end{aligned} \tag{4.24}$$

Finally, Y_{eq} can be expressed:

$$[Y_{eq}] = \frac{1}{F} \begin{bmatrix} Y_{11} + j\Delta Y x_{out} + \dfrac{F}{jx_{par}} & Y_{12} - \dfrac{F}{jx_{par}} \\ Y_{21} - \dfrac{F}{jx_{par}} & Y_{22} + j\Delta Y x_{in} + \dfrac{F}{jx_{par}} \end{bmatrix} \tag{4.25}$$

From the equivalent admittance matrix of the two-port embedded in a three elements LLE, one can calculate the expression of A_{eq}.

$$A_{eq} = \frac{jx_{par}Y_{21} - F}{jx_{par}Y_{12} - F} = \frac{j\Delta Y x_{par}Y_{21} - (Y_{22} + j\Delta Y x_{in})(Y_{11} + j\Delta Y x_{out}) + Y_{12}Y_{21}}{j\Delta Y x_{par}Y_{12} - (Y_{22} + j\Delta Y x_{in})(Y_{11} + j\Delta Y x_{out}) + Y_{12}Y_{21}} \tag{4.26}$$

The solution to the optimal LLE problem is the solution to the following equation:

$$A_{eq} = -G_{max} \tag{4.27}$$

Which can be more detailed into:

$$\begin{cases} real(A_{eq}) = -G_{max} \\ imag(A_{eq}) = 0 \end{cases} \tag{4.28}$$

The system (4.32) is solved using (4.30) for X_{par} and X_{out}.

The development is tedious and leads to the following solution:

$$\begin{cases} \boldsymbol{X_{out}} = C\boldsymbol{X_{par}} + B & (a) \\ \boldsymbol{X_{par}} = \dfrac{\left(-b - \sqrt{\{b^2 - 4ac)}\right\}}{2a} & (b) \end{cases} \quad (4.29)$$

$$\begin{cases} C = \dfrac{Im\{Y_{12}\overline{Y_{21}}\}}{\boldsymbol{X_{in}} Re\{\Delta Y\overline{(Y_{21} - Y_{12})}\} + Im\{Y_{12} - Y_{21} + Y_{22\overline{(Y_{21}-Y_{12})}}\}} & (a) \\ B = \dfrac{\boldsymbol{X_{in}} Im\{(Y_{21} - Y_{12})\overline{Y_{11}}\} + Re\{Y_{21} - Y_{12}\}}{\boldsymbol{X_{in}} Re\{\Delta Y\overline{(Y_{21} - Y_{12})}\} + Im\{Y_{12} - Y_{21} + Y_{22\overline{(Y_{21}-Y_{12})}}\}} & (b) \end{cases} \quad (4.30)$$

$$\begin{cases} a = \alpha_{21}\alpha_{12} + \gamma_{21}\gamma_{12} + G_{max}(\alpha_{12}^2 + \gamma_{12}^2) \\ b = (\alpha_{21} + \alpha_{12})\varepsilon - (\gamma_{21} + \gamma_{12})\xi + 2G_{max}(\alpha_{12}\varepsilon - \gamma_{12}\xi) \\ c = (1 + G_{max})(\varepsilon^2 + \xi^2) \\ \alpha_{ij} = C(Im\{Y_{22}\} + \boldsymbol{X_{in}} Re\{\Delta Y\}) - Im\{Y_{ij}\} \\ \gamma_{ij} = Re\{Y_{ij}\} - C(1 + Re\{Y_{22}\} - \boldsymbol{X_{in}} Im\{\Delta Y\}) \\ \varepsilon = B(Im\{Y_{22}\} + \boldsymbol{X_{in}} Re\{\Delta Y\}) - 1 + \boldsymbol{X_{in}} Im\{Y_{11}\} \\ \xi = B(1 + Re\{Y_{22}\} - \boldsymbol{X_{in}} Im\{\Delta Y\}) + \boldsymbol{X_{in}} Re\{Y_{11}\} \end{cases} \quad (4.31)$$

The solution is long and has been factorized in multiple factors; however, the solution can be evaluated, and it allows an instantaneous calculation of an optimal embedding that gives G_{max}. Plus, with the dependence on X_{in}, there's an infinite number of solutions. After fixing X_{in}, the other parameters considered as the number of ATPs in parallel or the frequency can be swept. There is no need for a graphical representation or optimization algorithm.

4.5 Implementation of Gain-boosted Amplifier in 28nm FD-SOI

To illustrate the analytical design methodology presented in the previous part, a one-stage amplifier is designed to operate at 190GHz. Figure 4.23 presents the general LLE case and the CS transistor model performances at 190GHz with and without an LLE that boosts the gain to G_{max}. The model predicts a G_{ma} of 4.6dB and a G_{max} of *9.6dB*, hence a potential *5dB* gain improvement with the appropriate embedding.

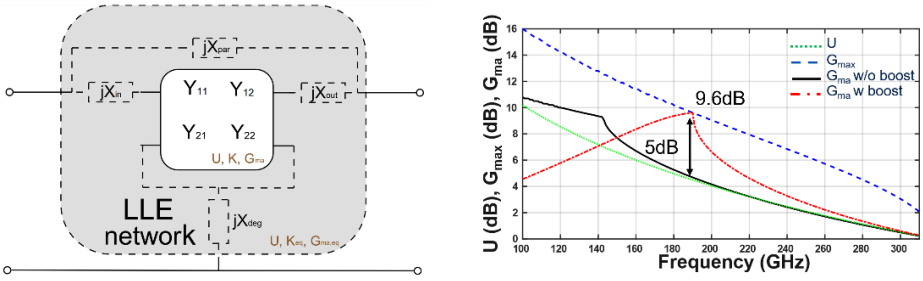


Figure 4.23 General representation of an LLE (a) simulations of transistor gain over frequency with and without gain boosting at 190GHz (b)

The first step is to design and layout the transistor up to thick metal layers while keeping a f_{max} (and thus a unilateral gain at the frequency of operation) as high as possible. The ATP model takes into consideration the full transistor model: front-end plus the proposed back-end interconnection.

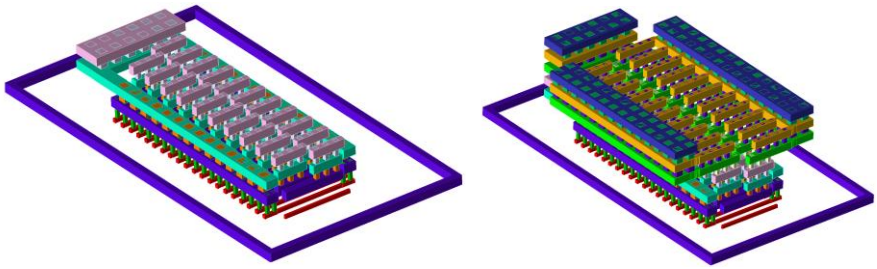


Figure 4.24 3D view of the optimized transistor staircase finger access layout: up to third thin metal layer (left) up to last sixth thin metal layer (right)

The transistor that has been used is a 28nm FD-SOI lvt-Nmos with *L=30nm* and *W=20*800nm=16µm*. The total width has been chosen to consume only *5 mA* of DC current while the 800 nm finger width allows maximizing the f_{max}. It is biased at $J_{opt,fmax}$*=0.3mA/µm* with a *1V* V_{DS}. A transistor with these dimensions has been used in previous work [Guil18] in this technology to design an oscillator at 200GHz and is optimized to achieve a high f_{max}. The layout has been kept the same and is presented in Figure 4.24. It uses two gate accesses and staircase accesses to the drain and source to minimize the drain/source coupling.

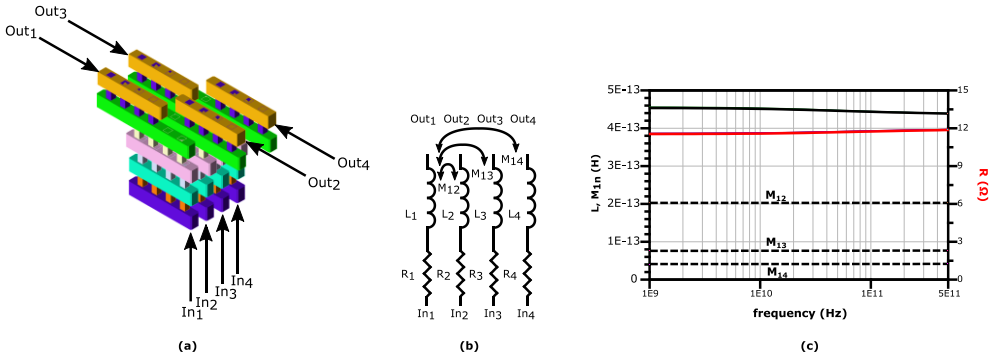


Figure 4.25 Drain/source finger accesses 3D view (a) electrical model (b) EM simulation (c)

Figure 4.25 presents the 3D view of the drain and source finger accesses through thin metal layers, the electrical model, and the result of the EM simulation of the fingers. The inductance per finger is about *0.5pH* and the mutual inductance between two adjacent fingers is *0.2pH* which is negligible in comparison to the near *12Ω* resistance per finger. The inductive part of the thin metal layers is negligible and consequently, the EM simulation of these accesses is not required to correctly model the transistor. The thin metal layers have been characterized by RCc extraction methods. The design kit's transistor model contains the front end of the line up to the first metal layer.

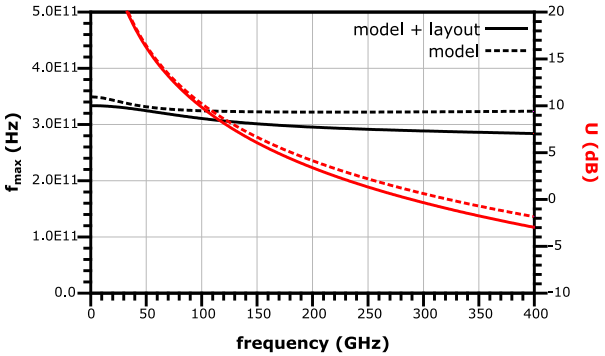


Figure 4.26 Effect of BEOL layout on transistor model, for f_{max} and U

Figure 4.26 presents the performances of the DK's transistor model without and with the layout presented in Figure 4.24. After layout, the f_{max} is reduced from 322GHz to 296GHz which represents a decrease of U at 190GHz from *4.6dB* to *3.9dB*. Because of the proximity of the operation frequency to the $f_{max,}$ the power gain is very sensitive to any additional parasitic element.

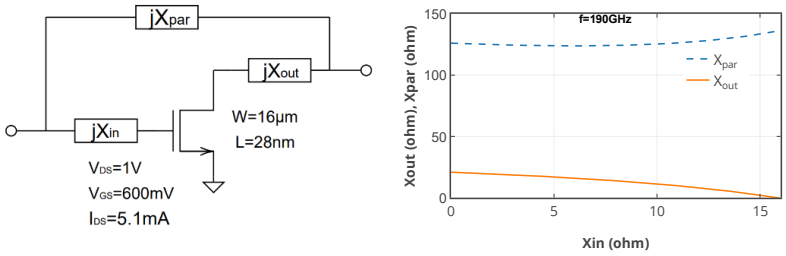


Figure 4.27 Schematic of the transistor with its embedding elements, calculated embedding elements combinations that boost the gain to G_{max} at 190GHz

Next, the *Y* parameters from the laid-out transistor are used to calculate the (4.33), (4.34), and (4.35) sets of equations to find the embedding values sets that boost the gain towards G_{max}. The schematic of the embedding and the calculated embedding values are presented in Figure 4.27. It offers multiple possible combinations. X_{in} can be set to 0 and one can only use the X_{out} element, or for practical layout reasons X_{in} and X_{out} can be chosen of equal values.

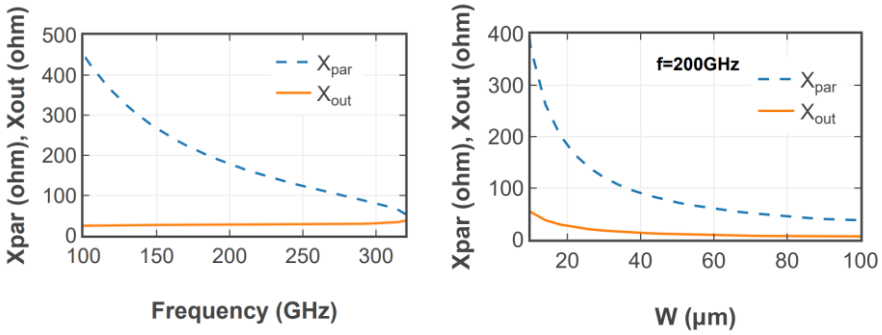


Figure 4.28 Calculated embedding solution values in function of frequency and transistor size when X_{in}=0

The methodology also enables the calculation of solutions for a wide frequency range and different transistor sizes (figure 4.28). We can notice the frequency and transistor size dependency with the values of X_{par}. Therefore, for lower frequencies and smaller transistors, this embedding may not be implementable on silicon due to the too-high required values. Alternatively, we can see that the implementation of gain boosting with larger transistors for amplifiers with higher output powers or at frequencies closer to f_{max} is easier given the decreasing passive elements values.

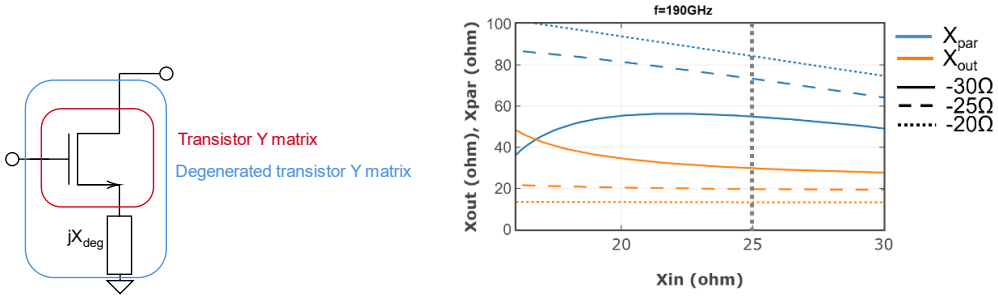


Figure 4.29 Pre-embedded CS transistor with degeneration (left) and calculation of embedding elements towards G_{max} for different degeneration impedances (right)

However, as presented in figure 4.27 the required parallel element is about 130Ω which is difficultly implementable given that embedding elements are made with transmission lines whose impedance values are limited by their characteristic impedance. The values that these obtainable impedances can take depend on the technology metal stack, metal thicknesses, and distance between the top and bottom metal layer. The transmission line impedances can take a maximal value of around 90-100Ω. The goal of this work is to propose a methodology to find an embedding to any ATP, the transistor has been kept the same knowing that from Figure 4.28, an increase of the transistor size or an increase in the frequency could reduce the required X_{par} value. Another solution is to pre-embed the transistor with a degeneration element and then find a solution to the modified ATP. Figure 4.29 presents the pre-embedded transistor with degeneration and the calculated embedding solutions for different degeneration element values. In the case of a transistor, a negative impedance is necessary to reduce the X_{par} value.
Hence, we decided to add an X_{deg}=-30Ω in combination with, X_{par}=55Ω, X_{in}=25Ω, and X_{out}=28Ω.
The passive elements used to implement the calculated embedding are presented next.

Transmission Line:

The *Y* matrix representation of a lossless TL of characteristic impedance Z_0 and length *l* is:

$$Y_{TL} = \begin{bmatrix} \frac{1}{jZ_0 tan(\beta l)} & -\frac{1}{jZ_0 sin(\beta l)} \\ -\frac{1}{jZ_0 sin(\beta l)} & \frac{1}{jZ_0 tan(\beta l)} \end{bmatrix} \quad (4.32)$$

Where β is the propagation constant.
For short transmission lines (i.e., when $l<\lambda/20$), the matrix can be approximated to

$$Y_{TL} \approx \begin{bmatrix} \frac{1}{jZ_0 \beta l} & -\frac{1}{jZ_0 \beta l} \\ -\frac{1}{jZ_0 \beta l} & \frac{1}{jZ_0 \beta l} \end{bmatrix} \quad (4.33)$$

Which is equivalent to a lumped element of impedance $jZ_0\beta l$ and acts as an inductance. Hence short transmission line can be used for series X_{in} and X_{out} LLE elements. For the parallel element, longer TL can be used and acts as a lumped element of impedance $jZ_0 \sin(\beta l)$. As stated before, a transmission line can present a maximal impedance of jZ_0 when $\beta l = \pi/4$.

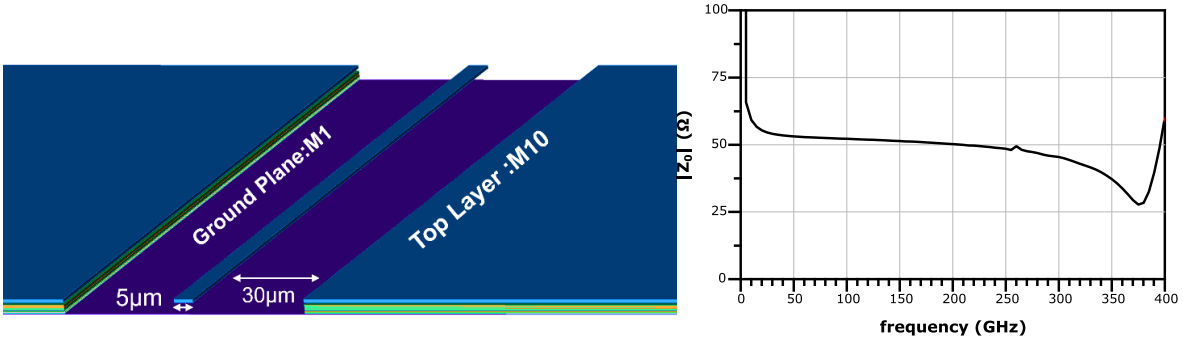


Figure 4.30 3D view of the 50Ω Microstrip transmission line used in the amplifier (left) and simulated characteristic impedance over frequency (right)

For the design of the amplifier, 50Ω Microstrip transmission lines as depicted in Figure 4.30 have been used. It allows for keeping a continuous ground plane around the TL contrary to a coplanar waveguide TL. In addition, side ground plane walls using all metal layers have been implemented at 30µm from the line to fulfill the density rules requirements without requiring metal tilling.

The simulated losses of the line are *1.7dB* per millimeter.

Capacitance:

The value of the required capacitance is small, and a precise model is necessary. The circuit behavior is sensible to this value. In addition, the interconnection to the capacitance must be as simple as possible to reduce additional parasitics. Consequently, the best capacitance for this circuit is the interdigitated one, which is realized with the top IB metal layer, the same as the one used for transmission lines.

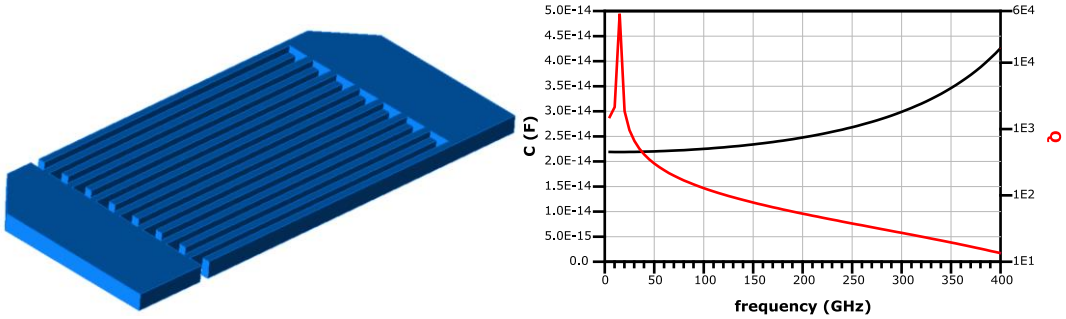


Figure 4.31 3D view of the interdigitated capacitance used for the degeneration (left) simulated capacitance value and quality factor (right)

Figure 4.31 presents the 3D view of the degeneration capacitance whose standalone value is *25fF* and presents the required impedance in situ, once interconnected to the circuit. The quality factor of the capacitance is equal to 65 at 190GHz.

RF pad:

The RF pad that has been used in this design is depicted in Figure 4.32 and is composed of a top copper and aluminum metal stack over a first metal ground that connects the two ground pads on the sides. The capacitance of the pad is about *21fF* at 190GHz and may be integrated into the impedance-matching design.

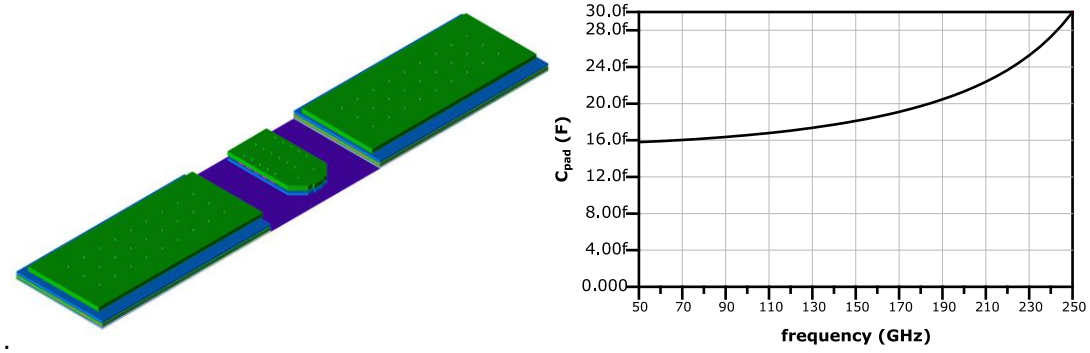


Figure 4.32 3D view of the used RF pad in the amplifier (left) and simulated pad capacitance over frequency (right)

It is a commonly used pad in this technology and is compatible with 50 and 100µm pitches. Further optimization to reduce the capacitance was not necessary for this design.

Core Boosted Amplifier:

The core amplifier elements are implemented as presented in Figure 4.33. An additional interdigitated capacitor is used in series with the parallel transmission line to separate the DC bias of the drain and source. The impedance of the capacitance is considered to tune the length of the parallel TL.

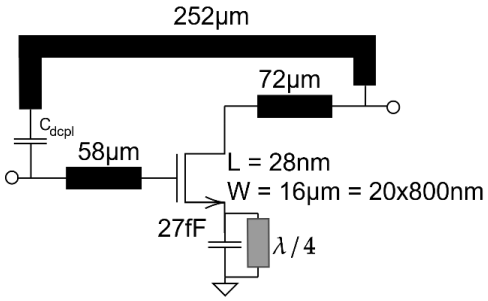


Figure 4.33 Schematic of the core gain boosted amplifier at 190GHz.

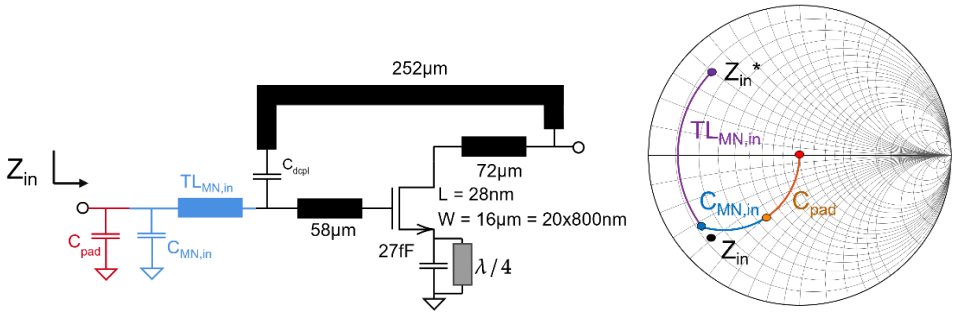


Figure 4.34 Addition of the input impedance matching network schematic (left) and impedance transformation represented on a Smith chart (right)

The impedance matching is kept as simple as possible to limit the losses, as depicted in Figure 4.34, a direct transmission line $TL_{MN,in}$ connecting the input of the core amplifier to the input of the circuit is used. The series same 50Ω transmission line is used to bring the real input admittance part to *1/50 S* and the imaginary part is then canceled thanks to the pad capacitance and an additional interdigitated capacitance $C_{MN,in}$ added next to the pad.

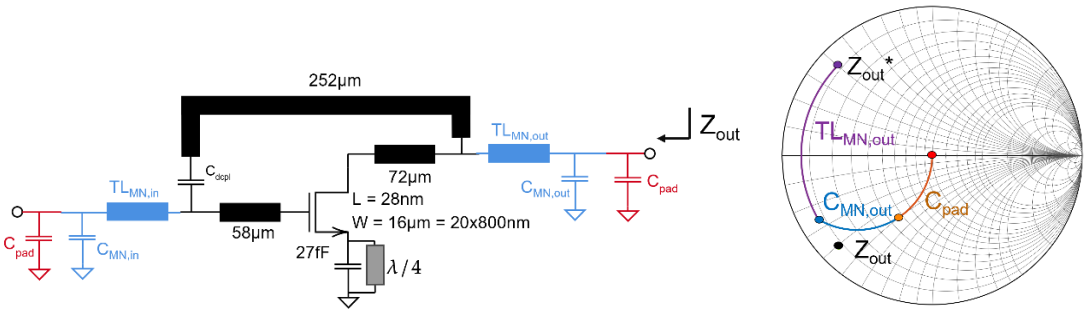


Figure 4.35 Addition of the output impedance matching network schematic (left) and impedance transformation represented on a Smith chart (right)

Similarly, the output-matching network employs the same strategy as depicted in Figure 4.35.

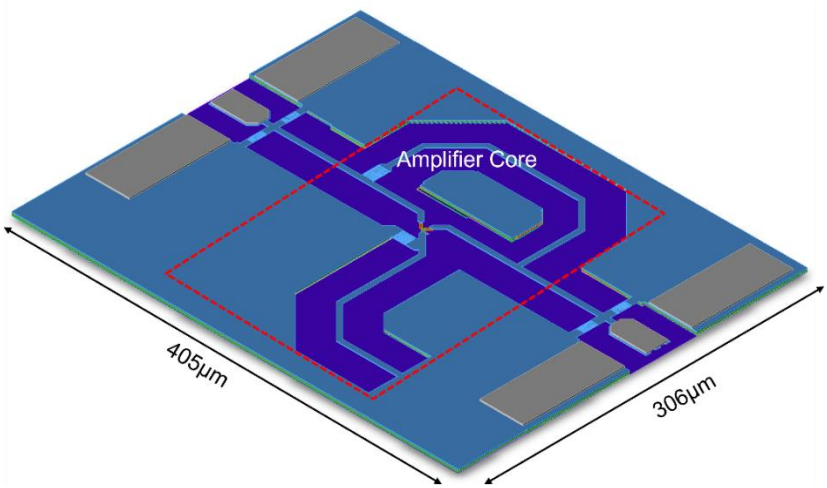


Figure 4.36 Gain Boosted Amplifier 3D view

The top layout of the amplifier is presented in Figure 4.36, it has been EM simulated with a 2.5D EM simulator except for the transistor that has been extracted separately as described before. The implemented layout is fully compliant with all metal density rules, and very stringent in a 28 nm node. The transmission line design and the passive devices position were chosen to not require any additional tilling for density rules purposes which could lead to performance shift or deterioration.

4.6 Measurement and analysis

The fabricated circuit is presented in Figure 4.35 (left), while the measurement test fixture is given in 4.37 (right). A Rohde&Shwartz ZVA 24 PNA with ZC220 frequency converters and Picoprobe 220 probes are used for the measurements.

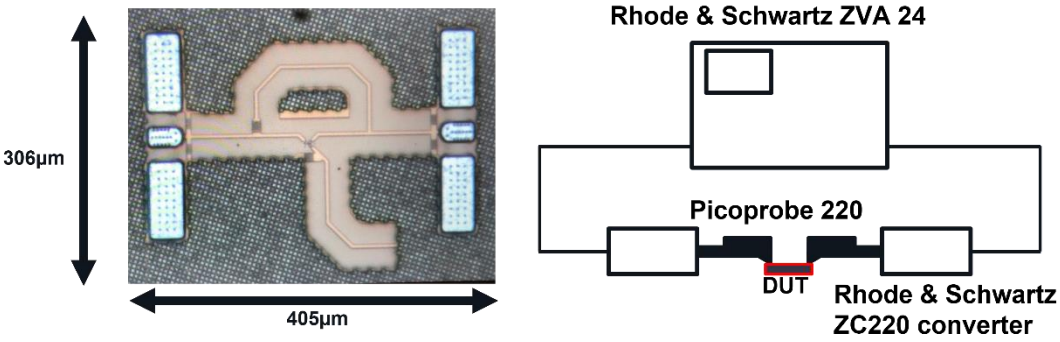


Figure 4.37 Chip microphotograph (left) measurement setup schematic (right)

In Figure 4.38, we present a wide frequency band S-parameters comparison between simulations and measurements. An excellent agreement between measurements and simulations can be observed.

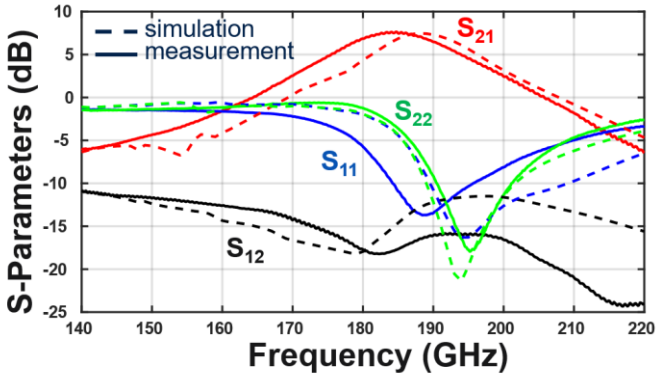


Figure 4.38 S-parameters measurement and simulations comparison [Sadlo21]

Nevertheless, there is a small shift in the input reflection coefficient that leads to a center frequency measured at 184GHz while simulations indicate 189GHz. The gain is equal to *7.6dB* at 184GHz and is in line with the simulations. The amplifier consumes *5.1mW* from a supply of *1V*. The amplifier has a 3dB bandwidth of 20GHz. The S-parameters are summarized in Figure 4.39.

	Sim.	Meas.
Frequency (GHz)	190	184
S_{21} (dB)	7.4	7.6
S_{11} (dB)	-10.8	-9.7
S_{22} (dB)	-12.2	-3
S_{12} (dB)	-12.2	-18

Figure 4.39 Simulation to measurement S-Parameters comparison

Figure 4.40 shows measured results from 5 chips on the same wafer putting in light very low dispersion (σ = *0:35 dB* for S_{21}) and demonstrating the process's low variability even at these very high frequencies.

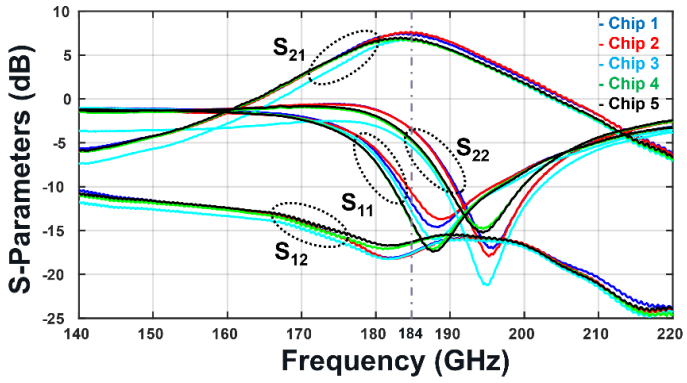


Figure 4.40 Amplifier measurements on multiple dies

The dispersion is also very low for the other S-parameters. Given the frequency, even slight variability in capacitances or resistivity could have led to significant performance variations. The calculated variability for all S-parameters is presented in Figure 4.41.

@184GHz	S21 (dB)	S12 (dB)	S11 (dB)	S22 (dB)
Average	7.12	-17.34	-11.50	-3.8
σ	0.35	0.64	1.18	0.76

Figure 4.41 Dispersion on multiple dies measurement

Large signal measurements have also been performed on the amplifier using the same measurement fixture as for the S-Parameters. The input power has been swept on the PNA to determine this circuit's linearity, which can generate up to around *2dBm* at 190GHz which is enough for this circuit.

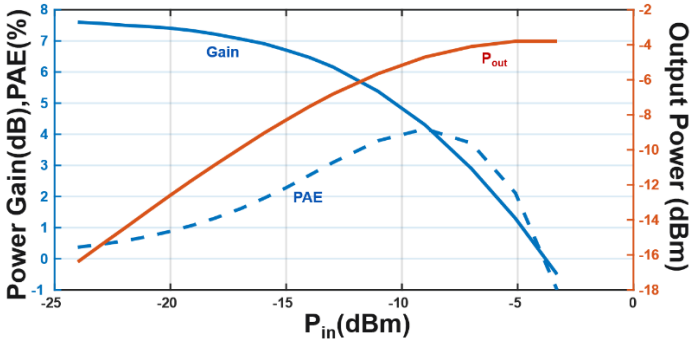


Figure 4.42 Large signal measurements at 184GHz: gain, output power, and PAE

The large signal measurements are presented in Figure 4.42 and Figure 4.41 summarizes the main large signal performances. The amplifier shows an *OCP1* of *-8 dBm*, a peak *PAE* of *4.2 %,* and a P_{sat} of *-3.7 dBm*.

OIP1 (dBm)	-8
Power Gain (dB)	7.6
P_{sat} (dBm)	-3.7
Max PAE	4.2%
Consumption (mW)	5.1

Figure 4.43 Large signal performances at 184GHz

Figure 4.44 presents the performance comparison with the state of the art. The existing amplifiers can be separated into two categories, with and without gain boosting. The difference between these two categories arises in the gain per stage metric. The comparison demonstrates the superiority of gain boosting for improving the overall gain. The presented amplifier offers the second lower power consumption per stage and the best gain per stage to power consumption tradeoff.

	Process	f_{max} (GHz)	Operation Freq. (GHz)	Gain Boosting	Structure	Gain (dB)	Gain/Stage (dB)	P_{sat} (dBm)	BW (GHz)	peak PAE (%)	P_{DC}/Stage (mW)	FOM	Area (mm^2)
This work	**28nm FD-SOI**	**390**	**184**	**Yes**	**1 CS**	**7.6**	**7.6**	**-3.7**	**20**	**4.2**	**5.1**	**1.30**	**0.12**
[Kath18a]	130nm SiGe	280	173	Yes	3 CE	18.5	6.2	0.9	8.2	-	**14**	**1.65**	0.40
[Kath18b]	130nm SiGe	280	183	Yes	1 Casc.	9.5	9.5	-2.8	8.5	-	**30**	**1.27**	0.16
[Park19]	65nm CMOS	395	242	Yes	4 CS	13.9	3.48	-3.3	29.7	1.6	**6**	**0.85**	0.142
[Bame17]	65nm CMOS	352	257	Yes	4 CS	9.2	2.3	-3.9	12.2	0.8	**6.9**	**0.91**	0.14
[Park20]	65nm CMOS	395	247/272	Yes	2 CS	18/15	9/7.5	0.09/-2.36	5	4.4/2.4	**10.8**	**3.11/2.64**	0.28
[Park18]	65nm CMOS	395	280/300	Yes	3 CS	12/9	4/3	-4.7/-5.3	-	1.6/1.4	**6**	**1.26/1.15**	-
[Ko13]	40nm CMOS	275	213	No	9 CS	10.5	1.17	-3.2	13	0.75	**4.7**	**0.79**	0.12
[Frit14]	130nm SiGe	450	200	No	2 Casc.	17	8.5	-3.5	44	-	**9**	**1.40**	0.24
[Parv18]	28nm FD-SOI	390	160	No	4 CS	15.7	3.93	1.3	23	-	**8**	**0.42**	0.34

$*FOM = \frac{\sqrt[n]{Gain}}{U(f)}$: where n is the number of stages and $U(f)$ is the unilateral power gain of the device at the frequency of operation.

Figure 4.44 Comparison table with the state-of-the-art

The designed amplifier also presents a large bandwidth and a promising peak *PAE* compared to the literature. Finally, compared to other 28 nm FD-SOI Amplifiers, the Gain Boosting allows *3.6dB* higher gain and *35%* less consumption per stage.

The objective of gain boosting is to produce the highest possible amount of gain from a given device without compromising its stability. Consequently, the common figure of merit (FOM) is the

ratio between the gain per stage and the unilateral gain at the frequency of operation. Thus, the FOM can be expressed as:

$$FOM = \frac{\sqrt[n]{Gain}}{U(f)} \tag{4.34}$$

The measured performance of the amplifier matches the fast corner of the technology that indicates a f_{max} of 390GHz versus 330GHz in the typical case. Previous works [Guil19] indicated an underestimation of the transistor performances at near f_{max} frequencies and thus a higher f_{max}. The 390GHz value has been chosen as the reference to calculate the FOM of this amplifier to demonstrate a realistic gain improvement with this presented methodology. Consequently, this amplifier performs an FOM of 1.3 which is competitive with the other reported works.

Measurements for body bias control:

Measurements have also been made to investigate the potentialities of the body biasing feature at such high frequencies. Figure 4.45 presents the measured amplifier current consumption as a function of the transistor gate voltage for different values of backgate voltage. This measurement was used to find a set of voltages that give the same current consumption. For lower frequencies, it has been demonstrated that the transistor's performances are dependent on the (V_{gs}-V_t) quantity and are de-correlated from the absolute gate or body. bias.

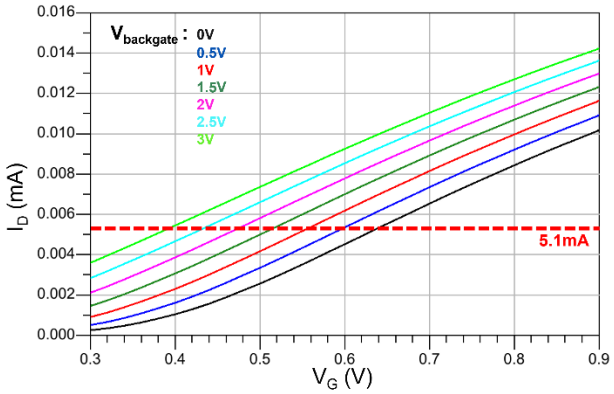


Figure 4.45 Measured current consumption versus gate voltage for different backgate biases

Figure 4.46 presents the different S-parameter measurements at a constant DC current of *5.1mA* for different backgate biases. The gain performances present low variations for backgate voltages up to *2V* and a slight decrease in the gain for *2.5V* and *3V*. The reverse isolation and reflection coefficients remain constant for all bias points.

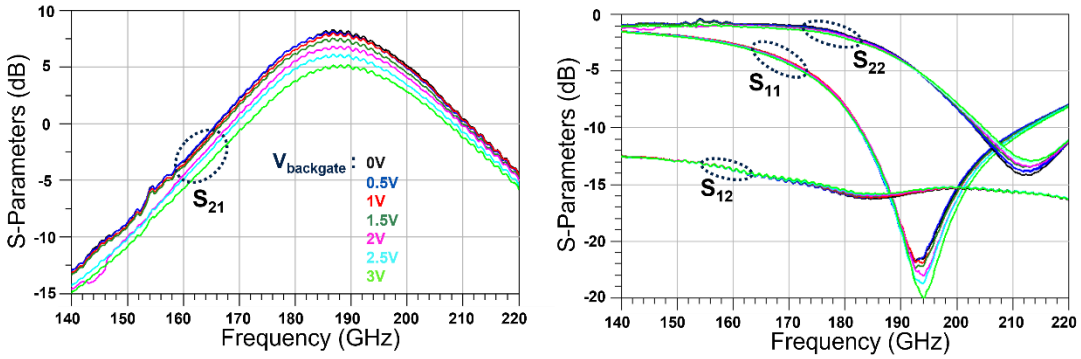


Figure 4.46 S-parameters measurement of the Gain boosted amplifier at constant DC current for different backgate voltages.

These measurements confirm that even at this frequency the gate bias can be replaced by the body bias that is not on the path of the RF signal and opens the potentialities for new amplification architecture in FD-SOI. This concept will be discussed in the last chapter on perspectives and illustrated with simulations.

4.7 Conclusion on results and goals for higher bandwidth and power

This chapter introduced the formalism of Gain boosting that enables the calculation of optimal conditions for an active two-port to achieve G_{max} while maintaining stability conditions. The study of CS neutralization highlighted the properties and limitations of this method. The generalization of the Gain boosting to any ATP with LLEs was introduced and an analytical methodology to design an optimal LLE was presented. With this method, a mmW design is significantly simpler compared to previously known methodologies and enables an analytic analysis over multiple frequencies or component parameters sweeps. To demonstrate this methodology, an 184GHz gain-boosted amplifier was designed and implemented in 28nm FD-SOI CMOS technology. The one-stage amplifier presents state-of-the-art power gain for the lowest power consumption while having a large bandwidth and a promising peak *PAE* for the given operation frequency. When considering both the gain per stage FOM and power consumption, the proposed solution performs the best-in-class compromise. This one-stage gain amplifier demonstrates the potential of the cited VLSI technology for >160 GHz transceivers for very high data rate communications.
This chapter studied the gain performances of the amplifier. The other important characteristics are the bandwidth and the maximal output power of the amplifier. The next chapter studies the design challenges regarding this characteristic and compares multiple solutions. In addition, thanks to the analytical design methodology of gain boosting LLE, the method is extended to design LLEs to boost the gain over very large frequency bands.

4.8 References

[Ster55] A. P. Stern, C. A. Aldridge, and W. F. Chow, "Internal Feedback and Neutralization of Transistor Amplifiers," in Proceedings of the IRE, July 1955

[Chen55] C. Cheng, "Neutralization and Unilateralization," in IRE Transactions on Circuit Theory, June 1955

[Sing64] A. Singhakowinta and A. Boothroyd, "On Linear Two-Port Amplifiers," in IEEE Transactions on Circuit Theory, March 1964

[Simi21] D. Simic and P. Reynaert, "Analysis and Design of Lossy Capacitive Over-Neutralization Technique for Amplifiers Operating Near f_{MAX}," in IEEE Transactions on Circuits and Systems I: Regular Papers, May 2021,

[Chan10] W. L. Chan and J. R. Long, "A 58–65 GHz Neutralized CMOS Power Amplifier with PAE Above 10% at 1-V Supply," in IEEE Journal of Solid-State Circuits, March 2010.

[Mome11] O. Momeni and E. Afshari, "A high gain 107 GHz amplifier in 130 nm CMOS," 2011 IEEE Custom Integrated Circuits Conference (CICC), San Jose, CA, USA, 2011

[Sing64b] Anek Singhakowinta, Ph.D. Thesis, “Gain, sensitivity and stability of linear two port amplifiers", University of London 1964

[Amak16] S. Amakawa and Y. I. Graduate, "Graphical approach to analysis and design of gain-boosted near-fmax feedback amplifiers," 2016 46th European Microwave Conference (EuMC), London, UK, 2016

[Bame17] H. Bameri and O. Momeni, "A High-Gain mm-Wave Amplifier Design: An Analytical Approach to Power Gain Boosting," in IEEE Journal of Solid-State Circuits, Feb. 2017

[Guil17] R. Guillaume et al., "Energy efficient distributed-oscillators at 134 and 202GHz with phase-noise optimization through body-bias control in 28nm CMOS FDSOI technology," in 2017 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), Honolulu, HI, 2017.

[Parv18] D. Parveg et al., "Design of a D-Band CMOS Amplifier Utilizing Coupled Slow-Wave Coplanar Waveguides," in IEEE Transactions on Microwave Theory and Techniques, March 2018.

[Kath18] H. Khatibi et al., "A 173 GHz Amplifier With an 18.5 dB Power Gain in a 130 nm SiGe Process: A Systematic Design of High-Gain Amplifiers Above fmax=2," in IEEE Transactions on Microwave Theory and Techniques, Jan. 2018.

[Park19] D. Park et al., "A 230–260-GHz Wideband and High-Gain Amplifier in 65-nm CMOS Based on Dual-Peak Gmax-Core" in IEEE Journal of Solid-State Circuits, June 2019

[Kath18] H. Khatibi et al., "A 183 GHz Desensitized Unbalanced Cascode Amplifier With 9.5-dB Power Gain and 10-GHz Band Width and -2 dBm Saturation Power," in IEEE Solid-State Circuits Letters, March 2018.

[Park20] D. -W. Park et al., "A 247 and 272 GHz Two-Stage Regenerative Amplifiers in 65 nm CMOS with 18 and 15 dB Gain Based on Double-Gmax Gain Boosting Technique," in 2020 IEEE Symposium on VLSI Circuits, Honolulu, HI, USA, 2020.

[Park18] D. Park et al., "A 280-/300-GHz Three-Stage Amplifiers in 65-nm CMOS With 12-/9-dB Gain and 1.6/1.4% PAE While Dissipating 17.9 mW," in IEEE Microwave and Wireless Components Letters, Jan. 2018.

[Ko13] C. Ko et al., "A 210-GHz Amplifier in 40-nm Digital CMOS Technology," in IEEE Transactions on Microwave Theory and Techniques, June 2013.

[Frit14] D. Fritsche et al., "A Broadband 200 GHz Amplifier with 17 dB Gain and 18 mW DC-Power Consumption in 0.13 m SiGe BiCMOS," in IEEE Microwave and Wireless Components Letters, Nov. 2014.

[Guil19] Raphael Guillaume, Ph.D. Thesis, «Millimeter-wave and terahertz frequency synthesis on advanced silicon technology», Université de Bordeaux, October 2019.

[Sadl21] S. Sadlo, et al., "One stage gain boosted power driver at 184 GHz in 28 nm FD-SOI CMOS," 2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC).

Chapter 5 Large bandwidth, High power PA

5.1 Specifications for the PA

The objective of this design is to challenge the limitations of the actual state of the art regarding the PA around 200GHz in CMOS technologies. The published works are summarized in Figure 5.1. After the gain enhancement that has been studied in the previous chapter, the P_{sat} and the bandwidth are two other critical metrics that must be addressed. The highest presented P_{sat} is of 9.4dBm in [Bame20], and the maximal bandwidth is above 12% for [Park19] and up to 36% for [Parv16]. However, no publication presents a high P_{sat} and bandwidth as [Bame20] bandwidth is 7%.

Reference	Technology	Node (nm)	F_0 (GHz)	BW (GHz)	BW (%)	Psat (dBm)	Gain (dB)	Gain/stage (dB)	No. of Stages	PAE max (%)	DC
[Park19]	CMOS	65	236	29.7	12.5	-3.3	13.9	3.48	4	1.6	23.8
[Ko13]	CMOS	40	213.5	13	6.1	-3.2	10.5	1.17	9	0.75	42.3
[Wang14]	CMOS SOI	32	210	14	6.7	4.6	15	5	3	6	40
[Bame20]	CMOS	65	202	14	6.9	9.4	19.5	4.44	8	1.03	732
[Simi21]	CMOS	28	190	14.4	7.6	1.5	14.3	2.86	5	2.6	45
[Sadl21]	FD-SOI	28	184	20	10.9	-3.7	7.6	7.6	1	4.2	5.1
[Parv16]	FD-SOI	28	170	61	35.9	n/a	10.1	2.53	4	n/a	31
[Seo09]	CMOS	65	150	27	18	6	8.2	2.73	3	9.5	25.5
[Li21]	CMOS SOI	45	140	19	13.6	17.5	24	6	4	13.4	410
[Simi18]	CMOS	40	140	17	12.1	14.8	20.3	5.08	3	8.9	305

Figure 5.1 State of the Art of PA around 200GHz in CMOS technologies

The targeted specifications for this work and the challenges that they represent are:

- P_{sat} > 12dBm: the 12dBm is defined by the last stage of the PA, it is targeted for a differential scheme, therefore each branch of the PA output stage must be sized for a 9dBm P_{sat}. Then the differential structure can be replicated in parallel branches to combine their output power to generate over 12dBm. The other advantages regarding differential schemes for high P_{sat} will be presented.
- Gain > 15dB: Given the previous study and the implemented gain-boosted amplifier, a 3-stage PA is required to produce this amount of gain at 200GHz. Gain boosting with TL feedback is required to achieve a high enough gain.
- Bandwidth > 20%: a 180-220GHz bandwidth is targeted which is challenging for the high-targeted P_{sat}. Impedance matching requires resonating high capacitive components of the transistor input and output impedances over a broad frequency range. Multi resonant matching networks are required and distributed power splitters and combiners are best suited for large bandwidth and low-loss operations. Finally, the gain-boosting methodology must be applied over the 180-220GHz band, and the strategy to achieve this result will be presented.

The resulting architecture is depicted in Figure 5.2.

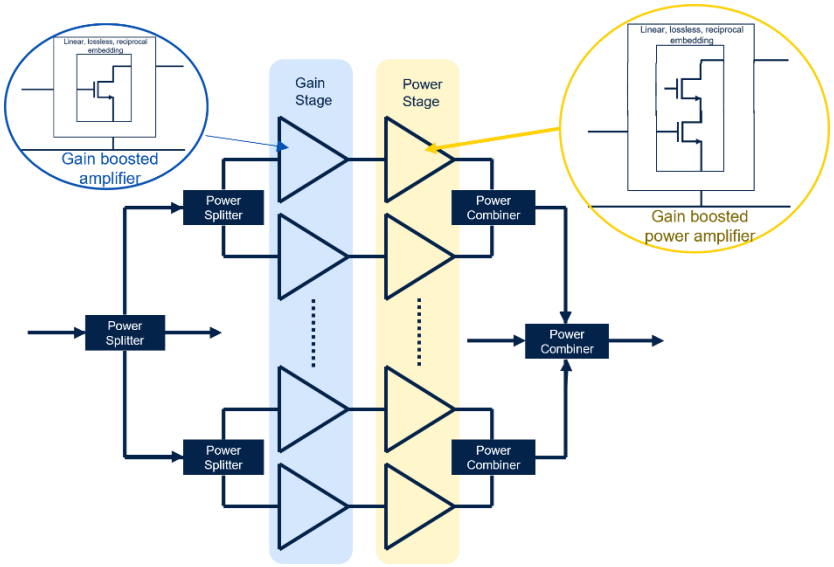


Figure 5.2 Targeted 200GHz PA architecture

5.2 Large signal impedance matching and device sizing

The maximal available power gain of a transistor, a circuit, or generally any component is achieved by presenting the conjugate impedance to its ports. An amplifier therefore is optimized regarding the power gain by conjugate-matching its output port. However, the condition that allows the amplifier to produce the highest possible amount of power is not necessarily the termination by the output conjugate impedance. Figure 5.3 illustrates an output port of an amplifier represented by a current source I_{out} and an output resistance R_{out}. It is terminated by a load R_L.

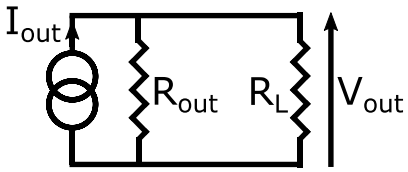


Figure 5.3

The circuit has a known maximal voltage and current that it can support and provide without damage, V_{max} and I_{max} respectively. The maximal power that can be therefore generated is:

$$P_{max} = \frac{V_{max} I_{max}}{2} \tag{5.1}$$

The optimal load that must be seen by the current source to achieve the P_{max} is consequently:

$$R_{opt} = \frac{V_{max}}{I_{max}} \tag{5.2}$$

In the configuration of Figure 5.3, the optimal load impedance is:

$$R_{L,opt} = \frac{1}{\frac{1}{R_{opt}} - \frac{1}{R_S}} \tag{5.3}$$

The current that flows out of the source is shared by the load and the internal resistance, a part of the delivered power is not received by the load and is lost. The maximal efficiency of the source is therefore limited by its impedance, and it can be expressed as:

$$\eta_{max} = \frac{P_{L,opt}}{P_{L,opt} + P_{S,opt}} = \frac{1}{1 + \frac{R_{L,opt}}{R_S}} \tag{5.4}$$

The higher the source resistance, the higher the efficiency. An ideal power amplifier may have an infinite output resistance to maximize its intrinsic efficiency.
When the amplifier is conjugate matched, the presented resistance is equal to the source impedance.

$$R_{L,\mathrm{match}} = R_S \tag{5.5}$$

The power efficiency is then equal to 50% from (5.4),

$$\eta_{max,match} = 50\% \tag{5.6}$$

The maximal power that can be generated by the current source is lower and can be limited either by the maximal voltage V_{max} or the maximal current I_{max} depending on the value of R_S:
When $R_S \geq 2R_{L,opt}$, the power is limited by V_{max}, therefore:

$$\mathrm{P}_{\mathrm{max,match}} = \frac{\mathrm{V}_{\mathrm{max}}^2}{R_S} \tag{5.7}$$

When $R_S \leq 2R_{L,opt}$, the power is limited by I_{max}, therefore:

$$\mathrm{P}_{\mathrm{max,match}} = \frac{\mathrm{I}_{\mathrm{max}}^2 R_S}{4} \tag{5.8}$$

The difference between the optimal power load and the conjugate matching regarding the maximal power and efficiency has been described. However, conjugate matching is necessary to obtain the highest possible gain of the amplifier, the G_{ma}. Because the optimal load is different, a lower gain is therefore achieved for an optimal power load. It can be expressed and calculated using the reflection coefficient. The maximal available gain for the optimal power load can be referred to as $G_{ma,opt}$:

$$G_{ma,opt} = G_{ma}\left(1 - \left|\Gamma_{Rs,RL,opt}\right|^2\right) = G_{ma}\left(1 - \left|\frac{R_S - R_{L,opt}}{R_S + R_{L,opt}}\right|^2\right) \tag{5.9}$$

The G_{ma} can be achieved with an optimal power load when it is equal to the output impedance of the source. It is the opposite of the efficiency from equation (5.4) that is maximized when the optimal power load is much smaller than the output impedance. Equation (5.9) can be rewritten to make visible the maximal efficiency and highlight the required tradeoff:

$$G_{ma,opt} = G_{ma}(1 - |2\eta_{max} - 1|^2) \tag{5.10}$$

Consequently, (5.10) shows that the higher the amplifier efficiency the higher the gain reduction. However, R_S is a consequence of the intrinsic device characteristics that are used in the amplifier. The I_{max} and V_{max} are imposed by the available active components as well.
This section aims to highlight the device characteristics that lead to an efficient high P_{sat} and to discuss the available gain reduction when the devices are optimized for power output and not power gain.
The design of the power amplifier faces this additional challenge at high frequencies. The gain already drops intrinsically in the devices and the design for the highest P_{sat} requires sacrificing a part of the low available gain. This explains in part the low achieved efficiency performances in these frequency bands.
To increase the power output of an amplifier, V_{max} and or I_{max} must be increased. A transistor has a maximal operation voltage, typically 1V DC for a common source in 28nm FD-SOI and 2V dynamically. Therefore, to increase this limit, the use of stacked transistors is required. To increase the maximal current, the number of unitary transistors (transistor fingers) in parallel must be increased. The increase in the number of transistor fingers is challenging as the interconnection

parasitic reduces the f_{max}. The study to find an optimal geometry with several fingers is discussed in the next section.

The maximal power efficiency of a transistor or a more complex amplifier depends on other factors than the efficiency discussed here. A part of the additional efficiency loss comes from the ability of the amplifier to function as an ideal current source as represented in this study.

The transistors are not linear components when they are operating through their full voltage and current ranges. Therefore, the maximal voltage and current of a device gives the order of magnitude of the optimal load impedance but it is necessary to consider the real behavior of the device for iterative research of the optimal impedance.

The load-pull is used for this purpose on amplifiers; it consists of a load impedance sweep with an input signal adapted to push the amplifier into its current and voltage limits. It is illustrated in Figure 5.4. In practice, a load-pull is done on a unitary component in the simulation if the model captures enough of the nonlinearity of the devices or in measurement. The load-pull can be done to achieve the highest efficiency from an amplifier or the highest output power. Then the unitary component is replicated to increase the current to the desired current and the load is scaled.

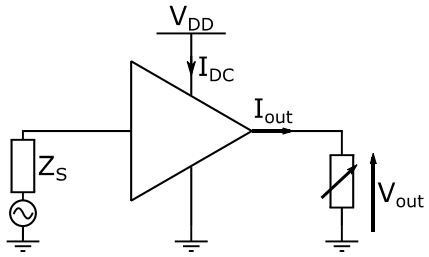


Figure 5.4 Load-pull principal schematic

The imaginary part of the output impedance of an amplifier is canceled similarly for a conjugate match as for a power match. The difference comes from the real part of the load impedance, this is why the previous discussion only considered real impedances.

5.3 Large transistor design optimization for sub-THz operation

To increase the P_{sat} of an amplifier, multiple transistors can be stacked or the number of transistors in parallel can be increased. During this scaling process, the additional parasitic elements must be as small as possible to keep as much as possible of the intrinsic device performance. This part presents the optimal geometry of a unitary transistor for f_{max} in the 28nm FD-SOI process. Then the geometry of the multi-transistor matrix and the simulation methodology is presented.

The best f_{max} performances of a 28nm FD-SOI transistor have been achieved through a meticulous optimization that is described in Appendix 3, is 313GHz and the transistor dimensions are L = 30nm, W=20x600nm=15.6µm. The transistor layout uses a two-side gate access, and it is depicted in Figure 5.5.

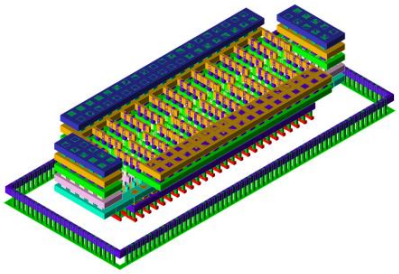


Figure 5.5 3D view of a 26x600nm transistor cell laid out up to 6th thin metal layer.

A 6-transistor matrix using the previous unitary cell has been designed and its optimization is presented in Appendix E as well. The matrix is depicted in Figure 5.6. The simulation comprises the RCc extraction of the 6 first thin metal layers and an EM simulation of the superior thick metal layer interconnections.

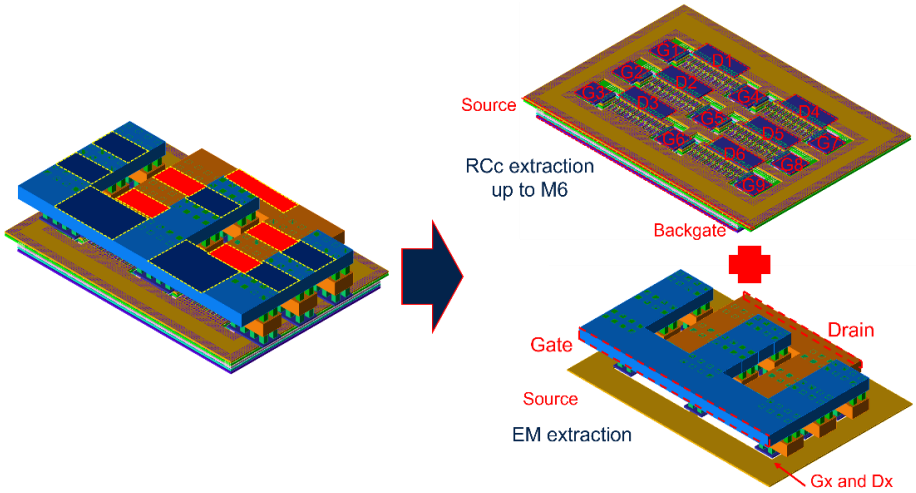


Figure 5.6 Strategy of a complete simulation of the transistor matrix with the RCc extracted lower part and EM simulated higher part

The f_{max} of this structure is 291GHz, the results are summarized in Figure 5.7.

Transistor size (W)	15.6µm=26x600nm	93.6µm=6x15.6µm	
Extraction type	RCc	RCc	EM+RCc
F_{max} @200GHz	313 GHz	313 GHz	291GHz

Figure 5.7 Summary of transistors' f_{max} for L=30nm

5.4 Impedance Matching and power combining techniques for large BW

The impedance matching between amplification stages and the source and load must be done in an efficient and broadband manner. An amplifier impedance can be represented as a resistance in parallel with a capacitance as presented in Figure 5.28 (a). In the matching process, the capacitance impedance must be resonated then the resistance must be transformed to another value. The higher the resistance transformation usually is, the more challenging the impedance matching and the lower the efficiency of the network. The resonance of the capacitance becomes more difficult as the frequency increases because of the lower required inductance dimensions. At one point, the inductance is too small to be physically implemented and any additionally needed interconnection in the circuit alters the inductance total value. As an example, if the minimal implementable inductance is 10pH, it can resonate 25.3pF, 253fF, and 63.3fF at 2GHz, 100GHz, and 200GHz, respectively.

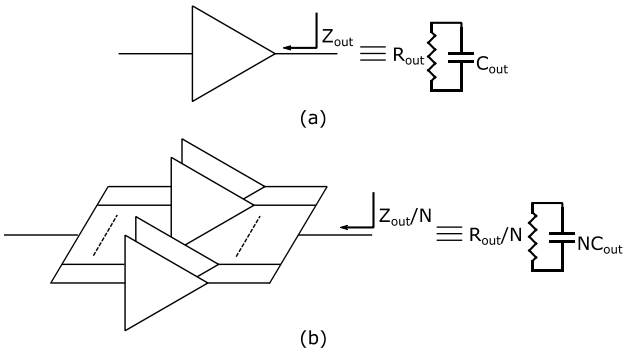


Figure 5.8 Equivalent RC parallel output impedance of an amplifier (a) and of N parallel amplifiers (b)

Figure 5.8 (b) represents N amplification cells connected in parallel to produce a large amplifier. Its equivalent impedance is divided N times compared to the unitary cell. Therefore, the real part is divided by N times as well and the capacitance is multiplied by N times.
Up to a given frequency, the resistive part of the output impedance limits the number of parallel components that can be implemented but above it is the capacitance that limits it. In comparison, the maximal capacitance from the previous example at 2GHz is 400 times higher than at 200GHz. Therefore, a far larger amplifier can be implemented at 2GHz.

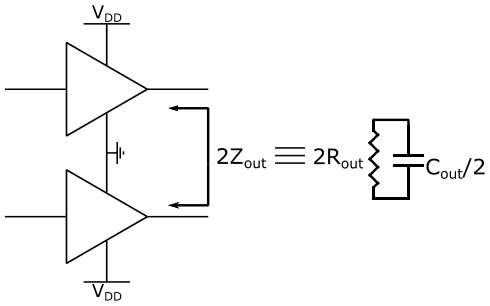


Figure 5.9 Equivalent RC parallel output impedance of a differential amplifier

The differential operation consists of operating two amplifiers with out-of-phase signals. The advantage of this operation mode is the equivalent impedance that is double that of the single amplifier cells. In the differential scheme, the voltages are summed for the same current of a single cell the parallel scheme the total current is doubled while the voltage is the same as for one cell. Consequently, for the same output power, the impedance of a differential scheme has a 4 times higher impedance than for a parallel implementation.
For a maximal acceptable output capacitance, the differential operation allows a 4 times higher output power compared to a single operation.

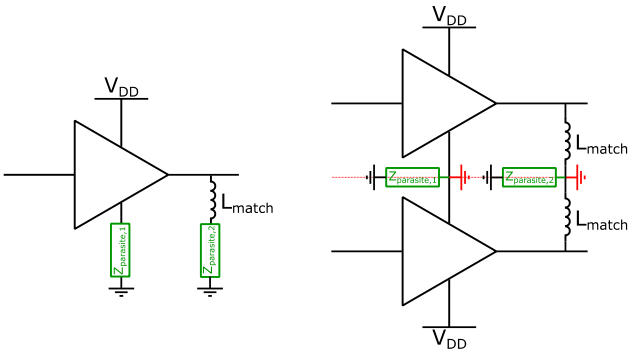


Figure 5.10 Impact of parasitical access components on single scheme operation (a) and differential scheme operation (b)

Another important property of the differential operation is the virtual ground which is a virtual symmetry point between the two unitary amplifiers. In this differential operation, the voltages are therefore referred to as the middle point between the two circuits. The additional access impedances to the circuit ground are transparent and do not alter the RF function of the circuit. As presented in Figure 5.10 (a) the reference of the single amplifier is connected to the ground through a certain path that presents an undesired impedance $Z_{parasite,1}$. In addition, the depicted inductance L_{match} cannot be directly connected to the reference, and it suffers from a parasitical interconnection which degrades the quality and characteristics of the inductance. Conversely, for

the differential scheme presented in Figure 5.10 (b), all these parasitical components are transparent once they pass the middle virtual ground.

The strategy to bias the transistors was transparent so far but voltages and current must be supplied to the transistors, and thus for each successive amplification stage. As presented in Figure 5.11, an amplifier composed of N successive common source amplification stages requires two biases per stage, one for the gate and one for the drain. Consequently, at each one of these nodes, an operation must be realized to separate the DC component from the RF component.

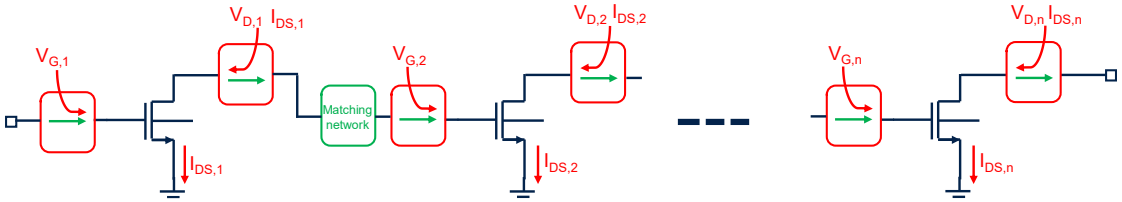


Figure 5.11 Cascaded N CS stages with DC to RF couplers present on each gate and drain

One convenient solution is the use of DC blocking elements and DC feeding elements, that either let pass the DC and block the RF or block the DC and let the RF flow. Usually, DC blocking elements are capacitors that can present a low series impedance for RF signals. However, as the frequency increases the quality factors degrade and the losses are not negligible, in addition, the parasitic capacitances to the ground are not negligible at sub-THz frequencies. For the DC feed elements, a high-value resistance can be used for nodes that do not require current as the gates. It presents therefore a high impedance at RF, occupies a small area, and doesn't have drawbacks. However, for the DC feed elements that carry the supply current as for the drain of transistors, it must present a very low resistance to not dissipate energy and diminish the amplifier efficiency. Inductances can be used for this purpose, but it is challenging to design high-impedance inductances at sub-THz frequencies because of the resonance of the components. Lumped inductors are for this reason not suitable for this frequency range. Quarter wavelength transmission lines are used in return because of their theoretically infinite impedance at the designed frequency. However, despite the high frequency a quarter-wave transmission line is large, narrowband, and adds additional losses at each stage. This solution is usually used for sub-THz single-ended circuits.

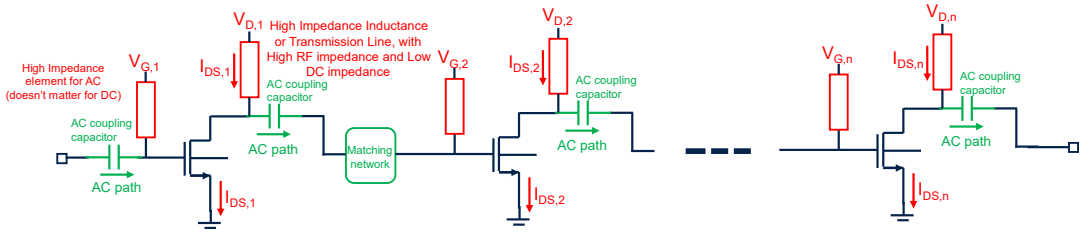


Figure 5.12 Cascaded N CS stages with DC pass and DC block components

Matching networks are also necessary besides the DC coupling and DC blocking elements, but the DC elements can also be integrated into the matching networks. It limits the choice of matching network architectures but allows for a reduction in the number of passive elements.

Matching networks can benefit from the differential operation properties regarding symmetry. In the previous differential example, the parasitic paths were transparent to the RF operation, but they can also be used for the supply. One of these implementations using this symmetrical point for biasing is the use of coupled inductors transformer-based architectures as depicted in Figure

5.13. In the example, the gate and drain voltages are supplied through the midpoint of the inductors used in the transformers. Therefore, the impedances presented at these nodes are not important for differential operation. However, the impedance presented at this node is seen by the common mode that can appear in the amplifier. If the signal is not of equal amplitudes in the amplifier, it can be represented as a sum of two perfectly symmetrical signals plus a signal in phase. Therefore, this in-phase signal does not benefit this symmetrical property. In addition, it can lead to common mode instabilities if the common mode impedances are not controlled. Therefore, common mode terminations are either high impedance for gate voltage paths using large resistances or decoupled with large capacitors at supply nodes to present a low impedance.

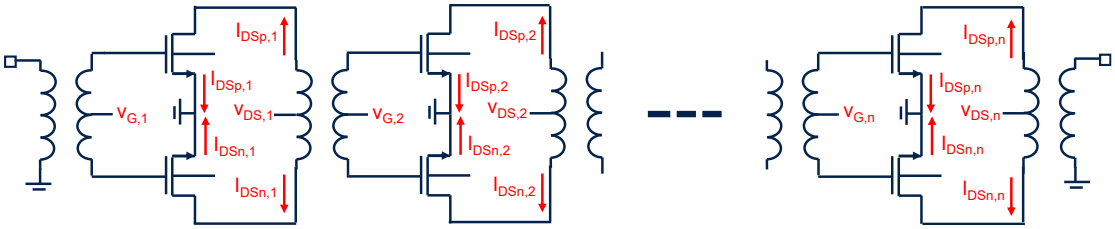


Figure 5.13 Coupled inductors transformer-based N CS stage amplifier

The differential operation using coupled inductor transformers is practical for DC biasing, offers an advantageous impedance transformation, and is robust to interconnection parasites. The frequency response of the transformer can be studied by considering the schematic of Figure 5.14.

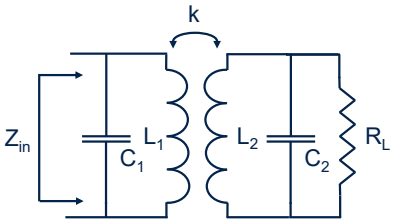


Figure 5.14 Schematic of a transformer with an R//C load

The expression of the input impedance is seen at one port of the transformer while the second is terminated by a resistance load R_L in parallel with a capacitance C_2 that can represent a port from an amplifier or a transistor node. The second port is connected to the inductor L_2 from the transformer. L_2 is magnetically coupled to the inductor L_1 from port 1. A capacitance C_1 is added on the first port for the study of the double LC resonance. The C_1 capacitance can also represent the capacitance of an amplifier output node for an interstage matching case. The impedance Z_{in} can be expressed as:

$$Z_{in} = sL_1 \frac{1 + \frac{s}{\omega_2}\frac{\sqrt{1-k^2}}{Q_s} + \left(\frac{s}{\omega_2}\right)^2 (1-k^2)}{1 + \frac{s}{\omega_2}\frac{1}{Q_s\sqrt{1-k^2}} + \left(\frac{s}{\omega_2}\right)^2 (1+\zeta) + \left(\frac{s}{\omega_2}\right)^3 \frac{\zeta^2\sqrt{1-k^2}}{Q_s} + \left(\frac{s}{\omega_2}\right)^4 \zeta^2(1-k^2)} \quad (5.11)$$

Where:

- $s = j\omega$
- $\omega_1 = \frac{1}{\sqrt{L_1 C_1}}$
- $\omega_2 = \frac{1}{\sqrt{L_2 C_2}}$
- $\zeta = \left(\frac{\omega_1}{\omega_2}\right)^2$

- $Q_s = R_L\sqrt{\frac{C_2}{L_2}}\frac{1}{\sqrt{1-k^2}}$

Equation (5.15) presents the general expression of the impedance of a transformer, and solutions for impedance matching exist for multiple combinations of the transformer parameters. The transformer presents an interesting frequency response when $\backslash zeta = 1$ and the previous equations simplify to:

$$Z_{in} = sL_1\frac{1+\frac{s}{\omega_2}\frac{\sqrt{1-k^2}}{Q_s}+\left(\frac{s}{\omega_2}\right)^2(1-k^2)}{1+\frac{s}{\omega_2}\frac{1}{Q_s\sqrt{1-k^2}}+2\left(\frac{s}{\omega_2}\right)^2+\left(\frac{s}{\omega_2}\right)^3\frac{\sqrt{1-k^2}}{Q_s}+\left(\frac{s}{\omega_2}\right)^4(1-k^2)} \quad (5.12)$$

The equation is still not clear to interpret but three characteristic frequencies can be identified at which the equation simplifies:

$$\omega_s = \frac{\omega_2}{\sqrt{1-k^2}} \quad (5.13)$$

$$\omega_L = \frac{\omega_2}{\sqrt{1+|k|}} \quad (5.14)$$

$$\omega_H = \frac{\omega_2}{\sqrt{1-|k|}} \quad (5.15)$$

The three characteristic frequencies are situated around the LC resonance frequency as presented in Figure 5.15 (the Figure represents the pulsations that appear in the Z_{in} equation).

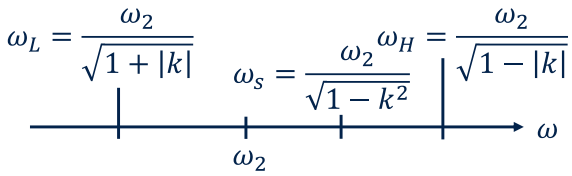


Figure 5.15 Characteristic pulsations of a transformer

The impedance is equal at ω_L and ω_H and its expression is:

$$Z_{in}(s = j\omega_{L,H}) = \frac{L_1}{L_2}R_L\left(1 + j\frac{1}{Q_s}\frac{\sqrt{(1\pm|k|)(1-k^2)}}{1-(1\pm|k|)(1-k^2)}\right) \quad (5.16)$$

The impedance is symmetrical around ω_2 and is proportional to the inductance ratio and the load resistance. In addition, when $Q_s \gg 1$ The imaginary part vanishes, letting:

$$Z_{in}(s = j\omega_{L,H}) \approx \frac{L_1}{L_2}R_L \quad (5.17)$$

For ω_S, the impedance is equal to:

$$Z_{in}(s = j\omega_s) = \frac{L_1}{C_2}\frac{1-k^2}{k^2}\frac{1}{R_L} \quad (5.18)$$

In this case, the resulting impedance is inversely proportional to the load resistance.

Hence, the higher the coupling factor k, the furthest apart are ω_L and ω_H, and the impedances lower greatly at ω_S.

By equating the resonance frequencies, the use of the two symmetrical resonance frequencies is convenient for operating broadband impedance matching. Once the inductance and capacitance values are fixed, the proximity between the inductances permits adjustment of the distance between the resonance frequencies and achieving the desired broadband matching.

To evaluate the efficiency of the transformer the resistivity of the inductors is represented with series resistances. The G_{ma} which represents the intrinsic losses of the passive element can be calculated from the coupled inductors two port. The two-port model is presented in Figure 5.16.

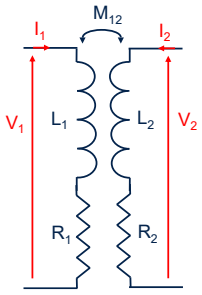


Figure 5.16 Coupled inductors with series resistance two port

The impedance matrix of these two ports is:

$$Z = \begin{bmatrix} j\omega L_1 + R_1 & j\omega M_{12} \\ j\omega M_{12} & j\omega L_2 + R_2 \end{bmatrix} \quad (5.19)$$

Using the G_{ma} definition, its expression is:

$$G_{ma} = 1 + \frac{2}{x}\left(1 - \sqrt{1+x}\right) \quad (5.20)$$

With,

$$x = k^2 Q_1 Q_2 \quad (5.21)$$

The efficiency is dependent on the $k^2 Q_1 Q_2$ quantity. The higher this product is the better the efficiency. Figure 5.17 presents the evaluation of G_{ma} as a function of this product. It indicates that at constant efficiency, the lower the coupling factor, the higher the requirements on the inductors' quality factors.

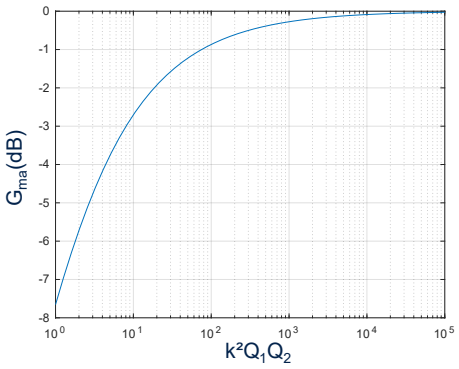


Figure 5.17 Transformer G_{ma} ad a function of the product of the coupling factor and quality factors

In the *x* expression, the geometrical mean of the inductor quality factor can be used to evaluate the efficiency as a function of the magnetic coupling factor k.

$$Q = \sqrt{Q_1 Q_2} \quad (5.22)$$

For the constant inductor quality factor, the efficiency can be calculated as a function of the coupling factor and is presented in Figure 5.18.

In the previous characteristic frequency considerations, the coupling factor k is determined for a broadband operation, hence it is imposed by the design. To keep the loss of the transformer below a certain value, the constraints on the quality factor can be calculated. A graph in Figure 5.18 traces the G_{ma} as a function of the coupling factor for different mean quality factors. Therefore, for

k=0.8 only a quality factor of 10 is necessary to achieve a G_{ma}=-1dB. However, for k=0.5, the constraint increases to 17.5 and for k=0.2 it is above 40.

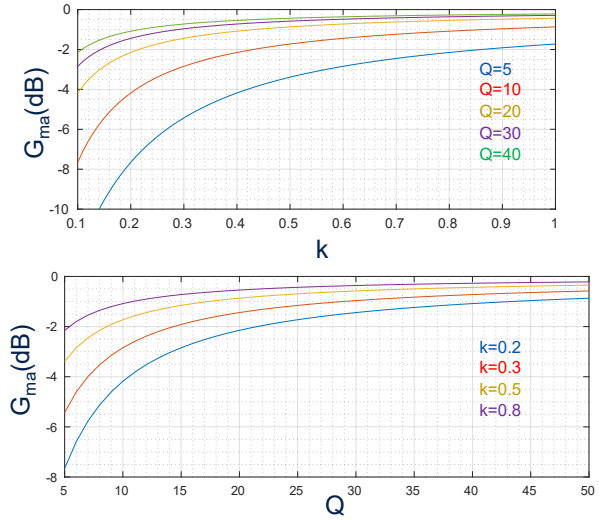


Figure 5.18 Transformer G_{ma} ad a function of the coupling factor (up) and geometrical mean quality factor (down)

To illustrate the previous analysis, Figure 5.19 depicts the design of a coupled inductor transformer. An interdigitated capacitor is designed on the 50Ω side to equate the resonances on both sides of the transformer. The width of the second inductance is larger to produce the transformation impedance. The resulting transformer presents a bandwidth larger than 80GHz around 200GHz, with a loss between -1.2 and -2.5 dB in this frequency range. Efficiency can be improved at the cost of bandwidth. But efficiency is moderate considering the large, achieved bandwidth. The respective inductors are designed to present the peak of quality factors around the frequency of operation as discussed in Chapter 3 on passive devices.

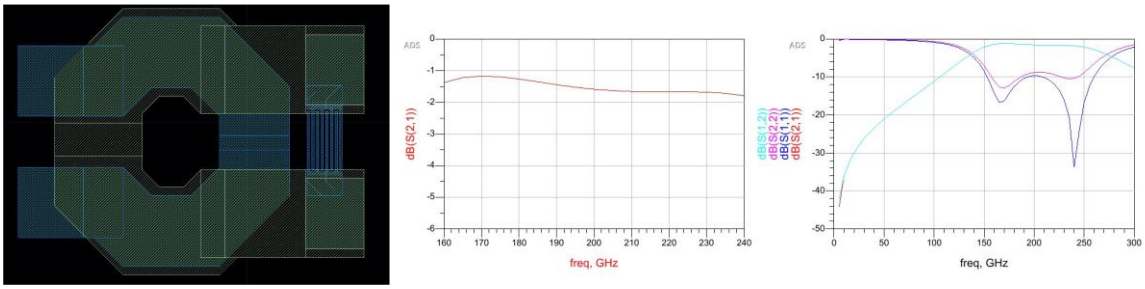


Figure 5.19 Layout view and simulations of an example of broadband operating transformer for a 90Ω//30fF to 50Ω matching

The coupled inductors transformer is an efficient matching network to resonate capacitive components, generate out-of-phase signals, and bias the circuits. For pure resistive impedances, multiple power combiners exist based on transmission lines and coupled transmission lines. Their responses are broadband, offer moderate losses, and allow impedance transformations (on resistive impedances). One convenient power combiner/divider is the Marchand balun [Marc44] which can also be used for differential operation. It can make a broad range of impedance transformations over a large frequency range.

Figure 5.20 presents a schematic of the Marchand balun; it is composed of two successive λ/4 coupled transmission lines. The terminations of some of the terminals as depicted in the figure produce an interesting operation.

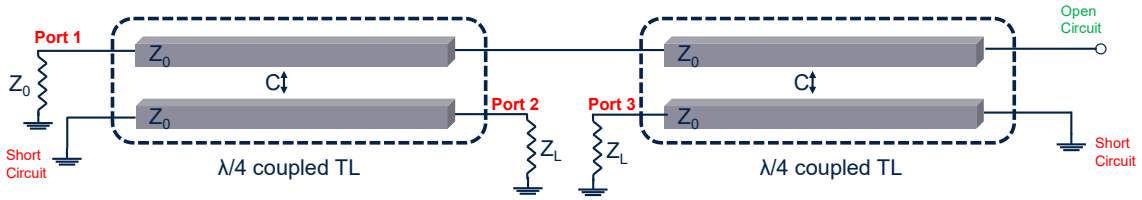


Figure 5.20 Marchand Balun principle schematic

It results in a 3 ports component which S-parameters can be expressed as [Kian01]:

$$S_{balun} = \begin{bmatrix} \frac{1-C^2\left(\frac{2Z_L}{Z_0}+1\right)}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & j\frac{2C\sqrt{1-C^2}\sqrt{\frac{Z_L}{Z_0}}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & -j\frac{2C\sqrt{1-C^2}\sqrt{\frac{Z_L}{Z_0}}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} \\ j\frac{2C\sqrt{1-C^2}\sqrt{\frac{Z_L}{Z_0}}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & \frac{1-C^2}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & j\frac{2C^2\sqrt{\frac{Z_L}{Z_0}}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} \\ -j\frac{2C\sqrt{1-C^2}\sqrt{\frac{Z_L}{Z_0}}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & j\frac{2C\frac{Z_L}{Z_0}}{1+\left(\frac{2Z_L}{Z_0}-1\right)} & \frac{1-C^2}{1+\left(\frac{2Z_L}{Z_0}-1\right)} \end{bmatrix} \quad (5.23)$$

The S-parameters are a function of the characteristic impedance Z_0 which is also the port 1 impedance, the load impedance Z_L, and the transmission line coupling coefficient C.

The coupled transmission lines' coupling coefficient C is a function of the even and odd impedances of the lines. It is defined as [Mong99]:

$$C = \frac{Z_e - Z_o}{Z_e + Z_o} \quad (5.24)$$

The S-Parameters of the Marchand balun simplifies for a certain value of coupling factor:

$$C = \frac{1}{\sqrt{\frac{2Z_L}{Z_0}-1}} \quad (5.25)$$

S-Parameters are then equal to:

$$\mathrm{S_{balun}} = \begin{bmatrix} 0 & \mathrm{j}\frac{1}{\sqrt{2}} & -\mathrm{j}\frac{1}{\sqrt{2}} \\ \mathrm{j}\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ -\mathrm{j}\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (5.26)$$

Equation (5.30) shows that when (5.29) is verified, port 1 is perfectly matched and the signal power is equally divided between ports 2 and 3 with a 180° phase difference. Therefore, the coupling factor of the Marchand balun can be adjusted using (5.29) to operate a required impedance transformation.

Figure 5.21 presents the implementation of a Marchand balun in a 28nm-FD-SOI process at 200GHz to operate a power division from a 50Ω port to two 50Ω ports. The design comprises coupled transmission lines with a coupling factor of 0.577 to fulfill the (5.29) condition. It is realized with coplanar coupled transmission lines to benefit from a higher signal conductor to ground distance than with a Microstrip configuration and achieve the best efficiency. It integrates the RF pad for measurement purposes. The RF PAD does not face a ground plane metal beneath it to

reduce the parasitic capacitance. Therefore, the pad capacitance is only 6.5fF at 200GHz against 36fF for the RF pad with the ground plane that has been used for the previous circuit. For optimal operation, port 1 of the Marchand balun must present the closest impedance to 50Ω. The two load ends of the Marchand balun are connected to microstrip lines that route the signals to two paths. A Microstrip line has been used for this section to allow a ground plane continuity under the line. The circuit simulation results are presented for a 2-port configuration, with a single-ended 50Ω on port 1 and a differential 100Ω between ports 2 and 3. The overall circuit presents a -1.15 dB loss at 200GHz and has less than -1.5dB of loss over a large 100GHz bandwidth. The Marchand balun is designed to divide and combine the power of multiple PAs ended with 50Ω to achieve high power levels.

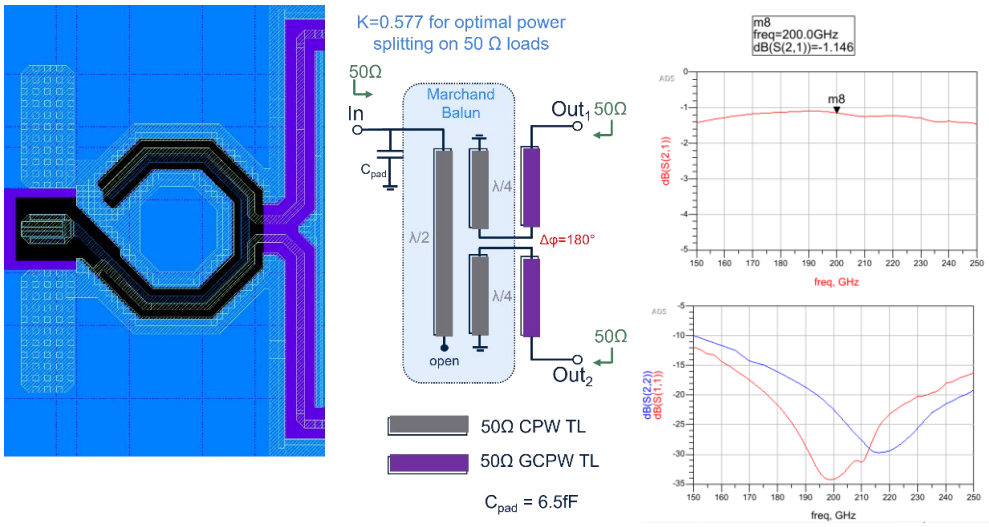


Figure 5.21 Designed Marchand balun at 200GHz in 28nm-FDSOI with layout view, schematic, and simulated S-parameters (with a differential 100Ω port between ports 2 and 3)

The Marchand balun dimensions and its distributed nature allow the absorption of the capacitive and inductive elements of the RF pad without altering the main behavior of the component. In addition, the coplanar transmission line can be easily connected to the RF pad. Conversely, the balun based on coupled inductors is sensible to parasitics as its resonances are tuned. Because of the targeted transistors with a large area in the designed PAs, the inductors from the balun have a low value and any additional parasitical inductance from interconnection to the pad prevents the balun from working at all. As a last drawback, the distance between the signal and ground pads is dictated by the measurement instruments and isn't compatible with the balun ports' distance. The solution to this challenge is presented in Figure 5.22, the pad is connected to a transmission, whose characteristic impedance has been tuned to absorb the pad parasitics. The distance to the circuit allows bringing the transmission line signal and ground ends to the required positions for the balun. The additional distance minimizes the eventual coupling between the RF probes and the balun structure. With a differential port situated at the required position in the layout view of Figure 5.39, the transition presents an impedance close to 50Ω and only a -0.2dB loss. The simulated baluns operate properly with this transition to the RF pads.

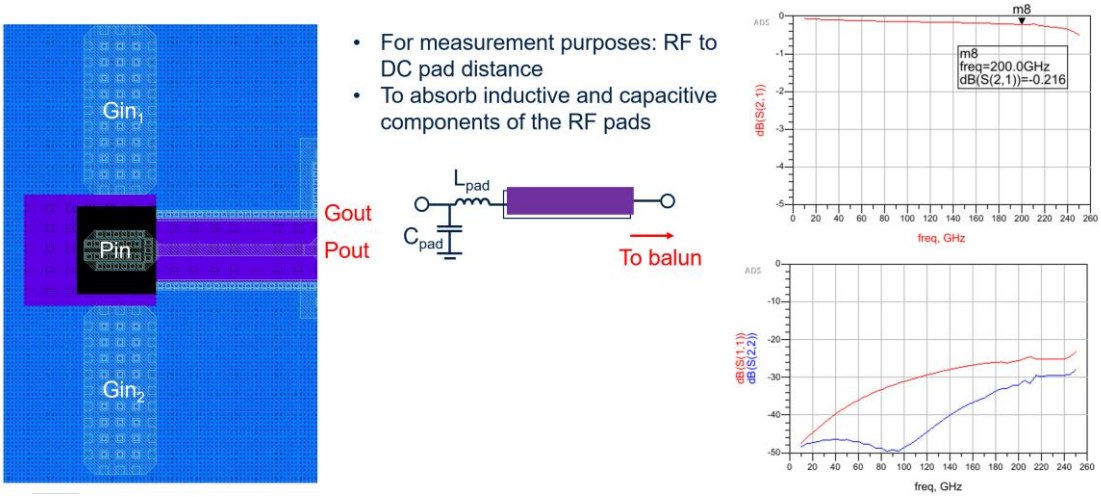


Figure 5.22 Designed and simulated transition between the RF pads and the balun

5.5 High power Amplification cells, CS, and Cascode comparison

Based on the previous analysis of the design of high output power amplification structures, multiple differential amplifier stages have been designed, two sizes of CS, one enhanced cascode structure to operate at 200GHz, and CS and Cascode structures using gain boosting

5.5.1 Differential neutralized Common Source

Two differential common source amplification stages have been designed using 3-transistor and 6-transistor matrix cells (the 6-transistor matrix has been presented in the previous section and the 3-transistor is half of this structure). The neutralization is done using transistors whose gate and sources have been shorted, to present a capacitance of the same process nature as the functional transistors from the amplifier. The neutralization transistors are smaller due to the sum of their C_{gd} and C_{ds} capacitances. The dimensions of the different transistors are given in Figure 5.23. The amount of neutralization is chosen to offer unconditional stability and thus good isolation at all frequencies.

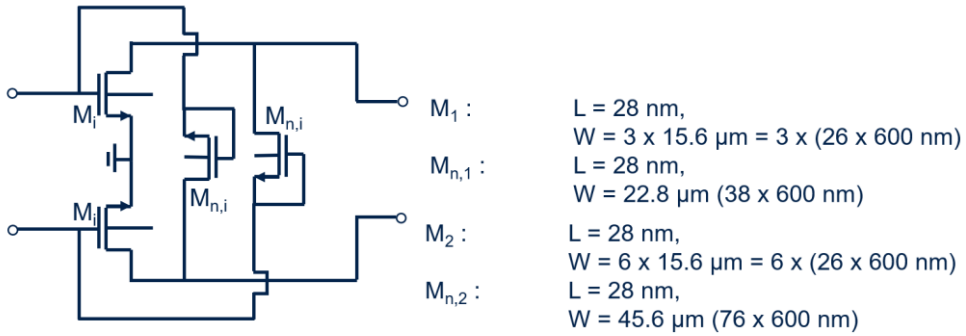


Figure 5.23 Schematic of the designed differential neutralized CS amplifier and the transistor dimensions

The simulation strategy is the same as for the transistor matrix simulation. The thin metal and active devices are RCc extracted, and the thick metal is EM simulated. The sum of the two simulations is used to simulate the behavior of the amplifiers. The 3-transistor matrix-based CS layout is presented in Figure 5.24.

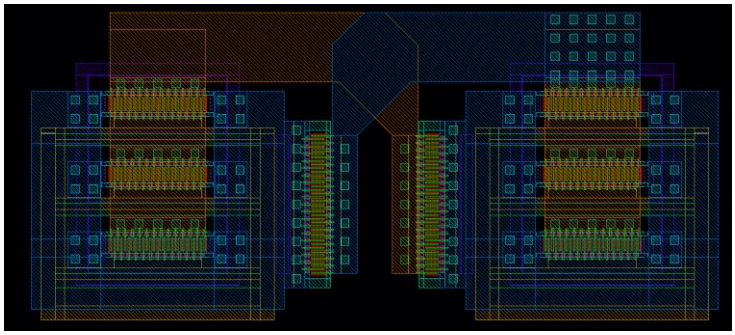


Figure 5.24 3-transistor matrix differential neutralized CS amplification stage

The distance between the two paths is kept as short as possible to reduce inductive degeneration thanks to the virtual ground and the parasitical inductances in the cross-coupled capacitances. The latter inductance can alter the frequency response of the neutralization operation. The same design and simulation process has been applied to the 6-transistor matrix-based amplifier and its layout is presented in Figure 5.25.

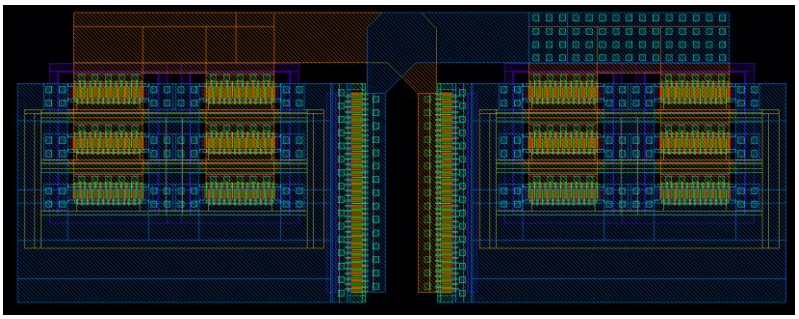


Figure 5.25 6-transistor matrix differential neutralized CS amplification stage

The gain performances of these two design stages are presented in Figure 5.26. For the smaller version, the maximally available gain G_{ma} is between 6.6dB and 4.5dB between 180GHz and 220GHz which is the targeted frequency band. For the larger version, the G_{ma} is similar and is between 6.1 and 2.8dB between 180GHz and 220GHz.

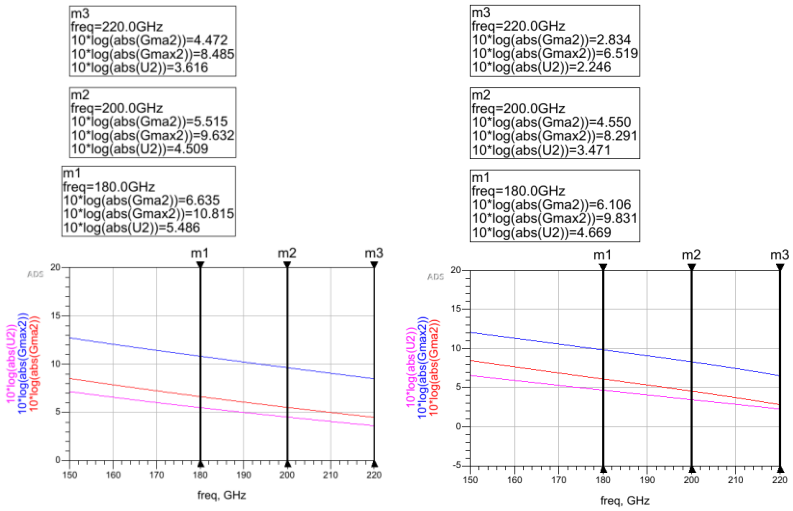


Figure 5.26 Simulated gain metrics of the CS stages: 3-transistor matrix (left) and 6-transistor matrix (right)

Figure 5.27 reports the R and C parallel input and output impedances of the two amplifiers. These values are used to design the input and interstage transformers with the methodology described previously to operate between 180 and 220GHz.

	R_{in} (Ω)	C_{in} (fF)	R_{out} (Ω)	C_{out} (fF)
Ampli n°1	92	29.9	113	32.6
Ampli n°2	32	72.25	46.8	66.7

Figure 5.27 Equivalent RC components of the input and output impedances from the designed amplification stages

5.5.2 Differential Cascode with internal matching

The capacitive components of the CS structure that have been described are high and an additional increase in transistor width is not practical. To design a structure that can produce an even higher output power, the maximal voltage must be increased. The Cascode structure uses two stacked transistors, hence a 2V DC can be used as the supply. The designed Cascode structure is depicted in Figure 5.28. It has been optimized to overcome multiple challenges that arise at these frequencies with a Cascode structure. Neutralization is used to improve the performance of the CS transistors and cancel the Miller effect. Without neutralization, an internal resonance occurs between the impedances that are presented by the CS and the CG transistors. However, with neutralization, the capacitance at the node joining the CS and CG increases as has been presented in Chapter 4. Hence, with neutralization this equivalent capacitance to the ground induces an impedance mismatch between the two transistors and the resulting gain is close to 0dB. To alleviate this phenomenon, this equivalent capacitance is resonated using the L_{match} inductance from the schematic. Thanks to the differential operation, the inductance is connected between the positive and negative paths to operate as an inductance connected to the virtual ground in differential operation. The two gates from CG transistors are connected to create a perfect connection to the virtual ground in the differential. Because the cascode is prone to instability, especially with the impedance presented at this gate node from the CG transistor, the common mode is not decoupled but only connected to the supply voltage using a high-value resistance. In addition, this allows for the reduction of the common mode gain of the stage. The 3D view of the resulting Cascode stage is presented in Figure 5.28.

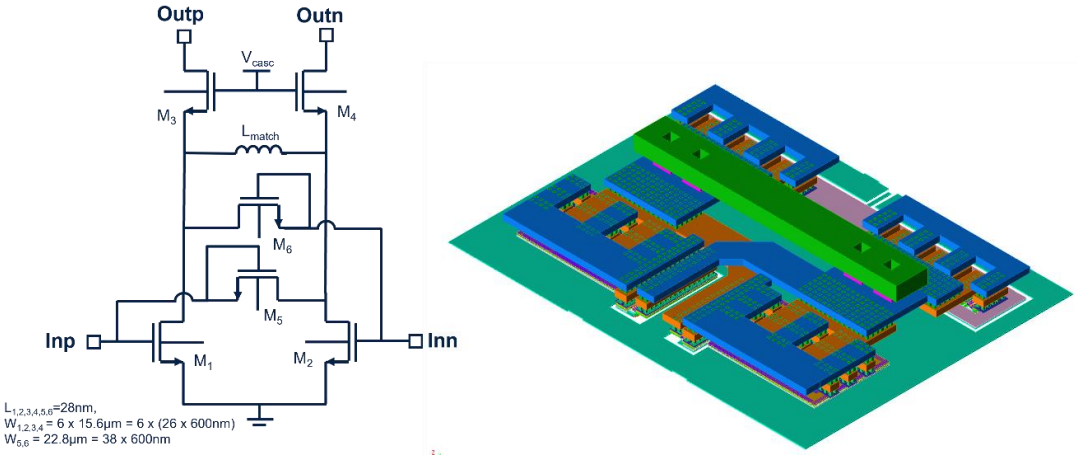


Figure 5.28 Differential Cascode with internal matching schematic and 3D view

The simulation methodology is the same as previously and the resulting performances are presented in Figure 5.29. The available gain presents low variation along the 180Ghz to 220GHz band, and its form indicates the internal inductive resonance of the Cascode stage. The P_{sat} of the amplifier is 12dBm and has a maximal PAE of 3.5%

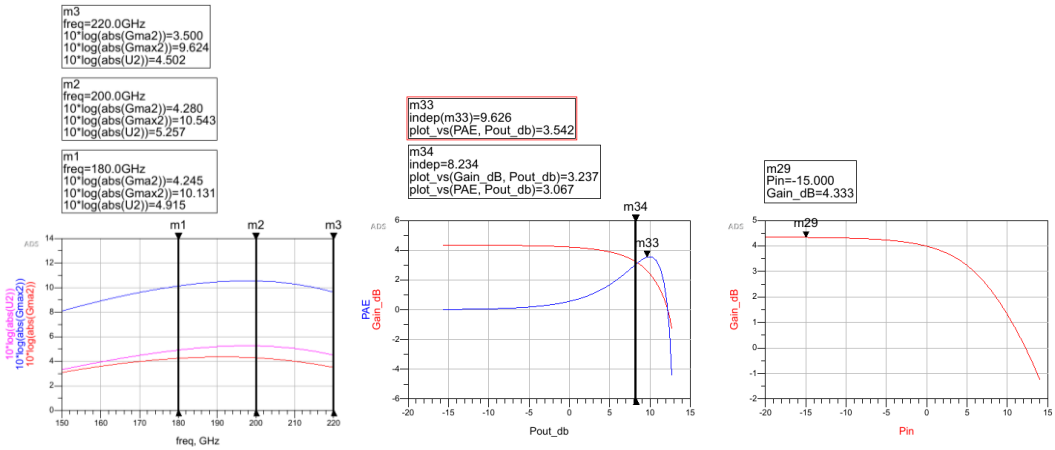


Figure 5.29 Simulations of the Gain performances and Large signal performances of the Cascode stage.

5.6 Large band gain boosting

The designed CS and Cascode structures are limited in terms of the gain boosting required to increase the available gain. The embedding equation solution that has been presented in Chapter 4 is evaluated for the parameters of the transistor matrixes. Figure 5.30 presents the required X_{par} impedance to boost the Gain to G_{max} when X_{out}=0. X_{in} is close to 0 and only X_{par} is enough to significantly increase the G_{ma}. Figure 5.30 presents the view of the studied 6-transistor matrix and the required X_{par} values for 180GHz, 200GHz, and 220GHz. The designed feedback impedance must present these impedances at these frequencies to achieve the targeted large bandwidth gain boosting. A transmission line parallel impedance is a sinusoidal function of the electrical length, at low length it is inductive, and the impedance is proportional to frequency but passed λ/4 the impedance decreases up to λ/2 where the period repeats. Therefore, the TL length can be chosen so that its impedance matches the variations of the requirements on X_{par}. A TL of characteristic impedance 24Ω and relative wavelength 0.318λ presents the required impedance variations as presented in Figure 5.30.

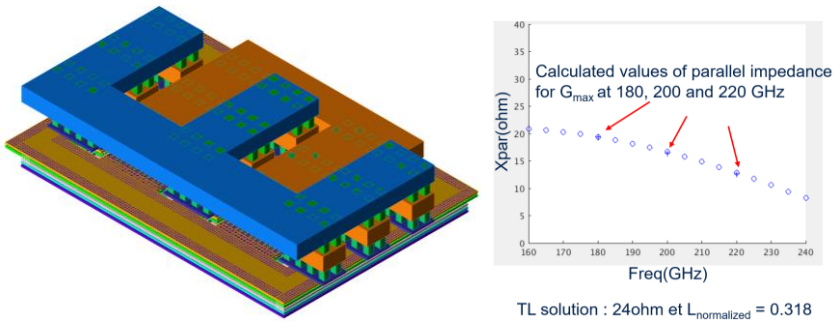
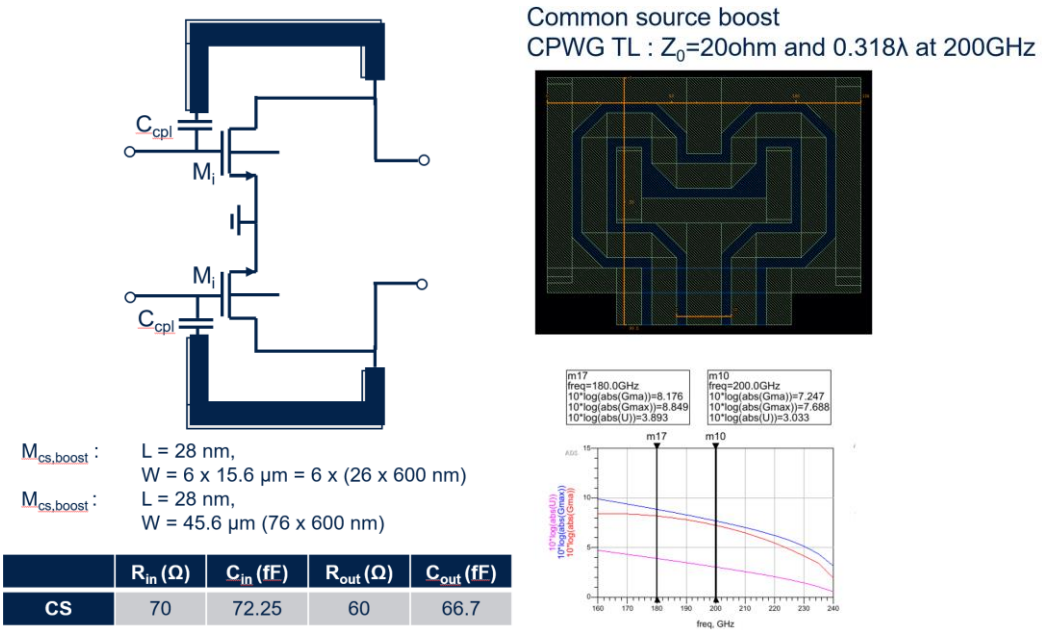


Figure 5.30 6-transistors matrix used for broadband gain boosting (left) calculated requires values of X_{par} for this transistor and the impedance over frequency of the chosen TL

The TL is implemented using a CPWG to achieve the required characteristic impedance and a coupling capacitance has been added in series with the line for DC biasing. When the extracted transistor is simulated with the simulated feedback network, it achieves a 7.28dB gain at 200GHz. The G_{ma} is close to G_{max} over the 170-230GHz band.



$M_{cs,boost}$: L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)

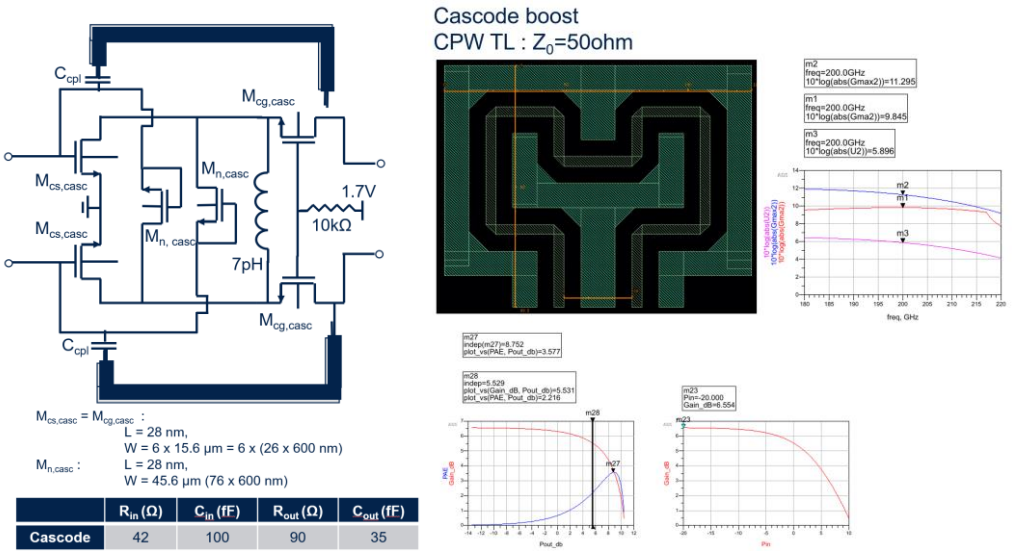
$M_{cs,boost}$: L = 28 nm, W = 45.6 µm (76 x 600 nm)

	R_{in} (Ω)	C_{in} (fF)	R_{out} (Ω)	C_{out} (fF)
CS	70	72.25	60	66.7

Figure 5.31 Designed boosted CS stage schematic, designed feedback TL, and simulated resulting gain of the amplifier stage.

The analytical boosting design solution permits to analysis of the embedding requirement variations along frequency and then the design of the feedback TL with the appropriate frequency variation. This procedure can be applied to single transistors but also to the amplification stage on the same principle by representing them by a two-port.

A large bandwidth gain boosting feedback has been designed for the differential Cascode stage. As presented in Figure 5.32 the feedback TL is added above the internal Cascode matching. The same coupling capacitance is used for DC purposes at the end of the TL. The ideal slope of the TL to achieve the large bandwidth gain boosting is near 0.318λ as well but the amplitude requires a 50Ω TL. It has been implemented using a CPW TL to maximize its quality factor. The resulting simulated performances are presented in this figure. It presents a G_{ma} of 9.8dB from 180GHz to 217GHz. This flat response is possible because of the flat gain response of the Cascode stage that is then boosted equally all over the band.



$M_{cs,casc} = M_{cg,casc}$: L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)

$M_{n,casc}$: L = 28 nm, W = 45.6 µm (76 x 600 nm)

	R_{in} (Ω)	C_{in} (fF)	R_{out} (Ω)	C_{out} (fF)
Cascode	42	100	90	35

Figure 5.32 Designed boosted Cascode stage schematic, designed feedback TL, and simulated resulting gain and large signal performances of the amplifier stage

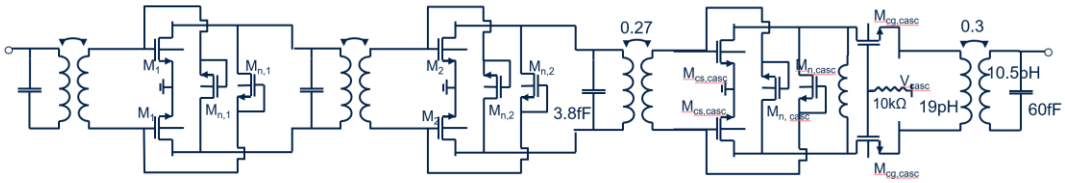
The large signal performances present an almost 11dBm P_{sat}, a maximal efficiency of 3.6%, and a small signal gain of 6.5dB. These large signal performances are achieved at 200GHz for a load that has been optimized through a load-pull on the structure. The presented optimal load is 90Ω with an inductive parallel component that is frequency-dependent. This inductive component cancels the capacitive component from the PA stage. Therefore, the optimal load impedance through the 180GHz to 220GHz is close to the conjugate of 90Ω in parallel with 35fF. It can be referred to as the “large signal” output impedance of the amplifier.

5.7 Implementation of PAs in 28nm FDSOI

Multiple versions of PAs have been designed using the previously presented circuit blocks to operate at 200GHz in 28nm FD-SOI technology.
Two approaches were pursued: a structure without gain boosting, composed of a neutralized differential CS stage and a Cascode stage. A boosted structure also uses three stages with one neutralized CS stage, followed by one CS differential gain boosted stage and a gain boosted Cascode stage. A third stage has been added at the input for the isolation intention, especially for the gain-boosted version as the gain boosting intrinsically lacks isolation, and the impedances inside the amplifier interstage nodes can vary depending on the RF inputs impedances. This solution allows us to improve the overall isolation and provide additional gain.

5.7.1 Classic 3 stages PA

The “classic” implementation, without gain boosting, is presented in Figure 5.33. The transformers have been designed following the presented methodology to present broadband matching. The power capability of the three stages is increasingly doubling. The second stage has the double of transistors compared to the first stage and the cascode structure uses the same transistor size as the second stage but uses two of them stacked. Because of the low gain, a greater size difference is not preferred for this design.



M_1 : L = 28 nm, W = 3 x 15.6 µm = 3 x (26 x 600 nm)
$M_{n,1}$: L = 28 nm, W = 22.8 µm (38 x 600 nm)
M_2 : L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)
$M_{n,2}$: L = 28 nm, W = 45.6 µm (76 x 600 nm)
$M_{cs,casc} = M_{cg,casc}$: L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)
$M_{n,casc}$: L = 28 nm, W = 45.6 µm (76 x 600 nm)

Figure 5.33 Schematic of the classic PA implementation in 28nm-FDSOI

Figure 5.34 presents the small signal simulations of the amplifier. It has a good S_{11} over 80GHz around 200GHz. The S_{22} does not depict the output matching as it has been designed for large signal operation. The achieved power gain is 15.44dB at 167GHz and decreases linearly to 6dB at 220GHz. This response is due to the intrinsic natural response of transistors. The fall of the gain translates to the technological limit.

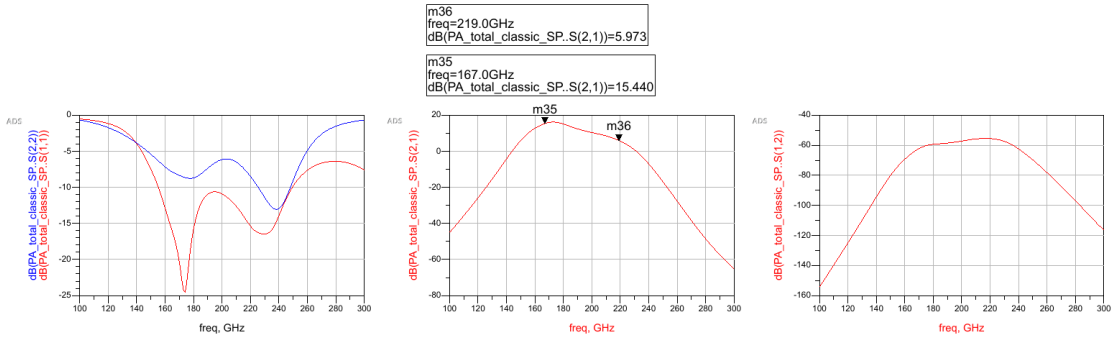


Figure 5.34 S-parameters simulations of the classic PA implementation

The large signal simulations at 200GHz are presented in Figure 5.35. The amplifier presents 11.5dBm P_{sat} and a maximal PAE of 4.75%. The resulting OCP_1 is equal to 6.5dBm.

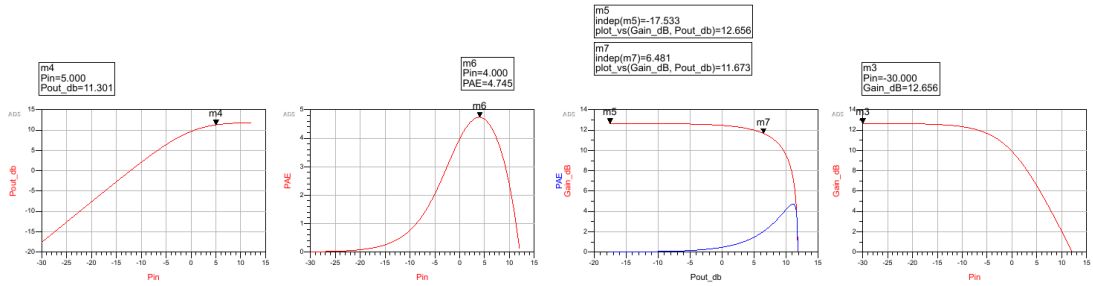


Figure 5.35 Large signal simulations at 200GHz of the classic PA implementation

Figure 5.36 depicts the implemented layout of the amplifier. The transition to the RF pad that has been presented previously has been used to connect to the input and output balun. The structure is symmetrical, with decoupling capacitances at the north and south of the structure. Similarly for gate biasing, the biasing is symmetrical up to high-value resistances at north and south.

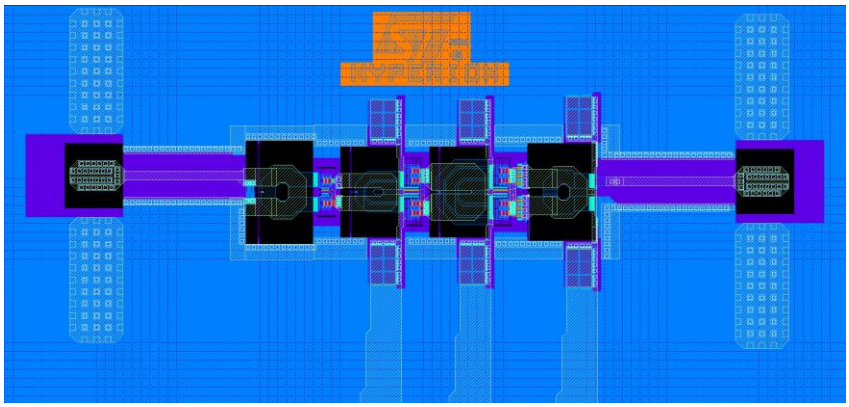
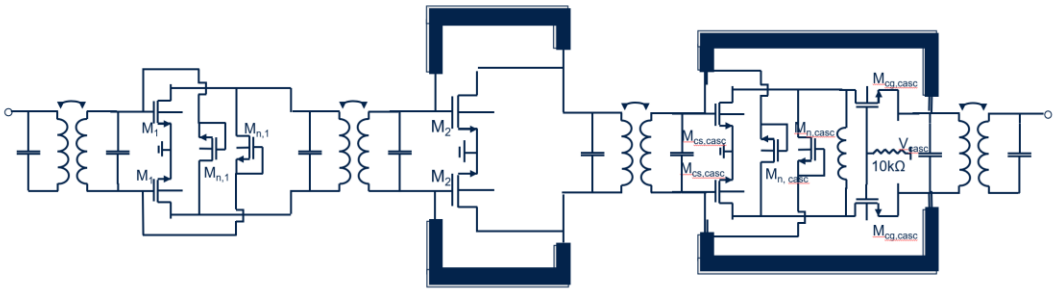


Figure 5.36 Layout view of the classic PA

5.7.2 Boosted 3 stage PA

The complementary design to the classical approach has been designed using two gain-boosted amplification stages. The schematic of the resulting amplifier is presented in Figure 5.37.



M_1 : L = 28 nm, W = 3 x 15.6 µm = 3 x (26 x 600 nm)
$M_{n,1}$: L = 28 nm, W = 22.8 µm (38 x 600 nm)
M_2 : L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)
$M_{cs,casc} = M_{cg,casc}$: L = 28 nm, W = 6 x 15.6 µm = 6 x (26 x 600 nm)
$M_{n,casc}$: L = 28 nm, W = 45.6 µm (76 x 600 nm)

Figure 5.37 Schematic of the boosted PA implementation in 28nm-FDSOI

Figure 5.38 presents the S-parameters simulations. The achieved S_{11} is adapted over the large, targeted bandwidth. The S_{21} presents a higher performance than the classic approach. A 17.9 dB is achieved at 185GHz, and it is still equal to 14.23dB at 214 GHz. The 3dB bandwidth is almost 40GHz wide as targeted. However, the gain falls faster around the band because of the gain boosting. Around the targeted frequency band, the gain is not boosted and falls. In addition, the impedance varies rapidly outside of this gain-boosted frequency range. Finally, the S_{12} simulation is 20dB higher than the classical approach. This is because of the positive feedback and the isolation is due to the first neutralized CS stage.

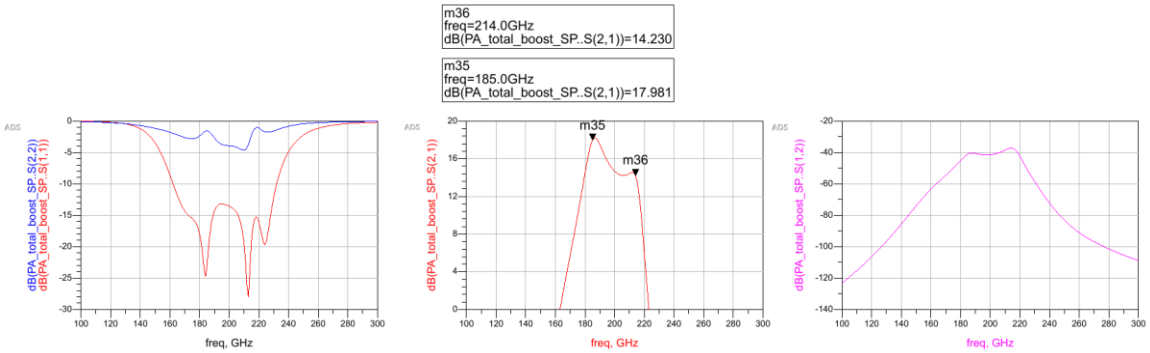


Figure 5.38 S-parameters simulations of the boosted PA implementation

The large signal simulations at 200GHz are presented in Figure 5.39. The P_{sat} is 10.6dBm and the maximal PAE is 4%. The OCP_1 is at 5dBm. The small signal gain at 200GHz is 15dB.

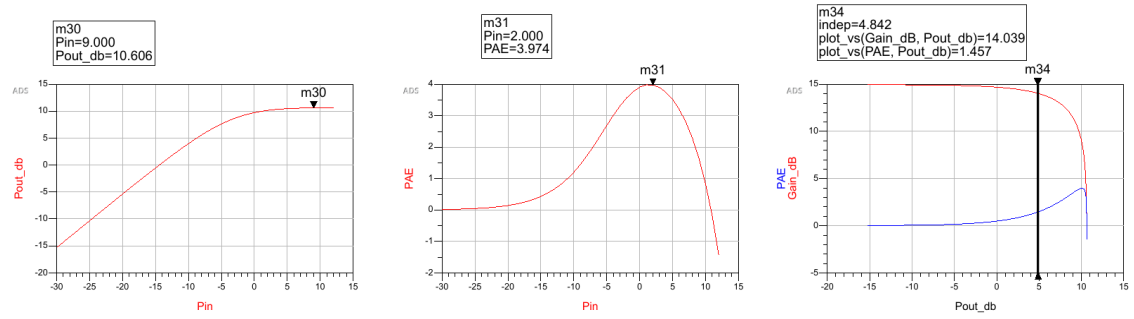


Figure 5.39 Large signal simulations at 200GHz of the boosted PA implementation

Figure 5.40 presents the layout view of the implemented boosted PA at 200GHz.

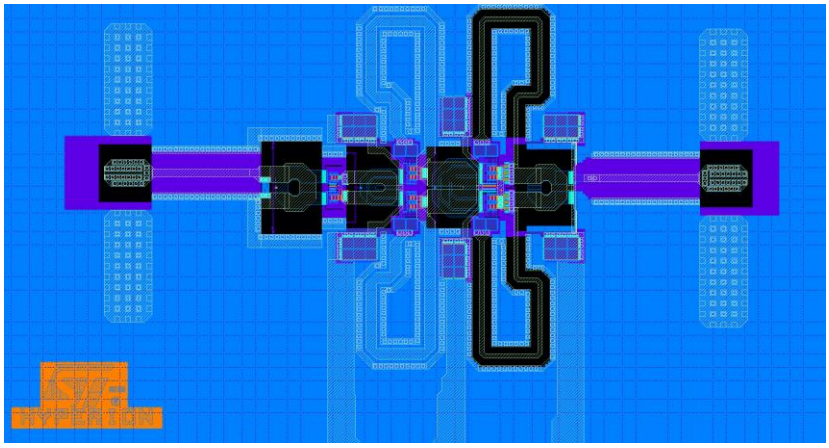


Figure 5.40 Layout view of the boosted PA

To increase the power capability of the PA, two paths combined PA have been implemented for both classic and boosted approaches. The previously described 50Ω Marchand Balun was used as a power splitter and combiner.

The symmetry of the PAs and the 50Ω matching allowed us to implement the two-path design without any modification on the unitary PAs.

Figure 5.41 presents the classic approach and Figure 5.42 the boosted solution.

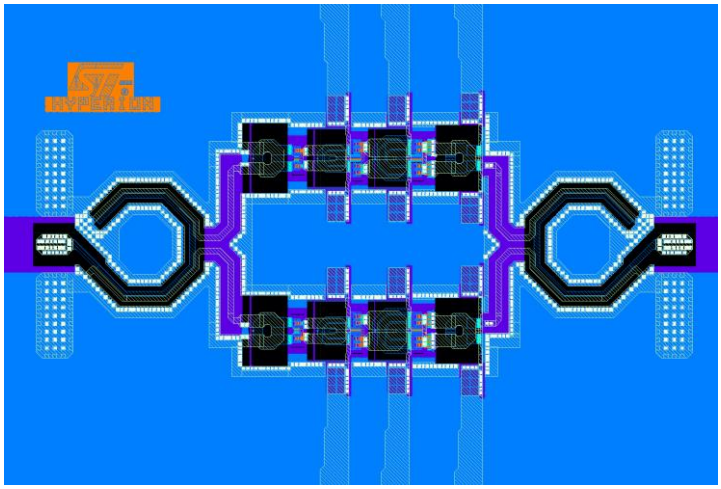


Figure 5.41 Layout view of the classic two-way PA

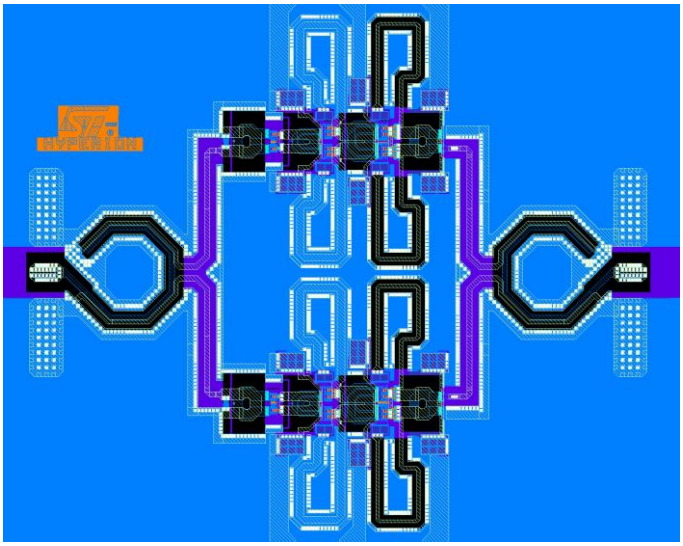


Figure 5.42 Layout view of the boosted two-way PA

5.8 Measurements and analysis of results

The circuits have been fabricated in the 28nm FD-SOI process from STMicroelectronics and the measurements have been proceeded on-wafer with the same Rhode & Schwartz ZVA 24 VNA as for the previous circuit in the IMS laboratory of the University of Bordeaux.
Figure 5.43 depicts the chip microphotograph of the simple classic PA and the S-parameters measurements.
The achieved gain is 3.5dB with a 170-205GHz bandwidth. The gain is lower than expected and the reasons principally come from the design of the interstage matching transformer between the second and third stages. As presented before, the input impedance of the Cascode structure presented an important 100fF capacitance that was resonated using the transformer. However, the required inductance of low value was designed using very large conductors and was at the limit of the feasibility. The electrical coupling of this transformer with its environment that was not precisely captured in the EM extractions produced a frequency shift in operation but also a resonance frequency decrease of the transformer. This resonance frequency limits the operation frequency of the transformer, and the performance falls rapidly as this resonance decreases.

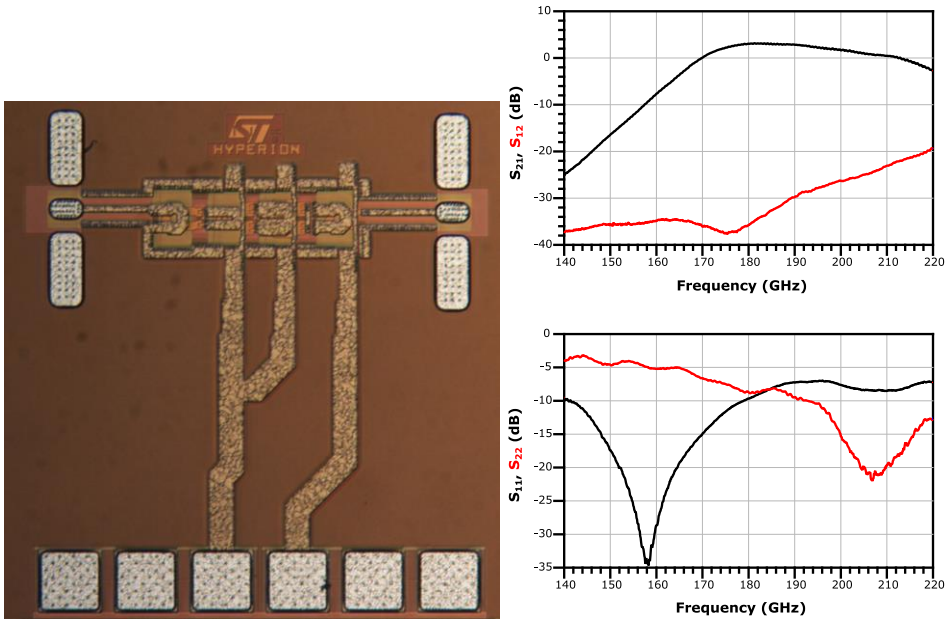


Figure 5.43 Chip microphotograph of the simple classic PA (left) and the measured S-parameters (right)

The chip microphotograph of the double paths classic PA is presented in Figure 5.44 next to its S-parameters measurements. The results are coherent with the simple implementation. The performances are a couple of dBs apart and the frequency behavior in terms of “bandwidth” is coherent. The necessary modification to improve these performances and overcome this transformer bottleneck comes from diminishing this input impedance capacitance. The size of the devices has been chosen with a margin regarding the possible matching, but the additional layout components and the compactness of the design significantly increased these parasitical capacitances. The designed amplifier was functional and at the limit of the possible capacitance resonance. Smaller devices presenting a 70fF capacitance could give enough margin to produce a better performance.

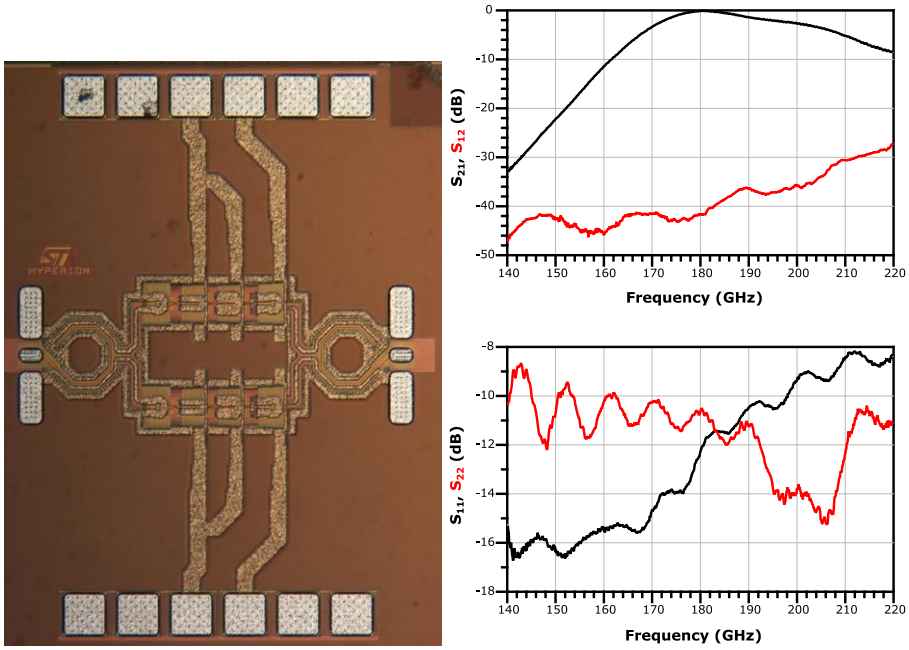


Figure 5.44 Chip microphotograph of the double classic PA (left) and the measured S-parameters (right)

Figure 5.45 presents the boosted PA version chip microphotograph and the measured S-parameters. The behavior was not expected in simulations and the bandwidth was not achieved. The PA shows signs of available gain and the boosting signature is visible. The side gain increase is due to the gain boosting signature on the band limits. Multiple points have been identified and that are the root of this behavior. The same problem of the interstage transformer changes the frequency response of the amplifier. The feedback transmission lines have been modified in form to fit with the supply connections and the small size of the transformers. Coupling between these elements that haven't been simulated properly changes the effective conditions of the feedback system. Therefore, the conditions to produce a G_{max} are not fulfilled and the gain can decrease rapidly.

To block the DC bias of the V_{DD} from the gate bias the coupling capacitors used in the feedback paths were used from the DK as their series values were not important (their values were high enough to be transparent). However, the parasitical capacitances of these capacitors were underestimated, and the real value where not considered. The complexity of the MiM capacitors prevented their integration into EM extractions.

To improve this performance, a custom interdigitated capacitance is required to be able to model the parasitical capacitances. The achievable capacitance values are not important enough to be transparent, the feedback transmission line must therefore integrate this capacitor impedance. The design was robust to this parasitical capacitance value, but a significant change could lead to an important frequency shift of the matching network.

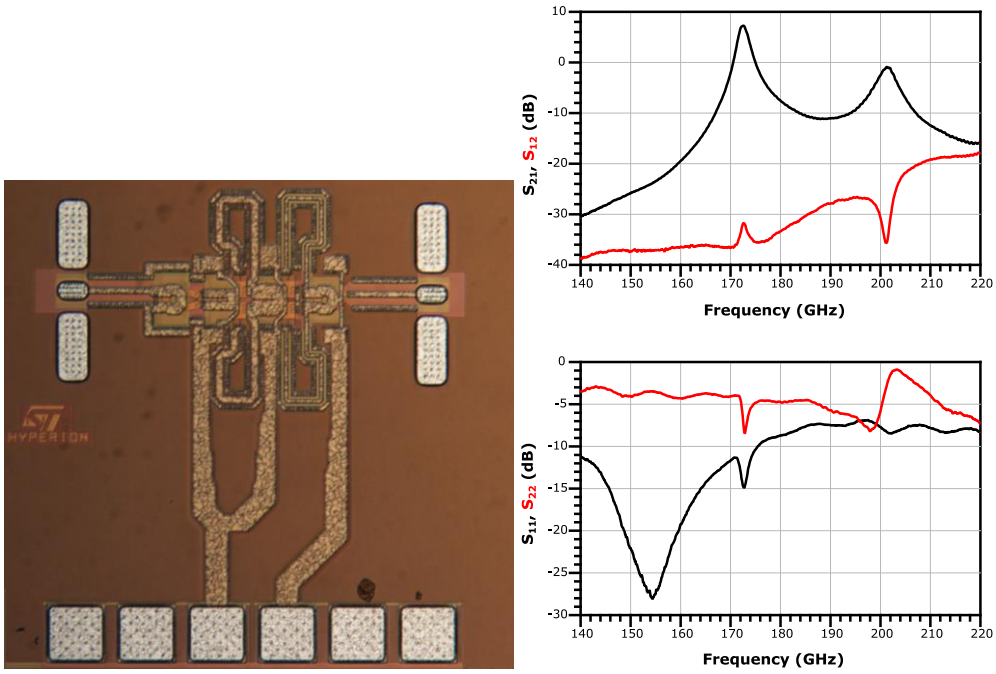


Figure 5.45 Chip microphotograph of the simple boosted PA (left) and the measured S-parameters (right)

The combined version of the boosted PA is presented in Figure 5.46. The frequency behavior of the gain is like the simple boosted PA.

The S_{11} and S_{22} are below -10dB for this double path design as for the classical PA combined design. It is the signature of the Marchand balun that presents a good matching along all the measurement setup 140-220GHz frequency band.

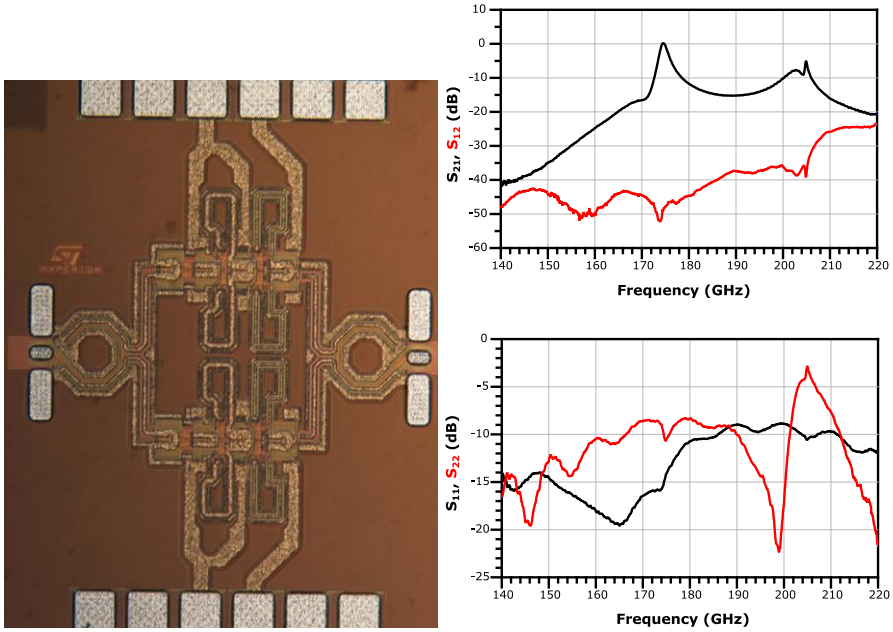


Figure 5.46 Chip microphotograph of the double boosted PA (left) and the measured S-parameters (right)

5.9 Conclusion on results

This Chapter presented the high-power challenges and limitations of a high-power output PA, especially at sub-THz frequencies where the intrinsic gain is limited. Design methodologies were presented, and an implementation highlighted the improvement aspects that must be corrected to improve the performance of the targeted PA

The boosting is vulnerable to model inaccuracies as the positive feedback brings the right amount of the signal back to the input. If the intrinsic gain is higher or lower than expected, the behavior of the amplifier is altered. The design was realized with the typical corner, but the previous measurements showed a slight difference in real-world operation than simulated with the model. With one transistor amplifier, the design was purposely robust to eventual variation but this time because of the large bandwidth gain boosting, it was designed for the simulated conditions of operation. Despite very good modeling, this kind of operation is sensitive to variations. The behavior of the circuit over corners showed an increase, a decrease, or a slight frequency shift in the operation but this measured response was not predicted. For classic implementation, the model is good enough to predict the transistor behavior as the difference between simulation and measurements in this case comes from the transformers.

To alleviate these problems, a design with half of the transistor sizes may be a solution at the cost of a lower P_{sat}.

The purpose of this design was to delimit what are the achievable design limits at these frequencies using techniques to allow the circuit to operate this high in frequency. The limit of P_{sat} is clear and with this size of transistors, it is near this limit as matching networks cannot resonate such high capacitances with a margin for simulation inaccuracies due to extraction mistakes.

The large bandwidth gain boosting is effective as the gain appears to increase at the designed frequencies, a less dense design with better top extraction may be the solution. The transformers and Marchand balun design showed the expected large bandwidth operation and confirmed at least partially the designed strategy.

5.10 References

[Marc44] N. Marchand, “Transmission line conversion transformers,” Electronics, 1944.

[Kian01] Kian Sen Ang and I. D. Robertson, "Analysis and design of impedance-transforming planar Marchand baluns," in IEEE TMTT, 2001

[Mong99] Rajesh Mongia - RF and Microwave Coupled-Line Circuits-Artech, House Publishers, 1999.

[Park19] D. Park et al., "A 230–260-GHz Wideband and High-Gain Amplifier in 65-nm CMOS Based on Dual-Peak Gmax-Core" in IEEE Journal of Solid-State Circuits, June 2019

[Ko13] C. Ko et al., "A 210-GHz Amplifier in 40-nm Digital CMOS Technology," in IEEE Transactions on Microwave Theory and Techniques, June 2013.

[Wang14] Z. Wang, P. -Y. Chiang, P. Nazari, C. -C. Wang, Z. Chen and P. Heydari, "A CMOS 210-GHz Fundamental Transceiver With OOK Modulation," in IEEE Journal of Solid-State Circuits, 2014

[Bame20] H. Bameri, et al. “An Embedded 200 GHz Power Amplifier with 9.4 dBm Saturated Power and 19.5 dB Gain in 65nm CMOS”, RFIC 2020

[Sadl21] S. Sadlo, et al., "One stage gain boosted power driver at 184 GHz in 28 nm FD-SOI CMOS," 2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC).

[Parv16] D Parveg et al., “A 124-184 GHz Amplifier Using Slow-wave Transmission lines in 28-nm FDSOI CMOS Process”, GSMM 2016

[Seo09] M Seo et al., “A 150 GHz Amplifier With 8 dB Gain and +6 dBm Psat in Digital 65 nm CMOS Using Dummy-Prefilled Microstrip Lines”, JSSCC 2009

[Li21] S. Li et al., “A 130-151 GHz 8-Way Power Amplifier with 16.8-17.5 dBm Psat and 11.7-13.4% PAE Using CMOS 45nm RFSOI”, RFIC 2021

[Simi18] D Simic et al., “A 14.8dBm 20.3 dB power amplifier for D-band applications in 40nm CMOS”, RFIC 2018

Chapter 6 Conclusions and perspectives

6.1 Summary

In this work, the technological and design limitations towards the implementation of sub-THz communication circuits in CMOS circuits have been studied. The precise analysis of the state-of-the-art power amplifiers above 100GHz highlighted the areas that required further research in Chapter 2. The theory of linear ports analysis and the characteristics of active and passive elements from CMOS processes have been recalled in Chapter 3. The limiting phenomenon in the transistor's power gain capability has been presented. The characteristics of the passive components whose efficiency is crucial have been presented and explained to try to improve their performances, especially in VLSI integrated circuit technologies. The analysis of the transfer function of the transistor-based common source amplifier in Chapter 4 was followed by the analysis of techniques such as neutralization to overcome the low performances at sub-THz frequencies. A comparison to other gain improvement techniques has been made to identify the advantages and drawbacks of each one of them. The gain boosting theory has been introduced and based on the existing theory a new design methodology based on an analytical solution calculation was presented and demonstrated with the design of a 184GHz amplifier in 28nm FD-SOI.

Finally, Chapter 5 addressed the other key characteristics of a power amplifier that are essential for a performant transmitter integration and to enable data rates over 100Gb/s. These characteristics are the saturation power of the amplifier and the operation bandwidth, which allow a high-distance wireless communication and a high symbol rate operation, respectively. To address these challenges the design of multi-transistor matrixes for sub-THz operations has been presented followed by the meticulous simulation methodology. Large bandwidth matching networks that can produce efficient power transmission at these frequencies have been recalled, as transformers for performant differential operations and coupled transmission line-based power combiner and dividers.

Multiple power amplifiers at 200GHz have been designed in a 28nm FD-SOI process and the measurements have been presented. The analysis of the measurements allowed us to identify perspectives to improve the performance. The intrinsic power capability limit caused by the component's capacitive parts was observed and it allowed us to understand the maximal performances of the used design techniques.

To conclude, the overall design methodology presented in this thesis can be applied to any other types of signal path mmW/sub-THz building blocks.

6.2 Perspectives

6.2.1 Improvement of the designed 200GHz PAs

The limitations of the designed PAs have been identified and an updated and improved version will be implemented to correct the following points:

- Smaller transistor matrixes (~30% smaller) will decrease the constrain of the transformer design, diminish the losses, and prevent a frequency shift
- Coupling capacitors for the positive feedback's transmission lines will be designed using interdigitated top metal layers to enable the EM modeling
- Top EM simulation of the final structure must be used to consider the inter-stage parasitical coupling

Therefore, based on the learnings from the measured first version, better performances can be expected.

6.2.2 Development of new passive devices geometries

This work's main objective was to find the design and technological limits at sub-THz frequencies. The existing impedance-matching structures have been studied and implemented and further research is required to propose new geometries that overcome the phenomena that are limiting the performances. Figure 6.1 presents a transformer geometry that reduces the capacitive coupling between the terminals and that can be used to overcome the stringent design rules in IC processes.

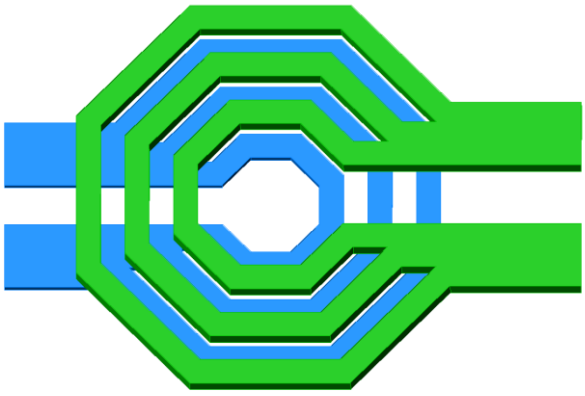


Figure 6.1 Parallel interleaved coupled inductors

6.2.3 Development of new amplifier architectures specific to FD-SOI technologies

In order to leverage the existing design technics issues at very high frequencies, new circuit architectures can be proposed. The positive feedback paths that have been proposed for gain boosting can also be used to bias multiple stages at the same time. The use of unique technological features such as body-biasing in FD-SOI technologies allows new operation strategies. Figure 6.2 presents a multi-stage gain-boosted amplifier that can operate at sub-THz frequencies. All nodes can be considered as shorted from a DC point of view, therefore only one DC feed structure is required. Concurrently, the use of FD-SOI permits to fix independently each transistor's body voltage, hence V_T, hence gain, without interfering on the RF signal path.

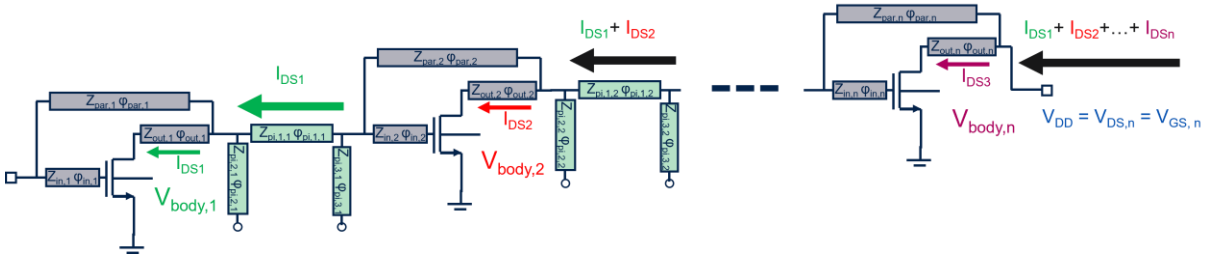


Figure 6.2 Gain-boosted multi-stage amplifier with body bias control

To conclude, the proposed perspectives are of course applicable to any other RF front-end blocks.

6.3 List of publications

This work has led to the following publications:

- S. Sadlo, M. De Matos, A. Cathelin and N. Deltimple, "One stage gain boosted power driver at 184 GHz in 28 nm FD-SOI CMOS," 2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC), 2021.
- N. Deltimple, S. Sadlo, A. Lhomel, F. Rivet, A. Cathelin, "Sub-THz Transceivers with Advanced CMOS technologies", 17th European Microwave Integrated Circuits Conference (EuMIC2022), Milan (Italy), Sept. 27, 2022
- S. Sadlo, A. Cathelin, and N. Deltimple "AMPLIFIER WITH TRANSISTOR VOLTAGE CONTROL", US Patent, filled 2023
- S. Sadlo, A. Cathelin, and N. Deltimple "INTERLEAVED COUPLED INDUCTORS TRANSFORMER", US Patent, filled, 2023
- S. Sadlo, "Development of integrated mmW/THz transmitters in 28FDSOI", ST Innovation iDNA PhD's days, Nov 2021 and Dec 2022

Appendix A Definition of PA metrics

A power amplifier is supposed to amplify the signal and is designed for a certain range of power amplitude. In addition to the previously studied linear gain that is characterized at low power levels where the linear approximation holds, the power amplifier is studied up to its power limits. The schematic of a generic PA is depicted in Figure A.1. The core of the PA is supplied through a constant voltage V_{DD} that supplies a current to the circuit. The signal is generated by a source of impedance Z_s with a signal of power level P_{in}. The frequency of operation is transparent as the discussion can be applied to any operation frequency that the PA is designed for. The PA therefore amplifies the signal and delivers the signal with a power level P_{out} to the load of impedance Z_L. Consequently, the amplifier is designed to operate optimally for these impedances.

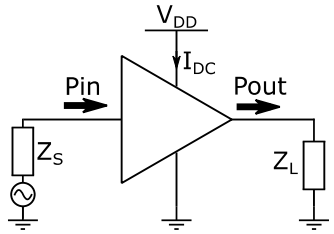


Figure A.1 Generic PA schematic with source and load impedance

An ideal power amplifier may be linear, and its response may be independent of the power levels. However, the characteristics of real amplifiers are non-ideal, and their behavior changes with the signal amplitude. In most use cases, the PA is expected to operate linearly as much as possible. Figure A.2 presents a typical PA response over input power levels. The PA is characterized by its linear gain G, the 1dB input compression point (ICP1), and 1dB output compression point OCP1 which represent the power at which the gain drops by 1dB. It delimits the linear range of operation of the PA, after which the power amplifier compresses further up to the maximal power that it can generate which is referred to as the saturation power P_{sat}. An alternative representation presents the Gain over the P_{in} in Figure A.2 (b) where the compression of the gain is better visible.

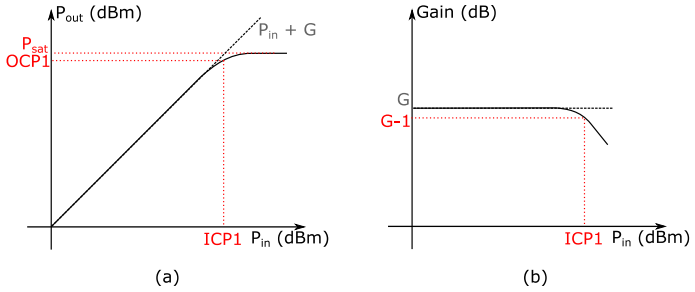


Figure A.2 P_{out} (a) and Gain (b) response of a generic PA as a function of input power P_{in}

As the P_{in} increases, the P_{out} increases linearly, therefore, to provide a higher signal power level, the power consumption of the PA increases with P_{in}. Figure A.3 (a) presents an example of the DC current consumption of a PA as a function of P_{in}. It can be characterized as base consumption, which is the power consumption of the PA from being turned on with no input signal or for low signal power levels. From a certain power level, the PA starts to consume more current from the supply as the P_{in} increases. The efficiency of the PA is commonly given with the power added efficiency (PAE) that is equal to:

$$PAE = 100\frac{P_{out} - P_{in}}{P_{DC}} \quad (A.1)$$

At low power levels, the PAE is close to 0% but as the output power increases, the power consumption increases slower than the P_{out} up to a peak of PAE followed by a decrease due to compression as depicted in Figure A.3 (b). The best efficiency range of operation of the PA is in its non-linear range of operation and close to the compression point. Therefore, a tradeoff must be made in the range of usage of a PA, between the linearity of the response and the efficiency.

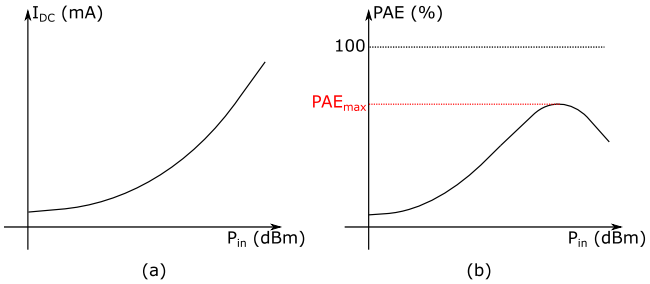


Figure A.3 Current consumption (a) and PAE (b) of a PA as a function of P_{in}

The scheme of a PA can be detailed and decomposed as successive amplification stages as presented in Figure A.4. In the example two successive amplification stages are present. Each of them has an optimal input and output impedance that allows its best functioning conditions to deliver the required output power and offer the required power gain. The source, Load, input, and output impedances are not necessarily compatible, and matching networks are necessary to interconnect the different PA parts. Therefore, three matching networks are necessary for the two-stage PA. The matching networks present a certain power loss that is represented by P_{L1}, $P_{L2,}$ and P_{L3}. Then, each PA stage has the same characteristics as a standalone PA. It has a compression point, a P_{sat}, and a power consumption. The base power consumption is proportional to the maximal power capability of the PA stage.

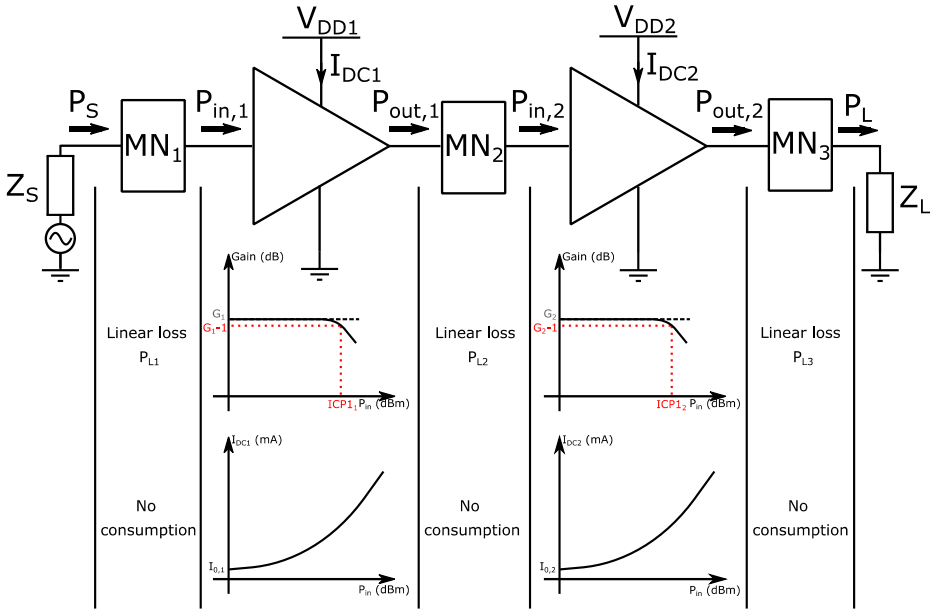


Figure A.4 Two-stage power amplifier schematic with individual component response as a function of power levels

The purpose of the first stage is to “drive” the signal to power levels that allow the second stage to operate through its power range, especially its maximal efficiency power range. Hence, it is referred to as the driver stage. It must be able to generate a P_{out} signal of at least the input compression point of the second stage while operating in the linear region to not alter the overall

linearity of the complete PA. However, the power consumption of the driver is added to the total power consumption and thus degrades the total efficiency. Therefore, the driver is ideally linear up to high power levels and consumes low power, but, both characteristics are exclusive from each other. Consequently, a tradeoff must be made between the driver's power consumption and its power capability. The last stage is referred to as the power stage as it is responsible for the generation of the targeted power level and operates through its power range.

Figure A.5 illustrates this tradeoff with two examples of implementation. Figure A.5 (a) presents the Gain response of a two-stage PA as above with a driver that compresses after the power stage. However, the total power consumption is significantly increased and thus reduces total efficiency. Figure A.5 (b) presents the complementary case where the driver can generate lower power levels but has a lower power consumption. Therefore, the total power consumption is lower, and the total efficiency is potentially better. However, the driver compresses before the driver stage, which degrades the total linearity and can even prevent the power stage from operating at its maximal output power if the driver saturates too soon.

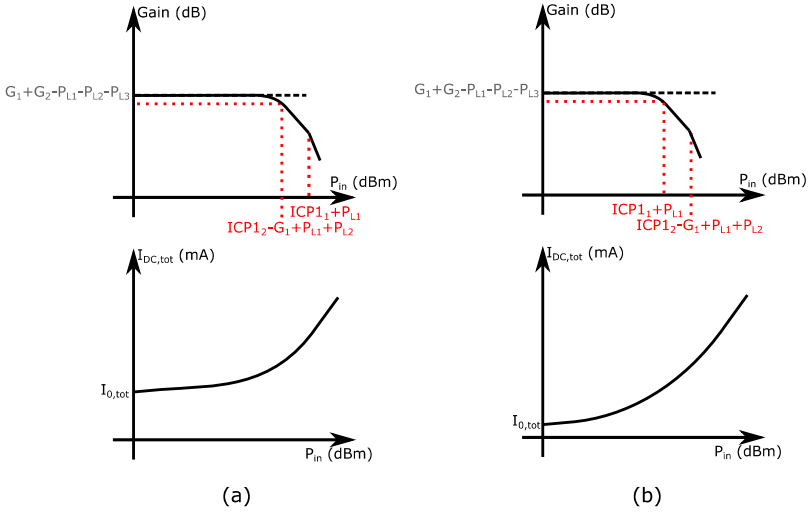


Figure A.5 Total gain response and current consumption of two stages PA for a high P_{sat} first stage (a) and a moderate P_{sat} second stage (b)

The constraint on the compression point of the driver stage can be calculated from the power stage parameters, the required power gain of the driver, and the matching network losses. The difference between the two input compression points must be positive and can be defined as:

$$ICP1_1 - ICP1_2 + G_1 - P_{L2} > 0 \quad (A.2)$$

It can be seen from the output compression point side:

$$OCP1_1 - OCP1_2 + G_2 - P_{L2} > 0 \quad (A.3)$$

Therefore,

$$OCP1_1 > OCP1_2 - G_2 + P_{L2} \quad (A.4)$$

Equation (A.4) allows to size of the driver and shows that the higher the gain of the power stage the smaller must be the driver stage. The loss of the matching network also adds the higher the loss the larger the driver must be. As the frequency increases the power gain of amplifiers and the passive elements efficiency decrease simultaneously, therefore the total efficiency of PA mechanically decreases. If the power gain of two stages is not sufficient, the use of a third, a fourth, or more stages is possible, and the sizing of each additional stage can be made recursively using the previous discussion.

Appendix B PA over 100GHZ State of the Art

1st Autor	Ref name	Paper name	Technology	Node (nm)	Fondeur	Frequence (GHz)	BW (GHz)	Relative BW (%)	Psat (dBm)	PAE max	OP1 (dBm)	Gain (dB)	Stages	Gain/stage (dB)	Stage topology	Boost	Differential	DC (mW)	Area (mm^2)
D-W Park	[Park18]	A 28?/-300-GHz Three-Stage Amplifiers in 65-nm CMOS With 12-/9-dB Gain and 1.6/1.4% PAE While Dissipating 17.9mW	CMOS	65	TSMC	300	10	3.3	-5.3	1.4	-6.4	8.95	3	2.98	CS	Yes	No	17.85	0.056
K Tolgoz	[Tolg19]	A 273–301-GHz Amplifier With 21-dB Peak Gain in 65-nm Standard Bulk CMOS	CMOS	65		287	28	9.8	-10		-13	21	16	1.31	CS	Yes	No	35.4	1.12
D-W Park	[Park20]	A 247 and 272 GHz Two-Stage Regenerative Amplifiers in 65 nm CMOS with 18 and 15 dB Gain Based on Double-Gmax Gain Boosting Tech nique	CMOS	65	TSMC	272	4	1.5	-2.36	2.37	-10.2	15	2	7.50	CS	Yes	No	21.5	0.283
D-W Park	[Park19]	A 230–260-GHz wideband and high-gain amplifier in 65-nm CMOS based on dual-peak Gmax-core	CMOS	65		245	29.7	12.1	-4	1	-6	13.9	4	3.48	CS	Yes	No	23.8	0.14
C-L Ko	[Ko13]	A 210-GHz Amplifier in 40-nm Digital CMOS Technology	CMOS	40	TSMC	213.5	13	6.1	-3.2	0.75	-7.2	10.5	9	1.17	CS	No	No	42.3	0.013
H Bameri	[Bame20]	An Embedded 200 GHz Power Amplifier with 9.4 dBm Saturated Power and 19.5 dB Gain in 65 nm CMOS	CMOS	65	TSMC	200	14	7.0	9.4	1.03	6.3	19.5	8	2.44	Cascode	Yes	No	732	0.92
D Simic	[Simi21]	Analysis and Design of Lossy Capacitive Over-Neutralization Technique for Amplifiers Operating Near fMAX	CMOS	28		190	14.4	7.6	1.5	2.6	-2.4	14.3	5	2.86	CS	No	Yes	45	0.31
M Seo	[Seo09]	A 150 GHz Amplifier With 8 dB Gain and +6 dBm Psat in Digital 65 nm CMOS Using Dummy-Prefilled Microstrip Lines	CMOS	65		150	27	18.0	6.3	9.5	1.5	8.2	3	2.73	CS	No	No	25.5	0.16
H-S Hsiao	[Hsia13]	Millimeter-Wave CMOS Power Amplifiers With High Output Power and Wideband Performances	CMOS	65		140	30	21.4	13.2	14.6	9.9	15	4	3.75	CS	No	No		0.38
Simic	[Simi18]	A 14.8dBm 20.3 dB power amplifier for D-band applications in 40nm CMOS	CMOS	40		140	17	12.1	14.8	8.9	10.7	20.3	3	6.77	CS	No	Yes	305	
J Zhang	[Zhan22]	A 124-152 GHz >15-dBm Psat 28-nm CMOS PA Using Chebyshev Artificial-Transmission-Line-Based Matching for Wideband Power Splittin g and Combining	CMOS	28		138	28	20.3	16.2	8.6	11.4	22.6	3	7.53	CS	No	Yes		0.33
K Katayama	[Kata13]	133GHz CMOS Power Amplifier with 16dB Gain and +8dBm Saturated Output Power for Multi-Gigabit Communication	CMOS	40		133	13	9.8	8.6	7.4	6.8	16.8	6	2.80	CS	No	No	89.1	0.3
Tang	[Tang20]	Design of D-Band Transformer-Based Gain-Boosting Class-AB Power Amplifiers in Silicon Technologies	CMOS	28		132	22	16.7	8	6.6	5.2	22.5	4	5.63	CS	Yes	Yes		0.027
Son	[Son18]	A D-band CMOS power amplifier for wireless chip-to-chip communications with 22.3dB gain and 12.2dBm P1dB in 65nm CMOS technology	CMOS	65	TSMC	118	17	14.4	14.5	10.2	12	22.5	4	5.63	CS	No	Yes		0.34
Son	[Son16]	A 109GHz CMOS power amplifier with 15.2dBm Psat and 20.3dB Gain in 65nm CMOS technology	CMOS	65	TSMC	109	17	15.6	15.2	10.3	12.5	20.3	4	5.08	CS	No	Yes		0.34
Xu	[Xu11a]	A W-band current combined power amplifier with 14.8 dBm Psat and 9.4% maximum PAE in 65nm CMOS	CMOS	65	TSMC	108	16	14.8	14.8	9.4	11.6	14.1	3	4.70	2 Cascode + 1 CS	No	Yes	200	0.106
Xu	[Xu11b]	A 100-117GHz W-band CMOS power amplifier with on-chip adaptive biasing	CMOS	65	TSMC	108	17	15.7	13.8	9.4	11.2	13.4	2	6.70	Cascode	No	Yes		0.041
Momeni	[Mome11]	A high gain 107GHz amplifier in 130nm CMOS	CMOS	130		107	5.4	5.0	2.3	4.4	-1	12.5	3	4.17	CS	Yes	No	31	0.36
Wang	[Wang14]	A CMOS 210-GHz Fundamental Transceiver With OOK Modulation	CMOS SOI	32		210	14	6.7	4.6	6	2.7	15	4	3.75	CS	No	Yes	40	0.06
S Li	[Li21]	A 130-151 GHz 8-Way Power Amplifier with 16.8-17.5 dBm Psat and 11.7-13.4% PAE Using CMOS 45nm RFSOI	CMOS SOI	45	GF	141	21	14.9	17.5	13.4	14.2	24	4	6.00	CS	No	Yes	410	0.43
S Sadio	[Sadi21]	One stage gain boosted power driver at 184 GHz in 28 nm FD-SOI CMOS	CMOS-FDSOI	28	ST	184	20	10.9	-3.7	4.2	-8	7.6	1	7.60	CS	Yes	No	5.1	0.12
D Parveg	[Parv16a]	Demonstration of a 0.325-THz CMOS Amplifier	FD-SOI	28	ST	325	15	4.6				4.5	4	1.13	CS	Yes	No	28	0.242
D Parveg	[Parv16b]	A 124-184 GHz Amplifier Using Slow-wave Transmission lines in 28-nm FDSOI CMOS Process	FD-SOI	28	ST	154	61	39.6				10.1	4	2.53	CS	No	No	31	0.295
E Rahimi	[Rahi22]	A 22nm FD-SOI CMOS 2-way D-band Power Amplifier Achieving PAE of 7.7% at 9.6dBm OP1dB and 3.1% at 6dB Back-off by Leveraging Ada ptive Back-Gate Bias Technique	FD-SOI	22		135	20	14.8	10.3	7.9	9.6	14.2	4	3.55	CS	No	Yes		
J S Chien	[Chien22]	A 111-149-GHz, Compact Power-combined Amplifier With 17.5-dBm Psat, 16.5% PAE in 22-nm CMOS FD-SOI	FD-SOI	22		130	38	29.2	17.5	16.5		13.5	3	4.50	CS	No	Yes		
B Philippe	[Phil20]	A 156dBm 12.8%-PAE Compact D-Band Power Amplifier with Two-Way Power Combining in 16nm FinFET CMOS	FinFET CMOS	16		135	22	16.3	15	12.8	9.2	20.5	3	6.83	CS	No	Yes	207	0.041
Q Yu	[Yu22]	An F-Band Power Amplifier with Skip-Layer Via Achieving 23.8% PAE in FinFET Technology	FinFET CMOS	16	Intel	110	20	18.2	11.8	23.8	11.2	17.1	2	8.55	CS	No	Yes	58.9	0.023
D Fritsche	[Frit14]	A Broadband 200 GHz Amplifier with 17 dB Gain and 18mW DC-Power Consumption in 0.13µm SiGe BiCMOS	SiGe	130	IHP	200	44	22.0	-2.9		-4.1	16.9	2	8.45	Cascode	No	No	18	0.24
X Li	[Li22]	A 110-to-130GHz SiGe BiCMOS Doherty Power Amplifier with Slotline-Based Power-Combining Technique Achieving >22dBm Saturated Out put Power and >10% Power Back-Off Efficiency	SiGe	130		121	28	23.1	22.7	18.7		21.8	2	10.90	Cascode	No	No		1.11
A Visweswaran	[Visw19]	A 112-142GHz Power Amplifier with Regenerative Reactive Feedback achieving 17dBm peak Psat at 13% PAE	SiGe	130	Infineon	130	30	23.1	17	13	14.6	34	5	6.80	CE	No	Yes		1.06
Kucharski	[Kuch17]	A 109-137 GHz Power Amplifier in SiGe BiCMOS with 16.5 dBm Output Power and 12.8% PAE	SiGe	130	IHP	120	28	23.3	16.5	12.8		26.5	3	8.83	Cascode	No	Yes	380	0.74
X Li	[Li22]	A High-Efficiency 142–182-GHz SiGe BiCMOS Power Amplifier With Broadband Slotline-Based Power Combining Technique	SiGe	130	IHP	161	40	24.8	18.1	12.4		30.7	3	10.23	Cascode	No	Yes	405	0.42
X Li	[Li20]	A 160 GHz High Output Power and High Efficiency Power Amplifier in a 130-nm SiGe BiCMOS Technology	SiGe	130	IHP	160	20	12.5	18	9.4		24	3	8.00	Cascode	No	Yes		0.84
Lin	[Lin14]	A 110-134-GHz SiGe Amplifier with peak output power of 100-120mW	SiGe	90	IBM	122	24	19.7	20.8	7.6	1.7	15	4	3.75	Cascode	No	No	1520	4.95
Hou	[Houd22]	A D-band cascode amplifier with 24.3dB gain and 7.7 dBm output power in 0.13µm SiGe BiCMOS technology	SiGe	130	IHP	130	20	15.4	7.7	6.8		24.3	3	8.10	Cascode	No	Yes	84	0.3
Bao	[Bao17]	A 100-145GHz area-efficient power amplifier in a 130nm SiGe technology	SiGe	130	Infineon	120	35	29.2	15.5	6.4		19	6	3.17	CE	No	No	480	0.53
Furgan	[Furq17]	A 15.5dBm 160GHz high-gain power amplifier in SiGe BiCMOS technology	SiGe	130	IHP	160	50	31.3	14	5.7		27	5	5.40	Cascode	No	Yes	440	0.48
A Ali	[Ali20]	168-195 GHz Power Amplifier With Output Power Larger Than 18 dBm in BiCMOS Technology	SiGe	130	IHP	182	27	14.8	18.7	4.4		23.6	3	7.87	Cascode	No	No	431	1.35
Yishay	[Yish15]	A 17.8dBm 110-130GHz power amplifier and doubler chain in SiGe BiCMOS technology	SiGe	120	IBM	120	20	16.7	17.8	4.3	13.5	32	3	10.67	Cascode	No	Yes	560	
Kucharsky	[Kuch19]	An 18 dBm 155-180 GHz SiGe Power Amplifier Using a 4-Way T-Junction Combining Network	SiGe	130	IHP	170	25	14.7	18	4	15.6	30.2	4	7.55	Cascode	No	Yes		0.85
Yishay	[Yish14]	A 17.5dBm D-band power amplifier and doubler chain in SiGe BiCMOS technology	SiGe	120	IBM	120	16	13.3	17.5	3.6	13.3	25.5	3	8.50	Cascode	No	Yes	560	
Daneshgar	[Dane15]	A 22dBm 0.6mm2 D-band SiGe HBT power amplifier using series power combining sub-quarter-wavelength baluns	SiGe	90	GF	120	35	29.2	22	3.6		7.7	2	3.85	Cascode	No	No	2220	0.624
P Stärke	[Stär20]	High-Linearity 19-dB Power Amplifier for 140-220 GHz, Saturated at 15 dBm, in 130-nm SiGe	SiGe	130	IHP	180	80	44.4	15	3.5	13	19	2	9.50	Cascode	No	Yes	830	0.92
Eissa	[Eiss22]	220–320-GHz J-Band 4-Way Power Amplifier in Advanced 130-nm BiCMOS Technology	SiGe	130	IHP	270	100	37.0	13.5	3.15	9.5	15	3	5.00	Cascode	No	Yes		
H Khatibi	[Khat18]	A 173 GHz Amplifier With a 18.5 dB Power Gain in a 130 nm SiGe Process: A Systematic Design of High-Gain Amplifiers Above fmax/2	SiGe	130	ST	173	8.2	4.7	0.9	3		18.5	3	6.17	CE	Yes	No	42	0.4
Eissa	[Eiss19]	A 13.5dBm Fully Integrated 200-to-255GHz Power Amplifier with a 4-Way Power Combiner in SiGe:C BiCMOS	SiGe	130	IHP	230	55	23.9	12	2.14	9	12.5	3	4.17	Cascode	No	Yes	740	2.15
M Sayginer	[Sayg23]	A 110-170 GHz Phase Invariant Variable Gain Power Amplifier Module with 20-22 dBm Psat and 30 dBm OIP3 Utilizing SiGe HBT RFICs	SiGe	130	IHP	140	60	42.9	22	2		14.5	3	4.83	Cascode	No	No	9000	
H Khatibi	[Khat18]	A 183 GHz Desensitized Unbalanced Cascode Amplifier With 9.5-dB Power Gain and 10-GHz Band Width and ~2 dBm Saturation Power	SiGe	130	ST	183	8.5	4.6	-2.8	1		9.5	1	9.50	Cascode	Yes	No	30	0.152
T Bücher	[Büch21]	A 239-298 GHz Power Amplifier in an Advanced 130 nm SiGe BiCMOS Technology for Communications Applications	SiGe	130	IHP	269	59	21.9	6.7	0.92	5.2	17.9	3	5.97	Cascode	No	Yes	417	0.26
S. Ghosh	[Ghos23]	305 GHz Cascode Power Amplifier Using Capacitive Feedback Fabricated Using SiGe HBT's with f max of 450 GHz	SiGe	130	IHP	305	13	4.3	8.3	0.7	6	14.5	4	3.63	Cascode	No	Yes	880	0.49
Samrah	[Sarm16]	A 200-225GHz SiGe power amplifier with peak Psat of 9.6dBm using wideband power combination	SiGe	130	Infineon	215	25	11.6	9.6	0.5	4	25	3	8.33	Cascode	No	Yes		
H Li	[Li22]	A 250-GHz Differential SiGe Amplifier With 21.5-dB Gain for Sub-THz Transmitters	SiGe	130	IHP	252	11	4.4	0	0.3	-3.7	21.5	3	7.17	Cascode	No	Yes	149	0.17

Appendix C Network analysis

The analysis of electrical systems consists of finding relations between currents and voltages at different points of the system. They can then be used to estimate the transfer functions from one point to another. Circuits can be composed of discrete components or matrixes that represent complex networks.

Matrix representation

When designing complex circuits, a level of abstraction for electrical components is useful to define the constraints for each circuit sub-part and to keep only information on the external nodes. The characteristics of the circuit must be summarized without a loss of information at the ports of interest.

A black box can be used to represent a circuit and a matrix of coefficients can represent the relation between the currents and voltages.

Multiple concepts such as gains, maximal gains, impedances, optimal impedances, and stability criteria can be generalized, and theorems can be demonstrated on a general matrix.

This method is commonly used for linear networks that represent a linear system or linear approximation of a non-linear system. The following considerations are about linear networks.

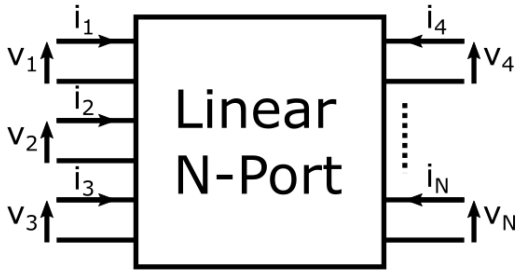


Figure C.1 Schematic of a linear N-port network

Figure C.1 is a representation of a linear N-Port network, its representative matrix can take different equivalent forms, and transformations exist from one to another.

One common representation matrix is the impedance matrix Z.

$$V = Z \times I = \begin{pmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \\ \vdots \\ V_N \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} & Z_{13} & Z_{14} & \dots & Z_{1\mathrm{N}} \\ Z_{21} & Z_{22} & Z_{23} & Z_{24} & \cdots & Z_{2\mathrm{N}} \\ Z_{31} & Z_{32} & Z_{33} & Z_{34} & \cdots & Z_{3\mathrm{N}} \\ Z_{41} & Z_{42} & Z_{43} & Z_{44} & \cdots & Z_{4\mathrm{N}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_{\mathrm{N}1} & Z_{\mathrm{N}2} & Z_{\mathrm{N}3} & Z_{\mathrm{N}4} & \cdots & Z_{\mathrm{NN}} \end{pmatrix} \begin{pmatrix} I_1 \\ I_2 \\ I_3 \\ I_4 \\ \vdots \\ I_N \end{pmatrix} \quad (\mathrm{C}.1)$$

Impedance matrix Z allows us to calculate the impedances from the currents on the ports. Naturally, the inverse matrix $\mathrm{Y} = Z^{-1}$ gives the currents from the voltages. The transformation is mathematically simple.

The matrix definition implies that each coefficient is defined as:

$$Z_{ij} = \left.\frac{V_i}{I_j}\right|_{\forall k \neq j, I_k = 0} \quad (\mathrm{C}.2)$$

And

$$Y_{ij} = \left.\frac{I_i}{V_j}\right|_{\forall k \neq j, I_k = 0} \quad (\mathrm{C}.3)$$

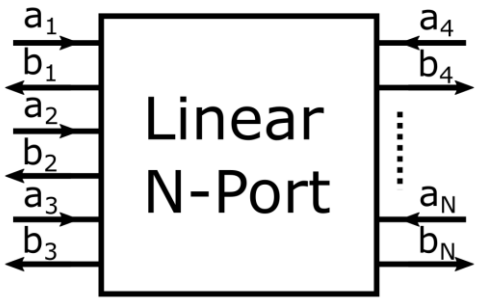


Figure C.2 Linear N-port network with the incident and reflected power waves.

Another common matrix representation of a component is the scattering parameters matrix. It describes the electrical behavior of linear electrical networks when ports are terminated by known impedances (Usually all ports are terminated with the same impedance noted Z_0). It differs from the previous representations, in the sense that the parameter S_{ij} gives the power transfer function between a source port *i* and load terminal port *j*. When *i=j* it represents the reflected power from the port *i*. This representation is convenient for measurement equipment, especially vector network analyzers that measure the transferred and reflected signal voltage magnitudes on the ports of a tested component for given known port impedances.

Once measured, these S parameters are transformed into any other matrix representation to treat the measurement results.

The S parameters are defined as the relations between a_i and b_i which represent respectively the incident and reflected power waves on each port.

They are defined as:

$$a_i = \frac{V_i + Z_i I_i}{2\sqrt{|real(Z_i)|}} \tag{C.4}$$

$$b_i = \frac{V_i - Z_i^* I_i}{2\sqrt{|real(Z_i)|}} \tag{C.5}$$

Where V_i and I_i are respectively the voltage and current at the i^{th} port and Z_i is the impedance looking out of the port (i.e., the impedance of the source connected to it).

To have some insight into the physical meaning of these power waves, we can apply them to the circuit in Figure 1.1-1.

$$V_L = V_S - Z_S I_S \tag{C.6}$$

$$a = \frac{V_L + Z_S I_L}{2\sqrt{|real(Z_S)|}} = \frac{V_S}{2\sqrt{|real(Z_S)|}} \tag{C.7}$$

Then,

$$|a|^2 = \frac{|V_S|^2}{4R_S} \tag{C.8}$$

Which is the maximal power emitted from the source.

The a_i power waves are the maximal power emitted by the sources independently from the impedance of the port.

$$|a|^2 - |b|^2 = \frac{|V_L + Z_S I_S|^2 - |V_L - Z_S^* I_S|^2}{4R_S}$$

$$= \frac{V_L^* I_S + V_L I_S^*}{2} = real(V_L I_S^*) = P_L \tag{C.9}$$

$|a_i|^2 - |b_i|^2$ is the power received by the load and it means that $|b_i|^2$ is the part of the power emitted by the source that is not received by the load, in other words, the reflected power.
Also, in the term $|a_i|^2 - |b_i|^2$ if the source *I* don't emit power, it is the power flowing out of the *i*th port.
Hence, the S parameter coefficients are defined as follows:

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{\forall k \neq j, a_k = 0} \quad (C.10)$$

Where $\left|S_{ij}\right|^2$ is the power that flows out the port *i* divided by the total power emitted by the source at port *j* when all other sources are turned off.
When $i \neq j$, it represents the power gain from the source at port *j* to the load at port *i*.
When $i = j$, it represents the ratio between the reflected power at one port and the total emitted power at the same port.

$$S_{ii} = \frac{V_i - Z_i^* I_i}{V_i + Z_i I_i} = \frac{Z_{L,i} - Z_i^*}{Z_{L,i} + Z_i} \quad (C.11)$$

Which is known as the reflection coefficient Γ.
S parameter coefficients and the power waves can be related in a matrix form:

$$b = Sa \quad (C.12)$$

Now that S parameters are defined and their matrix expression is known, it is practical to express it from *Z* or *Y* matrix and inversely.
From the definition of a_i and b_i we can write the matrix relation:

$$a = F(V + GI) \quad (C.13)$$

And

$$b = F(V - G^+ I) \quad (C.14)$$

With *F* and *G,* the diagonal matrices whose *i*th diagonal components are respectively $1/2\sqrt{|real(Z_i)|}$ and Z_i.
$^+$ indicates the complex conjugate transposed matrix.
Using the impedance matrix *Z* relating the voltages and currents of the network ports and from the previous relations we can write,

$$F(Z - G^+)I = SF(Z + G)I \quad (C.15)$$

Then by dividing on both sides,

$$S = F(Z - G^+)(Z + G)^{-1}F^{-1} \quad (C.16)$$

Finally, to obtain the equation giving Z we must rewrite the previous equation,

$$(F - SF)Z = (G^+ + SG)F \quad (C.17)$$

Then by dividing again,

$$Z = F^{-1}(1 - S)^{-1}(G^+ + SG)F \quad (C.18)$$

These relations allow to change from the representation of the linear network alone with its *Z* Impedance or *Y* Admittance matrix to the S-Parameters matrix which considers the sources and loads at its ports with their respective impedances.
S-Parameters represent *in fine* the measurement results when external sources and loads are connected to the device under test (DUT). Signals are sent from each port source and the output signals of all ports are measured, measuring the incident way the incident, and reflected power waves and thus calculating S-Parameters coefficients.

The *Z* and *Y* representations are often used for circuit analysis and designs based on RLCM representations.

For complex circuits, S-Parameters representation is, therefore, more convenient and is commonly used to represent and assemble complex blocks.

Usually, for circuits with more than 2 or 3 ports, S-parameters representation is used.

Finally, an

Two-Port Network

Definition

A Two-Port Network is a matrix representation of a linear two-terminal component. It is the most used Linear Component representation as it can be used for one input and one output terminal.

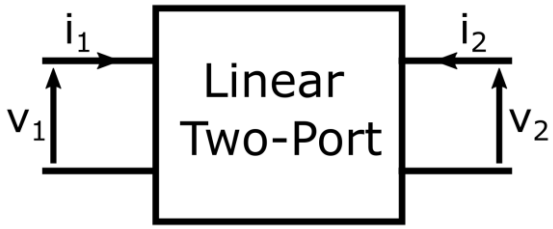


Figure C.3 Schematic of a linear two-port

The previous definition applies:

$$\begin{cases} V_1 = Z_{11}I_1 + Z_{12}I_2 \\ V_2 = Z_{21}I_1 + Z_{22}I_2 \end{cases} v \tag{C.19}$$

We can define impedances at each port when all other ports are ended by a load impedance.

$$Z_1 = \frac{V_1}{I_1} = Z_{11} - \frac{Z_{21}}{Z_{22} + Z_{Load,2}} Z_{12} \tag{C.20}$$

Inversely,

$$Z_2 = \frac{V_2}{I_2} = Z_{22} - \frac{Z_{21}}{Z_{11} + Z_{Load,1}} Z_{12} \tag{C.21}$$

For an admittance representation, the port admittances can be expressed as,

$$Y_1 = \frac{I_1}{V_1} = \frac{1}{Z_1} = Y_{11} - \frac{Y_{21}}{Y_{22} + Y_{Load,2}} Y_{12} \tag{C.22}$$

$$Y_2 = \frac{I_2}{V_2} = \frac{1}{Z_2} = Y_{22} - \frac{Y_{21}}{Y_{11} + Y_{Load,1}} Y_{12} \tag{C.23}$$

We observe that the impedances depend on the loads presented on the opposite port. The effect of the opposite port impedance influences the impedance proportional to the magnitude of the product $Z_{12}Z_{21}$. If at least one of the two components is equal to *0*, there is complete independence of the port impedances to the loads. This is the case for a unilateral two-port.

Alternative representations

Alternative representations of the LTP exist and are preferred depending on the required analysis. While *Z* and *Y* matrixes are either homogeneous to ohms or siemens, Cascade parameters (also named ABCD, chain, or transmission parameters), Hybrid and Inverse Hybrid parameters still relate to the voltages and currents of the LTP, but their parameters have different physical units. They are defined as follows,

Cascade Parameters

Cascade Parameters relate the voltage and current from one port to the one from the opposite port.

$$\begin{cases} V_1 = AV_2 - BI_2 \\ I_1 = CV_2 - DI_2 \end{cases} or \begin{pmatrix} V_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} V_2 \\ I_2 \end{pmatrix} \quad (C.24)$$

Thus,

$$A = \frac{V_1}{V_2}\Big|_{I_2=0} \quad B = -\frac{V_1}{I_2}\Big|_{V_2=0} \quad (C.25)$$

$$C = \frac{I_1}{V_2}\Big|_{I_2=0} \quad D = -\frac{I_1}{I_2}\Big|_{V_2=0} \quad (C.26)$$

Hybrid Parameters

Hybrid Parameters c

$$\begin{cases} V_1 = H_{11}I_1 + H_{12}V_2 \\ I_2 = H_{21}I_1 + H_{22}V_2 \end{cases} or \begin{pmatrix} V_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{21} \end{pmatrix} \begin{pmatrix} I_1 \\ V_2 \end{pmatrix} \quad (C.27)$$

Inverse Hybrid Parameters

Inverse Hybrid Parameters

$$\begin{cases} I_1 = G_{11}V_1 + G_{12}I_2 \\ V_2 = G_{21}V_1 + G_{22}I_2 \end{cases} or \begin{pmatrix} I_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{21} \end{pmatrix} \begin{pmatrix} V_1 \\ I_2 \end{pmatrix} \quad (C.28)$$

It is clear that [G] is the inverse matrix of [H].

Matrix transformations

Transformation formulas between the different matrix's representations can be calculated, their expressions are:

	[Z]	[Y]	[ABCD]	[H]
[Z]	$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$	$\frac{1}{\Delta Y}\begin{bmatrix} Y_{22} & -Y_{12} \\ -Y_{21} & Y_{11} \end{bmatrix}$	$\frac{1}{C}\begin{bmatrix} A & \Delta ABCD \\ 1 & D \end{bmatrix}$	$\frac{1}{H_{22}}\begin{bmatrix} \Delta H & H_{12} \\ -H_{21} & 1 \end{bmatrix}$
[Y]	$\frac{1}{\Delta Z}\begin{bmatrix} Z_{22} & -Z_{12} \\ -Z_{21} & Z_{11} \end{bmatrix}$	$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{21} \end{bmatrix}$	$\frac{1}{B}\begin{bmatrix} D & -\Delta ABCD \\ -1 & A \end{bmatrix}$	$\frac{1}{H_{11}}\begin{bmatrix} 1 & -H_{12} \\ H_{21} & \Delta H \end{bmatrix}$
[ABCD]	$\frac{1}{Z_{21}}\begin{bmatrix} Z_{11} & \Delta Z \\ 1 & Z_{22} \end{bmatrix}$	$\frac{1}{Y_{21}}\begin{bmatrix} -Y_{22} & -1 \\ -\Delta Y & -Y_{11} \end{bmatrix}$	$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$	$\frac{1}{H_{21}}\begin{bmatrix} -\Delta H & -H_{11} \\ -H_{22} & -1 \end{bmatrix}$
[H]	$\frac{1}{Z_{22}}\begin{bmatrix} \Delta Z & Z_{12} \\ -Z_{21} & 1 \end{bmatrix}$	$\frac{1}{Y_{11}}\begin{bmatrix} 1 & -Y_{12} \\ Y_{21} & \Delta Y \end{bmatrix}$	$\frac{1}{D}\begin{bmatrix} B & \Delta ABCD \\ -1 & C \end{bmatrix}$	$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{21} \end{bmatrix}$

Figure C.4 Conversion matrixes

The formulas are convenient to switch from one representation to the other when the calculation of multiple interconnected LTPs equivalent matrix is required.

Combination of LTPs

A combination of LTPs is used to calculate complex structures composed of elementary LTPs. The following cases illustrate the convenience of a specific type of representation compared to others when combining multiple LTPs.

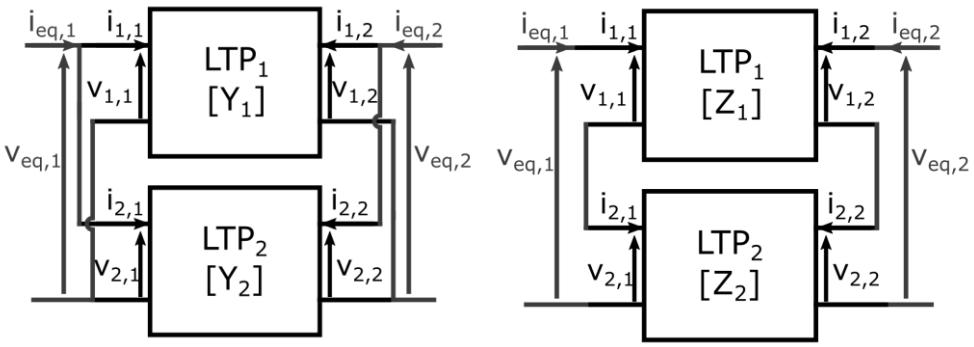


Figure C.5 Combination of parallel (left) and series (right) LTPs

Figure C.5 (left) depicts two LTPs connected in parallel. Using Admittance matrixes, the equivalent LTP is:

$$Y_{eq} = Y_1 + Y_2 \tag{C.29}$$

Similarly for series LTPs the convenient matrix form is the impedance matrix. Hence, the equivalent LTP parameters can be calculated with:

$$Z_{eq} = Z_1 + Z_2 \tag{C.30}$$

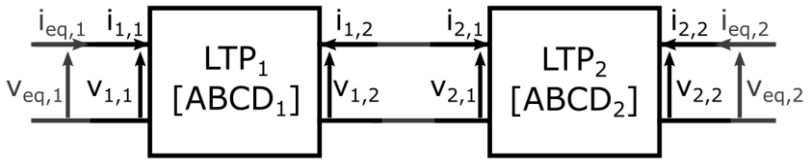


Figure C.6 Combination of cascaded LTPs

For two cascaded successive matrixes as presented in Figure C.6, the equivalent LTP can be expressed using the cascade parameters matrix. The matrix relates the current and voltage of one port to the other one. Hence the relation of the input of the first LTP to the output of the second is the product of their cascade matrixes:

$$ABCD_{eq} = ABCD_1 \times ABCD_2 \tag{C.31}$$

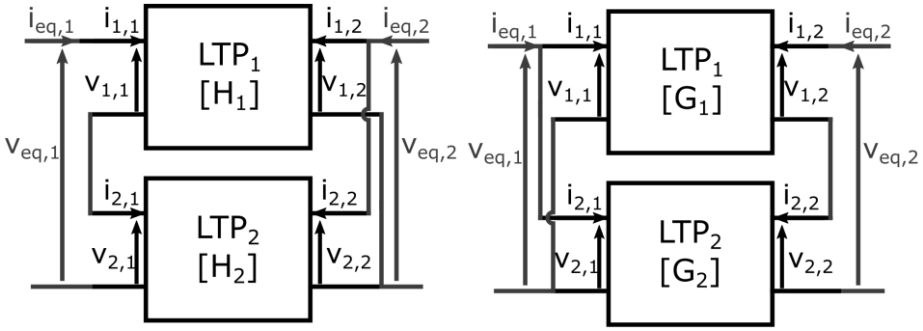


Figure C.7 Combination of hybrid parallel-series (left) and series-parallel (right) LTPs

For Hybrid configurations as presented in Figure C.7, series-parallel and parallel-series are adapted respectively to hybrid and inverse hybrid forms:

$$H_{eq} = H_1 + H_2 \tag{C.32}$$

$$\mathrm{G_{eq}} = G_1 + \mathrm{G_2} \tag{C.33}$$

Maximum power transfer theorem

The output of a circuit or any source of signal can be represented with a voltage source and an internal impedance that limits the current that the circuit can provide. The circuit is connected to

a load that receives the signal with a certain power amplitude. This configuration is depicted in Figure C.8.

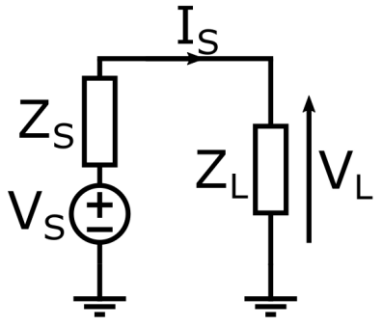


Figure C.8 Schematic of a voltage source V_S with an internal impedance Z_L connected to a load of impedance Z_L.

The voltage across the load and the current that flows into it is a function of the source and load impedances. Hence, the power that is generated by the source depends on the load impedance. The optimal load impedance that maximizes the power output of the source is the optimal load $Z_{L,opt}$.

The expression of the power received by the load is:

$$P_L = real(V_L I_L^*) \tag{C.34}$$

Then, the voltage and current are:

$$V_L = Z_L I_L \tag{C.35}$$

$$I_L = I_S = \frac{V_S}{Z_L + Z_S} \tag{C.36}$$

The power received is then expressed in terms of the circuit's parameters:

$$P_L = real\left(\frac{Z_L V_S}{Z_L + Z_S}\frac{{V_S}^*}{(Z_L + Z_S)^*}\right) = \frac{1}{2}|V_S|^2 \frac{real(Z_L)}{|Z_L + Z_S|^2} \tag{C.37}$$

If we separate the real and imaginary parts of the impedances as:

$$Z_S = R_S + jX_S \tag{C.38}$$

$$Z_L = R_L + jX_L \tag{C.39}$$

$$P_L = \frac{|V_S|^2 R_L}{(R_S + R_L)^2 + (X_S + X_L)^2} \tag{C.40}$$

The power is maximal when the imaginary part of the load impedance X_L is equal to the negative value of the imaginary part of the source X_S.

Thus, the expression becomes:

$$P_L = \frac{|V_S|^2}{\frac{R_S^2}{R_L} + R_L + 2R_S} \tag{C.41}$$

The function $\mathrm{f}(x) = \frac{a^2}{x} + \mathrm{x} + 2\mathrm{a}$ has the derivative $f'(x) = \frac{a^2}{-x^2} + 1$ which is zero for $\mathrm{x} = \pm a$.

The second derivative is equal to $f''(x) = 2\frac{a^2}{x^3}$ which is positive for positive values of x.

Meaning that *f* is minimal for $\mathrm{x} = \mathrm{a}$.

Finally, we can deduce that P_L is maximal for R_S=R_L,

$$Z_{L,opt} = R_S - jX_S = Z_S^* \tag{C.42}$$

$$P_{L,max} = \frac{|V_S|^2}{4R_S} \tag{C.43}$$

In summary, we demonstrated the maximum power transfer theorem which states that the transmitted power to a load is maximal when the load impedance is equal to the conjugate of the source impedance.
Therefore, in the design of a circuit, to maximize its performance, the impedances presented to the ports have to be matched to the port's impedances.

Gain definitions

We want to calculate the maximal power gain that can generate a Linear Two-Port under optimal conditions. Thus, the research of these optimal conditions is a prerequisite to obtaining maximal performances from a circuit.

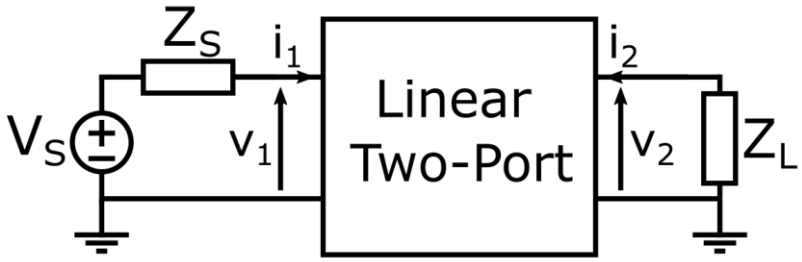


Figure C.9 Schematic of a linear two port with one port connected to a voltage source and the other connected to a load

From a representation of an LTP by a black box matrix, and knowing the source and load impedances, respectively Z_S and Z_L, it is possible to calculate the power gain of the circuit.
Multiple powers can be defined, either the maximal power that a source can generate, when the port's impedance is matched, or the received power by a load:
-the available power flowing out of the source,

$$P_{av,S} = \frac{|V_S|^2}{4ReZ_S} = \frac{|V_S|^2}{4} Re(Y_S) \quad (C.44)$$

-the power entering the LTP,

$$P_{in} = \frac{|V_1|^2}{4ReZ_1} = \frac{|V_1|^2}{4} Re(Y_1) \quad (C.45)$$

-the available power available for the load, flowing out of the LTP,

$$P_{av,L} = \frac{|V_2|^2}{4ReZ_2} = \frac{|V_2|^2}{4} Re(Y_2) \quad (C.46)$$

-the power received by the load,

$$P_L = \frac{|V_2|^2}{4ReZ_L} = \frac{|V_2|^2}{4} Re(Y_L) \quad (C.47)$$

Given the distinctions between the defined powers, multiple power gains can be defined as well:
-the power gain G_p, which is the gain between the power received by the LTP and the power received by the load,

$$G_p = \frac{P_L}{P_{in}} = f(Z_{ij}, Z_L) \quad (C.48)$$

Thus, it depends on the LTP parameters and the load impedance but is independent of the source impedance Z_S.
-the available gain G_a of the LTP, which is the power gain between the available power at the source $P_{av,s}$ and the output of the LTP $P_{av,L}$.

$$G_a = \frac{P_{av,L}}{P_{av,L}} = f(Z_{ij}, Z_S) \tag{C.49}$$

This time the gain depends on Z_S and is independent of Z_L.

-Finally, the transconductance gain G_t which considers the power received by the load P_L and the power available at the source $P_{av,S}$

$$G_T = \frac{P_L}{P_{av,S}} = f(Z_{ij}, Z_S, Z_L) \tag{C.50}$$

This metric considers both reflections, at the input and output of the LTP.

These gains can be related to the S-parameters of the LTP,

$$G_T = S_{21} \tag{C.51}$$

$$G_p = \frac{S_{21}}{1 - S_{11}} \tag{C.52}$$

$$G_a = \frac{S_{21}}{1 - S_{22}} \tag{C.53}$$

Optimal Matching and Maximum Available Gain

When ports are optimally loaded with conjugate impedances, $Z_S = Z_1^*$ and $Z_L = Z_2^*$, there are no reflections, and the three gains are equal. This maximal gain value under optimal conditions is named the maximally available gain G_{ma}.

$$G_{ma} = G_T(Z_S = Z_1^*, Z_L = Z_2^*) = G_p(Z_L = Z_2^*) = G_a(Z_S = Z_1^*) \tag{C.54}$$

Finally, the gains can be expressed in terms of Y parameters,

$$G_T = \frac{4Re\{Y_L\}Re\{Y_S\}|Y_{21}|^2}{|(Y_S + Y_{11})(Y_{22} + Y_L) - Y_{12}Y_{21}|^2} \tag{C.55}$$

$$G_p = \left|\frac{Y_{21}}{Y_{22} + Y_L}\right|^2 \frac{Re\{Y_L\}}{Re\{Y_1\}} \tag{C.56}$$

$$G_a = \left|\frac{Y_{21}}{Y_{11} + Y_S}\right|^2 \frac{Re\{Y_S\}}{Re\{Y_2\}} \tag{C.57}$$

By deriving the expression of G_p or G_a relative to Y_l and Y_S respectively, optimal $Y_{L,opt,}$ and $Y_{S,opt}$ can be derived in terms of Y parameters and thus can be expressed G_{ma}.

This analysis gives:

$$Y_{L,opt} = \frac{|Y_{21}Y_{12}|\left(K + \sqrt{K^2 - 1}\right) + Y_{21}Y_{12}}{2ReY_{11}} - Y_{22} \tag{C.58}$$

$$Y_{S,opt} = \frac{|Y_{21}Y_{12}|\left(K + \sqrt{K^2 - 1}\right) + Y_{21}Y_{12}}{2ReY_{22}} - Y_{11} \tag{C.59}$$

$$G_{ma} = \left|\frac{Y_{21}}{Y_{12}}\right|\left(K - \sqrt{K^2 - 1}\right) \tag{C.60}$$

With,

$$K = \frac{2Re\{Y_{11}\}Re\{Y_{22}\} - Re\{Y_{21}Y_{12}\}}{|Y_{21}Y_{12}|} \tag{C.61}$$

These expressions are defined for $K \geq 1$. When $K < 1$, the square root gives a complex value while it represents a power gain, thus a real quantity.

When these expressions are defined, they can give us instantly the value of G_{ma} and optimal impedances (or admittances) to present to the two ports, only using its Y parameters representation (or any other representation, given their equivalence).

It is convenient for non-unilateral ATP to instantly know the conjugate impedances to present simultaneously at its ports to match it. Otherwise, it requires an iterative input and then output impedance matching design optimization because of the dependence of the port impedances to loads.

The maximum available gain G_{ma} is a useful metric as it gives the maximal gain that can provide a LTP from its matrix parameters, without requiring to proceed to an impedance matching.

Appendix D Lumped modeling limit

To illustrate the frequency limit of a *π* representation, one calculates the component values from the admittance matrix of a *π* representation. Then apply these formulas to a double-*π* network to determine the frequency until which the two models are equivalent.

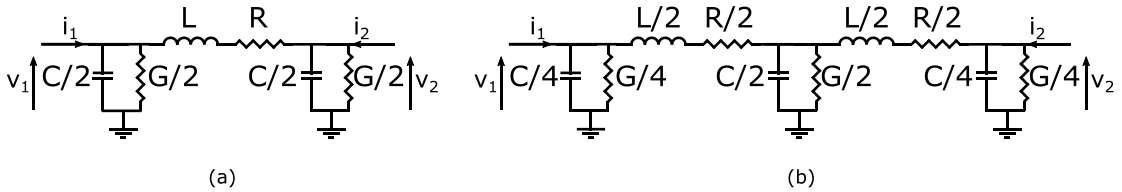


Figure D.1 π-model (a) and double-π-model (b) of an inductor

Figure D.1 (a) presents a π-model of an inductor. The admittance parameters of this two-port are:

$$Y_{\pi,11} = Y_{\pi,22} = \frac{j\omega C + G}{2} + \frac{1}{R + j\omega L} \tag{D.1}$$

$$Y_{\pi,12} = Y_{\pi,21} = -\frac{1}{R + j\omega L} \tag{D.2}$$

From which the lumped model component can be extracted with:

$$L_{extr} = -\frac{Imag\left(\frac{1}{Y_{12}}\right)}{\omega} \tag{D.3}$$

$$R_{extr} = -Real\left(\frac{1}{Y_{12}}\right) \tag{D.4}$$

$$C_{extr} = 2\frac{Imag(Y_{11} + Y_{12})}{\omega} \tag{D.5}$$

$$G_{extr} = 2Real(Y_{11} + Y_{12}) \tag{D.6}$$

Figure D.1 (b) presents a double-π lumped model of the inductor, where the total inductance, resistance, capacitance, and conductance are the same but divided into multiple components. The Y matrix parameters from the double-π model are:

$$Y_{2\pi,11} = Y_{2\pi,22} = (j\omega C + G)\left(\frac{1}{4} + \frac{2}{(j\omega C + G)(R + j\omega L) + 8}\right) + \frac{1}{R + j\omega L}\frac{8}{(j\omega C + G)(R + j\omega L) + 8} \tag{D.7}$$

$$Y_{2\pi,12} = Y_{2\pi,21} = -\frac{1}{R + j\omega L}\frac{8}{(j\omega C + G)(R + j\omega L) + 8} \tag{D.8}$$

Then the extraction formulas (D.3) to (D.6) can be applied to (D.7) and (D.8):

$$L_{extr} \approx L\left(1 - \left(\frac{\omega}{\omega_0}\right)^2\right) \tag{D.9}$$

$$R_{extr} \approx R\left(1 - \left(\frac{\omega}{\omega_0}\right)^2\right) \tag{D.10}$$

$$C_{extr} \approx C\left(\frac{1 - \frac{1}{2}\left(\frac{\omega}{\omega_0}\right)^2}{1 - \left(\frac{\omega}{\omega_0}\right)^2}\right) \tag{D.11}$$

$$G_{extr} \approx G\left(\frac{1 - \frac{1}{2}\left(\frac{\omega}{\omega_0}\right)^2}{1 - \left(\frac{\omega}{\omega_0}\right)^2}\right) \tag{D.12}$$

Where,

$$\omega_0 = \sqrt{\frac{8}{LC}} \tag{D.13}$$

Equations (D.9) to (D.12) confirm that the pi model can be used to represent a 2-*π* structure for frequencies smaller than ω_0. As the frequency increases towards ω_0, the ratio between the extracted parameters and the total values are:

$$\frac{L_{extr}}{L} = \frac{R_{extr}}{R} \approx \left(1 - \left(\frac{\omega}{\omega_0}\right)^2\right) \tag{D.14}$$

$$\frac{C_{extr}}{C} = \frac{G_{extr}}{G} \approx \left(\frac{1 - \frac{1}{2}\left(\frac{\omega}{\omega_0}\right)^2}{1 - \left(\frac{\omega}{\omega_0}\right)^2}\right) \tag{D.15}$$

Equations (D.14) and (D.15) can be evaluated to determine the frequency at which the error is equal to 1%. Figure D.2 presents the evaluation of the equations.

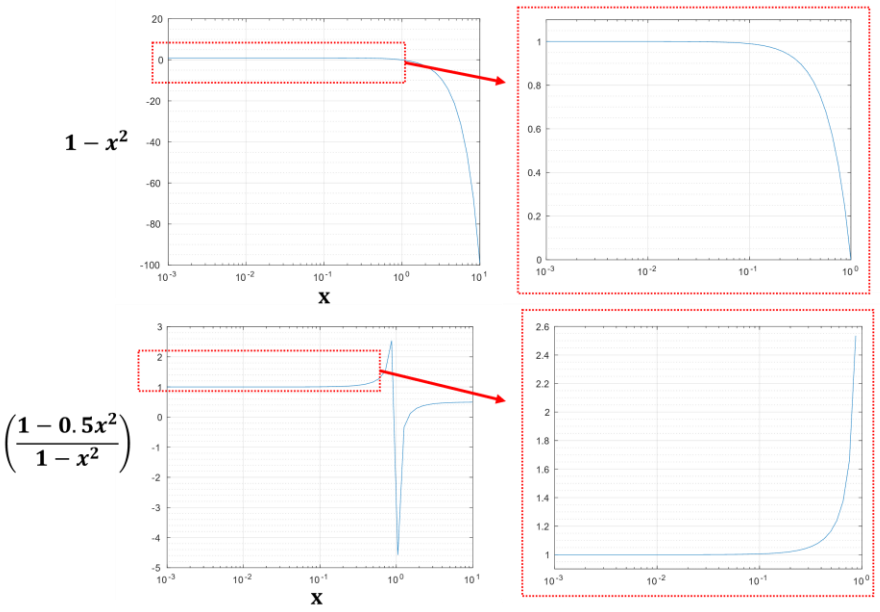


Figure D.2 Evaluation of the ratio between real and extracted component values as a function of normalized frequency

The extracted value of resistance and inductance starts to diverge at one-tenth of ω_0. Figure D.3 resumes the values for different orders of magnitude of the frequency.

x	0.001	0.01	0.1	0.5	0.75
$1 - x^2$	0.999999	0.9999	0.99	0.75	0.4375
$\left(\frac{1 - 0.5x^2}{1 - x^2}\right)$	1	1.0001	1.0051	1.1667	1.6429

Figure D.3 Evaluation of Equations D.14 and D.15 for the different order of magnitudes of frequencies

One can define the frequency $f_{1\%}$ at which the gap between to extracted value and the total value is equal to 1%. From figure D.3 it is equal to $\omega_0/10$.

$$f_{1\%} = \frac{1}{10\pi}\sqrt{\frac{2}{LC}} \approx \frac{1}{22\sqrt{LC}} \tag{D.16}$$

Equation (D.16) indicates that a π model is correct with less than 1% error up to 1/22 of the *LC* resonance frequency.

Consequently, the model of a component must be divided into multiple π networks to ensure that the model is correct at the highest frequency of operation. Equation (D.16) can be used recursively. A 2π model is correct up to $2f_{1\%}$, a 4π model, up to $4f_{1\%}$ etc.
Finally, for a maximal operation frequency $f_{op,max}$, to present less than 1% error, the model must be divided into at least n_{min} π networks, defined as:

$$n_{min} = \left\lfloor \frac{f_{op,max}}{f_{1\%}} \right\rfloor + 1 \qquad (\mathrm{D}.17)$$

The lumped element representation of components like capacitors or inductors is useful when the distributed behavior is not important and when the simple π or double π is sufficient. For higher frequencies, the lumped analysis becomes difficult because of the large number of required networks to capture the high-frequency effects.

Appendix E Optimization and simulation of large transistor cell for sub-THz PA application

Unitary transistor cell

The optimization of the unitary transistor f_{max} is the first step toward a multi-transistor matrix. Multiple parameters can be tuned on the transistor from the design kit, they are presented hereafter. To estimate the performances of a certain transistor topology and compare it to other possibilities, it is necessary to layout the transistor with at least the thin metal layers as the best combination at the first metal level is not necessarily the best once laid out.

Figure E.1 presents the Mason's Gain (a) and extrapolated f_{max} (b) over the frequency of a Pcell that considers interconnections up to the first metal layer, with an L=30nm, 20 fingers of W_{finger}=800nm each, hence, W_{tot}=16µm.

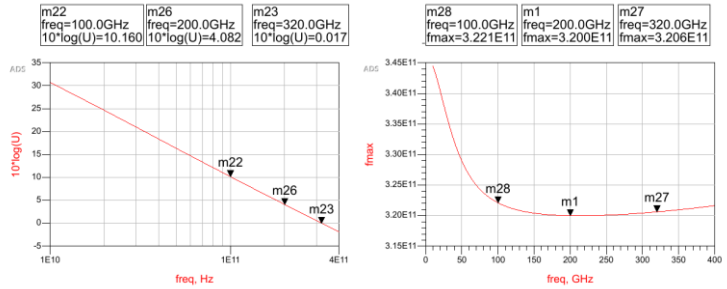


Figure E.1 Simulated Mason's gain (a) and extrapolated f_{max} (b) over frequency for a 28nm FD-SOI DK transistor model.

The model's U decreases at 20dB per decade above 100GHz which gives a constant extrapolated f_{max} of 320GHz as expected from the theory of the previous chapter. It corresponds to a 4.08dB gain at 200GHz. Figure E.2 presents the same simulated metric for the transistor with its 6 first metal layers interconnections. This time the Mason's gain decreases at a faster rate than 20dB per decade which produces a decreasing extracted f_{max} with frequency. At 200GHz the gain is 3.44dB and a 297GHz f_{max} when extracted at 200GHz. The f_{max} decrease is not important but because of the high frequency of operation, any decrease in gain is undesired.

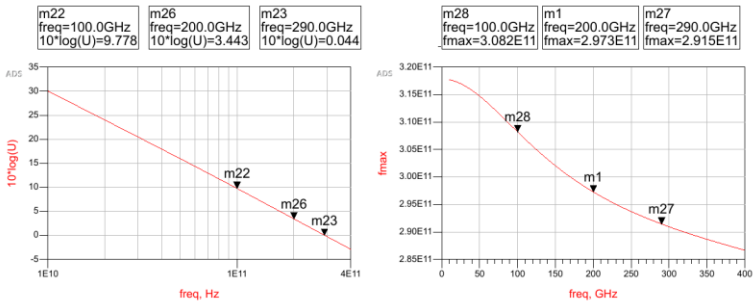


Figure E.2 Simulated Mason's gain (a) and extrapolated f_{max} (b) over frequency for a 28nm FD-SOI transistor with RCc extracted layout.

The targeted width for the unitary transistor is around 16µm to achieve a constant current but the manner of dividing the total width into fingers is not constrained. We aim to compare the combinations presented in Figure E.3:

- (a) W_{tot}=15.6µm, 26 fingers of 600nm
- (b) W_{tot}=16µm, 20 fingers of 800nm

- (c) W_{tot}=16µm, 16 fingers of 1000nm

Other parameters are of interest: the distance between the gate fingers and the drain and source fingers. It is a tradeoff between the C_{GD} and C_{GS} capacitances and the R_G access resistance that connects the gate fingers. The minimal allowed 33nm is compared to the maximal allowed 44nm.

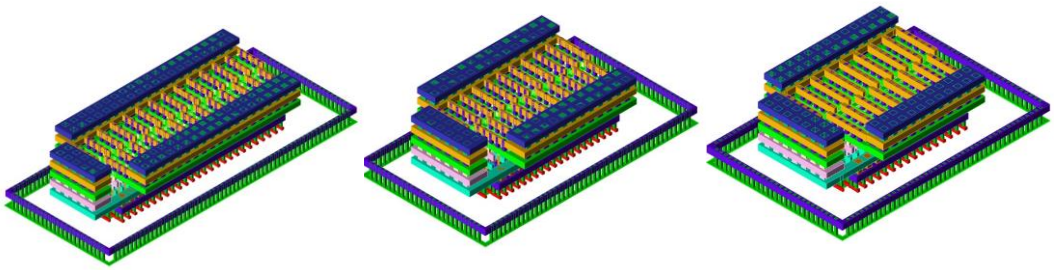


Figure E.3 3D views of multiple combinations of finger width and number of fingers to produce a W_{tot} of around 16µm: 26x600nm (a) 20x800nm (b) 16x1000nm (c)

In parallel the drain and source finger accesses can be optimized to reduce the C_{DS} capacitances by using only the half-width of the fingers to connect the higher metal layers as in Figure E.4 (a) or reduce the access resistances R_D and R_S as in figure E.4 (b).

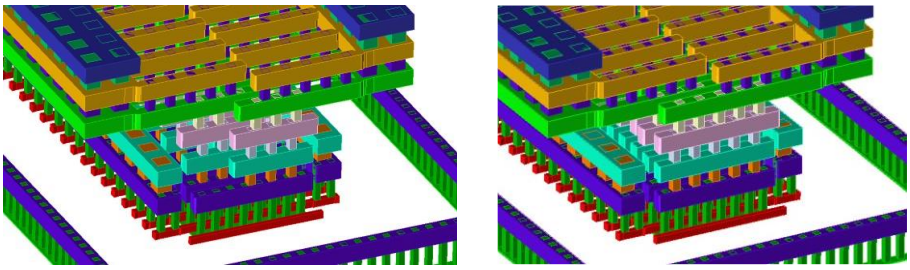


Figure E.4 3D view of the transistor staircase finger accesses using half the width (a) and full width (b) for the drain and source fingers access.

Each combination of the precedent parameters has been laid out as depicted in the previous figures, then the accesses have been extracted with an RCc model. Then all cases' f_{max} has been extracted at 200GHz and the results are summarized in Figure E.5. The 600nm gate fingers geometry presents the best f_{max} results, hence this tradeoff regarding finger width is preferred. Smaller 400nm and 500nm finger widths are not reported in the comparison but present equal or worse performances than the 600nm width. Next, the furthest distance between the gate and drain/source fingers offers the best results, therefore the reduction of the capacitances is a better tradeoff than the consequent gate interconnection resistance increase. Finally, the use of full-width drains and sources interconnections is better than the half-width. This comparison concludes with the different geometry optimization variables.

The best combination is the 26x600nm transistor with a 44nm gate to drain/source spacing and the full-width drain-source metal access. It offers a 306GHz f_{max} extracted at 200GHz.

Transistor size	15.6µm = 600n x 26				16µm = 800n x 20				16µm = 1000n x 16			
Drain/Source-Gate distance (nm)	33		44		33		44		33		44	
Plane contact on Drain/Source fingers	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
f_{max}(GHz)	301	302	305	306	297	299	300	303	290	292	294	295

Figure E.5 Comparison of f_{max} results of different transistor geometries.

All geometrical degrees of freedom from the transistor Pcell have been studied at this point and the observation is that the reduction in access resistances is usually the optimal to increase the f_{max}, especially the gate access resistance R_G. The part of the layout that presents the highest resistivity on the gate path is the M_1 and M_2 interconnection between the finger gates to the M_6 gate connection. Therefore, a final strategy that has been studied is the use of gate access on the two sides of the transistor cell as depicted in Figure E.6. It divides in two the highest resistance to a gate finger to a thick metal layer.

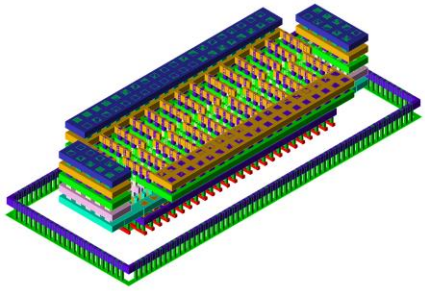


Figure E.6 3D view of a 20x600nm transistor cell laid out up to 6th thin metal layer with gate access on both sides.

Figure E.7 presents the simulated Mason's Gain a f_{max} of the two-side gate access transistor. It presents 3.88dB of gain and 313GHz f_{max} at 200GHz which is a significant improvement.

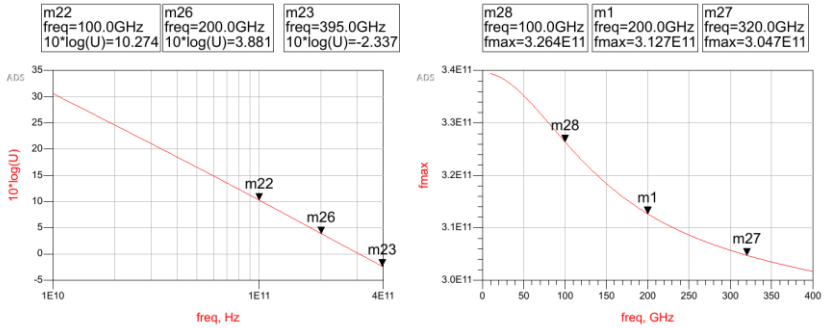


Figure E.7 Simulated Mason's gain (a) and extrapolated f_{max} (b) over frequency for a 28nm FD-SOI transistor with RCc extracted layout for a two-side gate access.

Multi transistor matrix

To support high currents the number of transistor fingers must be increased. However, the increase of transistor fingers at the thin metal layer level rapidly decreases the high-frequency performances because of the interconnection resistivity. However multiple unitary transistor cells that are optimized for the highest possible f_{max} can be interconnected into matrixes through thick top metal layers whose resistivities are low.

Figure E.8 presents a matrix of 6 unitary transistor cells of 26 times 600nm wide gate fingers that have been optimized in the previous section. A ground plane composed of the 5 first thin metal layers connects all the sources and also contacts the bodies of the transistors. Then the adjacent transistor's gates are connected through the 6 thin metal layers. This layout is designed for a common source topology use, common gate layout approach is presented later.

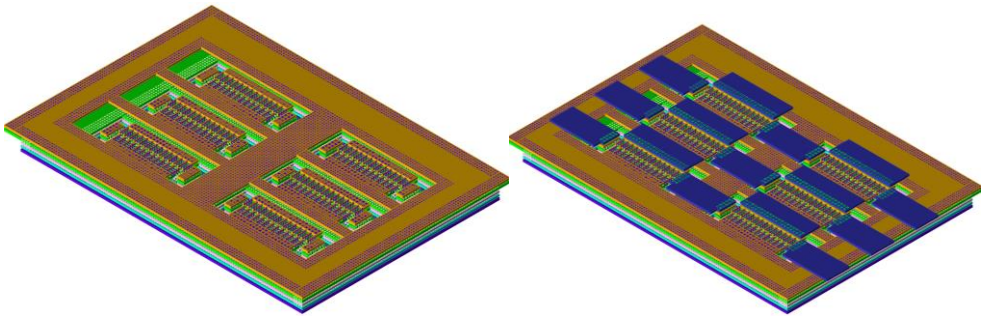


Figure E.8 3D view of a 6-transistor matrix up to the last thin metal layer

Then the input gates and output drains are each connected between them using the two top thick metal layers. Two different metal layers are used to reduce the electrical coupling. The result is a matrix of 6 transistors connected in parallel compactly with reduced interconnection parasitics, with one gate access, one drain access, and a ground plane. The result is a 156 transistor 600nm fingers matrix, which results in W_{tot}=93.6µm.

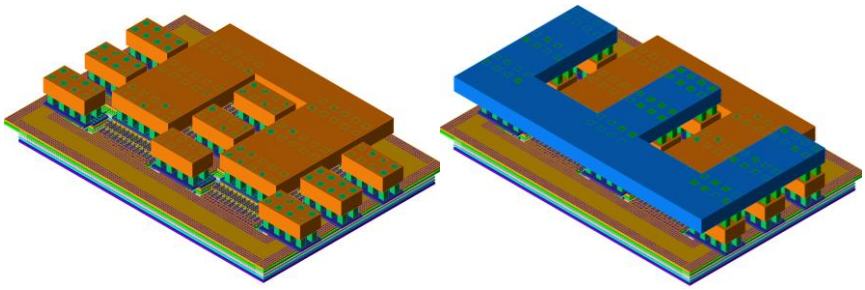


Figure E.9 3D view of a 6-transistor matrix up to the thick metal layers

The model of an RCc parasitic extraction of the transistor matrix is simulated to evaluate the resulting f_{max} and presents an f_{max} of 313GHz extracted at 200GHz which is equal to the performance of one unitary transistor cell. The meticulous matrix interconnection layout doesn't degrade the performance. Therefore, the limiting factor resides in the unitary transistor interconnection. As a matter of search for the optimal strategy, the 26 fingers unitary transistors are divided into 13 fingers unitary transistor cells. This way the gate average access resistance is even further reduced. The 12 transistor cells are interconnected similarly, and the 3D view of the layout is presented in Figure E.10. The f_{max} at 200GHz of a 13-finger transistor is equal to 317GHz and the resulting f_{max} extracted at 200GHz of the 12 transistors matrix with the RCc extracted model is also 317GHz.

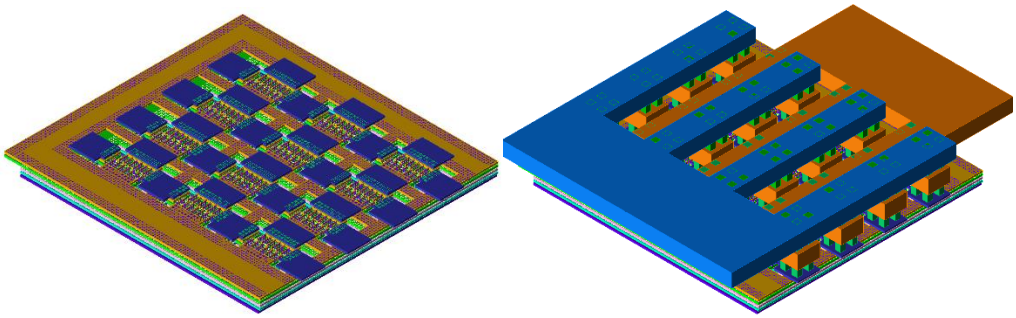


Figure E.10 3D view of a 12-transistor matrix

The division into smaller unitary transistor cells produces better performances in the conditions of the previous simulations and the reasoning can be pushed further by dividing each transistor cell into even smaller units. The complexity of the layout aside it seems to be an interesting solution.

However, the connections of the gate and drains with the thick metal layers are already thin in Figure E.11 with 12 units. The inductance whose effect becomes visible at the frequencies of operations is increased compared to the 6-transistor matrix case. It must be taken into account in the final modeling, and it is pertinent to consider it in the comparison between the 6 to 12-transistor matrix to decide whether a further division is interesting. Figure E.11 highlights the inductive sections that must be EM simulated and that are omitted with an RCc extraction.

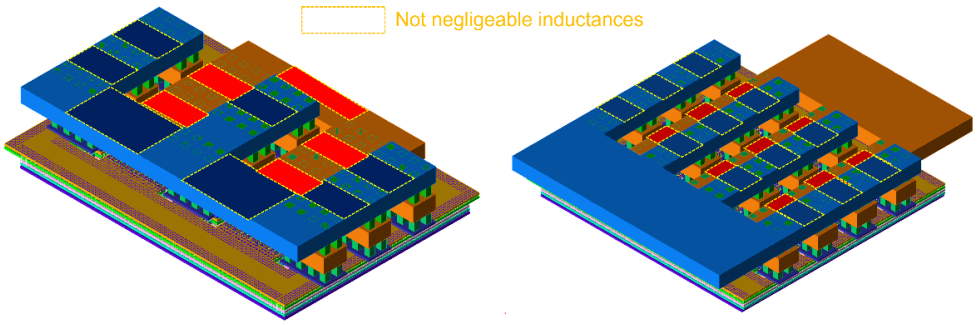


Figure E.11 3D view of a 6 and 12 transistor matrix with the non-negligible inductances highlighted

Consequently, a strategy to simulate the complex structure must be defined. As presented in Figure E.12 the transistor matrix is decomposed into the thin metal layers that are RCc extracted and the thick metal layers that are EM simulated. The finger accesses through thin metal layers have been shown to present a low inductance enough to be neglected in the extraction in the previous Chapter. The two parts are connected through the M6 that is present in both simulations and is used as the interface. In addition, a choice must be made regarding the electrical coupling of the thick metal layers to the ground plane. If the ground plane is present in both simulations, its resistivity is extracted twice and if it is absent in the EM simulation the electrical couplings are not captured. The decision has been made to keep the ground plane top metal in both extractions because of its already low resistivity and the need to capture all capacitive couplings.

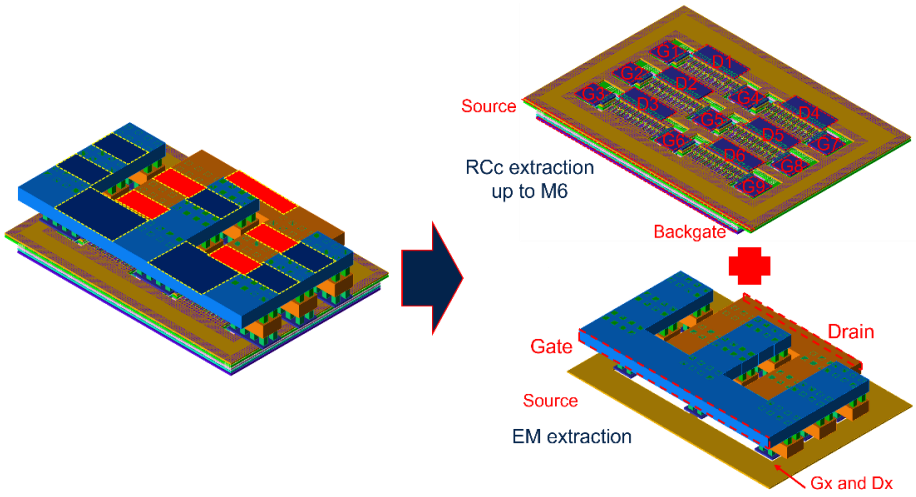


Figure E.12 Strategy of a complete simulation of the transistor matrix with the RCc extracted lower part and EM simulated higher part

The resulting model captures the resistive, capacitive, and inductive couplings. The procedure has been realized for the 6 and 12-transistor matrixes. The simulated Mason's gain and extracted f_{max} over frequency for the 6 transistors matrix is presented in Figure E.13. The gain and therefore

the f_{max} decreases faster than before because of the increasing impedance of the inductances with frequency. At 200GHz it results in an f_{max} of 291GHz.

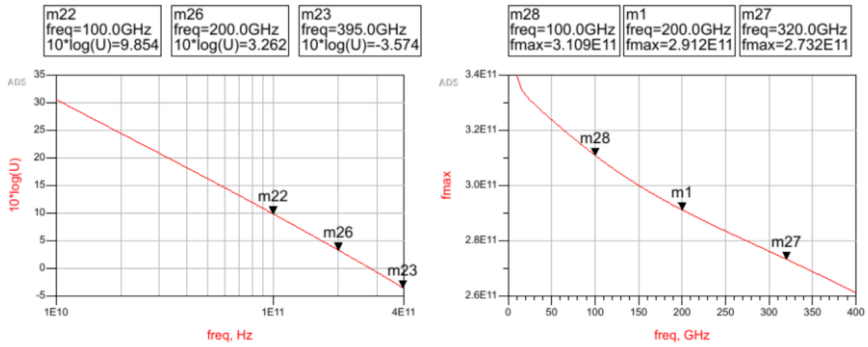


Figure E.13 Simulated Mason's Gain and f_{max} of 6 transistor matrix

In comparison, the 12 transistor matrix performances decrease even more and drop to a f_{max} of 276GHz at 200GHz. Therefore, there is an optimal transistor number in the matrix that is a tradeoff between the gate access resistance and the inductive accesses. The 6-transistor matrix is the optimal number here.

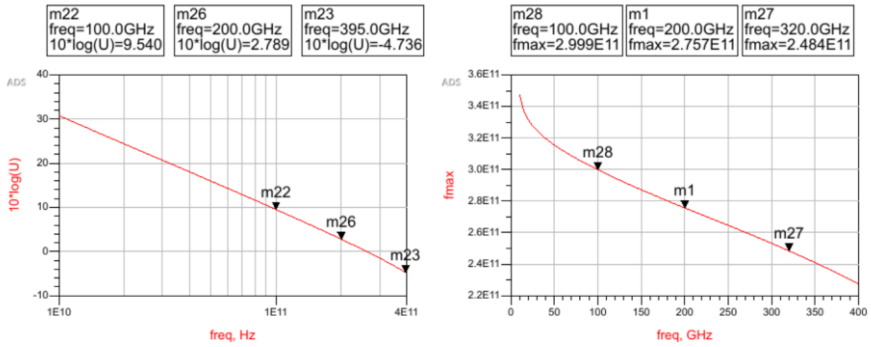


Figure E.14 Simulated Mason's Gain and f_{max} of 12 transistor matrix

Matrix type	6 transistors		12 transistors	
Extraction type	RCc	EM+RCc	RCc	EM+RCc
f_{max} (GHz)	313	291	317	276

Figure E.15 Comparison summary of the transistor matrixes' performances. L=30nm