



Cairo University

A 76-81 GHz FMCW Radar RF front-end for automotive applications in 65-nm CMOS Technology.

Team Members:

Belal Badr Mahmoud
Thabet Ahmed Hussien
George Wagih Fathi
Abdallah Essam-Eldin Hanafy
Nermeen Mohamed Abdelalim
Waheed Mohamed Ahmed

Under supervision of:

Dr. Ahmed Samir
Dr. Mohammed Ahmed El-Motaz Be-Allah

**Department of Electronics and Electrical Communications Engineering,
CAIRO UNIVERSITY.**

ABSTRACT

In the realm of automotive safety, radar plays a crucial role in collision avoidance and advanced driver assistance systems (ADAS). Radar sensors integrated into vehicles provide essential information about the surrounding environment, including the presence of other vehicles, pedestrians, and obstacles.

In this work, different radar system models are investigated with the design and implementation of the RF front-end of frequency-modulated, continuous-wave (FMCW) radar transceiver (TRX) in 65-nm CMOS technology for a frequency range of 76–81-GHz. Direct conversion architecture is used with different implementations for each block proposed to compare the different topologies.

The integrated receiver supports an input power range of -61dBm to -117dBm which corresponds to a detection range of 10m to 250m and achieves a total gain of 9dB, NF of 9dB, and *IIP3* of -11dBm. The Power amplifier transmits an output power of 13.7dBm.

In this thesis, the signal detection and processing techniques for radar systems are presented, In [chapter 2]. In [chapter 3], the direction of arrival estimation is discussed. In [chapter 4], interference and interference cancellation techniques are discussed. In [chapter 5], some important detection fundamentals are presented. In [chapter 6], system-level design with MATLAB simulations to determine the required specs for each block is presented. Chapters 7 to 9 include the results of each block in the following order LNA, Mixer, and PA. Finally, chapter 10 includes the integration results.

Sponsored by



AHEAD OF WHAT'S POSSIBLE™

Contents

INTRODUCTION	9
1.1. SENSOR TECHNOLOGY OVERVIEW	9
1.2. AUTOMOTIVE RADAR APPLICATIONS	10
1.2.1. Long-Range radar (LRR)	10
1.2.2. Short-Range radar (SRR)	11
1.3. FUNDAMENTALS OF RADAR SYSTEMS	11
1.4. LITERATURE REVIEW	12
1.4.1. historical background	12
1.4.2. Frequency band selection	13
1.4.3. Technology	14
REFERENCES	15
SIGNAL DETECTION AND PROCESSING	16
2.1. MATHEMATICAL TOOLS	16
2.1.1. Sampling Theorem	16
2.1.2. Discrete Fourier Transform (DFT)	17
2.2. CONTINUOUS WAVE RADAR	20
2.2.1. Unmodulated continuous wave radar	20
2.2.2. A frequency-modulated continuous wave radar	21
2.2.3. Triangular FMCW chirp signal	22
2.2.4. Multiple frequency shift keying waveform (MFSK)	23
2.2.5. Chirp sequence FMCW	24
2.2.6. Range and velocity resolution for FMCW radar systems	26
REFERENCES	27
DIRECTION OF ARRIVAL ESTIMATION	28
3.1. ANTENNA ARRAYS FUNDAMENTALS	28
3.1.1. Array Factor	30
3.1.2. Uniform Linear Array (ULA)	31
3.1.3. Linear Array in Rx mode	34
3.1.4. Analog VS digital beamforming	35
3.2. DOA ESTIMATION METHODS	36
3.2.1. Signal Model	37
3.2.2. Digital Beamforming (DBF) methods	38
Bartlett Beamformer	38
Capon Beamformer	39
Angle FFT method	41
Angle FFT Resolution	43
3.2.3. Multiple Signal Classifier (MUSIC)	44
3.2.4. Multi-dimensional DOA algorithms	46
REFERENCES	47
[2]. BALANIS, C.A. (2016) 'ARRAYS: LINEAR, PLANAR, AND CIRCULAR', IN ANTENNA THEORY: ANALYSIS AND DESIGN. 4TH EDN. HOBOKEN, NJ: WILEY	47
INTERFERENCE AND INTERFERENCE MITIGATION	48
4.1. AUTOMOTIVE RADAR INTERFERENCE	48

4.1.1.	Signal model for radar and the interference -----	49
4.1.2.	Characteristics of the automotive radar interference -----	51
4.1.2.1.	Types of the automotive radar interference-----	51
4.1.2.2.	Duration of the Interference-----	54
4.1.2.3.	Interference Effects in range profiles-----	56
4.1.2.4.	Summary-----	57
4.2.	INTERFERENCE MITIGATION -----	57
4.2.1.	Detection of interference -----	58
4.2.2.	Interference Mitigation and Avoidance -----	59
4.2.2.1.	Nulling and signal restoration -----	59
4.2.2.2.	Frequency altering and timing jitter -----	60
4.2.2.3.	Cooperative Time and frequency scheduling-----	60
4.2.2.4.	Digital beamforming and null steering-----	60
4.2.3.	Summary -----	62
REFERENCES	-----	63
DETECTION FUNDAMENTALS	-----	65
5.1.	PROBABILITY OF DETECTION AND FALSE ALARM DETECTION-----	65
5.2.	COHERENT AND NONCOHERENT INTEGRATION-----	66
5.2.1.	Coherent integration-----	67
5.2.2.	Non-coherent integration-----	67
5.3.	PEAK DETECTION -----	69
5.3.1.	CA-CFAR -----	69
5.3.2.	Cell-Averaging Greatest-Of CFAR (CAGO-CFAR)-----	70
5.3.3.	Cell-Averaging Smallest -Of CFAR (CASO-CFAR)-----	70
5.3.4.	OS-CFAR -----	70
REFERENCES	-----	72
SYSTEM-LEVEL DESIGN	-----	73
6.1.	PERFORMANCE PARAMETERS -----	73
6.1.1.	System-added Noise-----	73
6.1.2.	Linearity -----	74
6.1.2.1.	1-dB compression point (<i>P1dB</i>) -----	75
6.1.2.2.	Intermodulation -----	76
6.1.2.3.	Cascaded Non-Linearity-----	77
6.1.3.	Dynamic Range -----	77
6.1.4.	Quantization noise-----	78
6.2.	SYSTEM ARCHITECTURE -----	82
6.3.	SYSTEM SPECIFICATIONS -----	82
6.3.1.	System budget-----	82
Received power range	-----	82
IF frequency range	-----	83
SNR calculation for chirp sequence FMCW	-----	83
6.3.2.	Single object model-----	84
6.3.3.	Two object model-----	84
6.3.4.	MATLAB simulation-----	85
6.3.4.1.	LNA specs-----	85
6.3.4.2.	Mixer specs -----	86
REFERENCES	-----	88
LOW NOISE AMPLIFIER	-----	89

7.1.	NOISE -----	89
7.1.1.	Resistor thermal noise -----	89
7.1.2.	MOSFET noise sources-----	90
7.1.2.1.	Trasistor thermal noise-----	90
7.1.2.2.	Trasistor flicker noise-----	90
7.1.3.	Noise Figure -----	91
7.1.4.	Constant Noise Circles -----	94
7.2.	GAIN -----	95
7.3.	LINEARITY-----	95
7.4.	STABILITY -----	96
7.5.	CASCODE LNA AND BALUN DESIGN -----	98
7.5.1.	Target Specifications -----	98
7.5.2.	Topology Selection and Design Choices -----	98
7.5.3.	Post-Layout simulation results-----	99
7.5.3.1.	Gain -----	99
7.5.3.2.	Noise Figure-----	101
7.5.3.3.	Input/Output matching (<i>S₁₁</i> and <i>S₂₂</i>) -----	101
7.5.3.4.	Linearity-----	102
7.5.3.5.	Stability-----	103
7.5.3.6.	Reverse Isolation (<i>S₁₂</i>)-----	104
7.5.4.	Supply variations and Croners Simulation -----	104
7.5.5.	LNA Layout-----	109
	Performance Summary and Literature Survey -----	110
7.5.6.	Balun Design-----	111
7.6.	COMMON SOURCE-----	114
7.6.1.	Post-Layout Simulation Results-----	114
7.6.1.1.	Gain -----	114
7.6.1.2.	Noise Figure-----	115
7.6.1.3.	Input and Output Matching (<i>S₁₁</i> & <i>S₂₂</i>) -----	116
7.6.1.4.	Linearity-----	117
7.6.1.5.	Stability-----	118
7.6.1.6.	Reverse Isolation (<i>S₁₂</i>)-----	119
7.6.2.	Supply Variations and Corners Simulation-----	119
7.6.3.	LNA Layout.-----	122
	REFERENCES -----	124
	MIXER -----	125
8.1.	MIXER BASIC OPERATION -----	125
8.2.	MIXER PERFORMANCE METRICS-----	126
8.2.1.	Conversion gain/loss -----	126
8.2.2.	Noise -----	126
8.2.3.	Intermodulation and linearity-----	127
8.2.4.	Harmonic suppression -----	128
8.2.5.	Isolation -----	128
8.2.6.	Operating frequency range-----	128
8.3.	MIXER TOPOLOGIES-----	129
8.3.1.	Diode mixer-----	129
8.3.2.	Single-ended diode mixers-----	129
8.3.3.	Single-balanced diode mixers-----	129

8.3.4. Double-balanced diode mixers -----	129
8.3.5. Resistive passive mixer -----	129
8.3.6. Active FET mixer-----	130
8.3.7. Single-ended FET mixers -----	131
8.3.8. Single-balanced FET mixers-----	131
8.3.9. Double-balanced FET mixers -----	131
8.4. DOWN-CONVERTER MIXER -----	132
8.4.1. Mixer core -----	132
8.4.2. Mixer Core Simulation Results-----	133
8.4.2.1. Output (IF) matching-----	133
8.4.2.2. Input (RF) matching-----	133
8.4.2.3. Conversion gain-----	134
8.4.2.4. Noise Figure-----	134
8.4.2.5. OIP3-----	136
8.4.2.6. LO-RF Port Isolation-----	136
8.4.2.7. LO-IF Port Isolation-----	136
REFERENCES -----	140
POWER AMPLIFIER-----	141
9.1. POWER MATCHING AND LOAD-PULL TECHNIQUE -----	141
9.2. MAIN POWER AMPLIFIER SPECIFICATIONS -----	143
9.2.1. Output power-----	143
9.2.2. Efficiency -----	143
9.2.3. Linearity -----	143
9.3. CLASSES OF POWER AMPLIFIERS-----	144
9.3.1. Class-A power amplifier -----	145
9.3.2. Class-B power amplifier -----	145
9.3.3. Class-AB power amplifier -----	146
9.3.4. Class-C power amplifier -----	146
9.4. COMMON SOURCE IMPLEMENTATION -----	148
9.4.1. required specs -----	148
9.4.2. Structure -----	148
9.4.2.1. Neutralization capacitors -----	148
9.4.2.2. Transformer-based matching-----	149
9.4.2.3. Cascading stages-----	150
9.4.3. Design procedure -----	151
9.4.3.1. Current density selection-----	151
9.4.3.2. PA stage transistors Sizing-----	151
9.4.3.3. Output matching-----	151
9.4.3.4. Driver stages transistor Sizing-----	151
9.4.3.5. Interstage matching-----	151
9.4.4. Layout-----	152
9.4.5. Post-layout simulation results -----	153
9.4.6. Corners -----	161
9.4.7. Literature-----	169
9.5. CASCODE IMPLEMENTATION -----	170
9.5.1. Architecture -----	170
9.5.2. Structure -----	170
9.5.3. Topology -----	171

9.5.4.	Design methodology -----	171
9.5.4.1.	Power amplifier schematic-----	172
9.5.4.2.	Ideal Results -----	173
9.5.5.	PA layout -----	177
9.5.6.	Post-layout simulations: -----	178
9.5.7.	Corners: -----	184
REFERENCES	-----	189
RECEIVER INTEGRATION RESULTS-----		190
10.1.	SIMULATION RESULTS -----	190
10.2.	TRANSCEIVER LAYOUT -----	194
10.3.	LITERATURE -----	195
REFERENCES	-----	196

1

INTRODUCTION

In recent years, the automotive industry has witnessed significant technological advancements, leading to the development of autonomous cars and advanced driver assistance systems (ADAS). These systems aim to enhance the safety, efficiency, and overall driving experience, relying on the input signals from all the available sensors such as cameras, ultrasonic, LiDAR, radar, etc.

A radar, radio detection and ranging, is a system used to locate, track and identify different objects within a particular area by transmitting electromagnetic energy towards them and observing their received echoes. In the early days, radar systems were restricted to military applications due to their bulky size and high cost. Thanks to the advance of high-frequency integrated circuits (ICs), modern radar systems can be realized on printed circuit boards (PCBs) [1] or even on an IC [2, 3]. Owing to the high integration density of the CMOS technology, it is now possible to integrate RF front-end circuitry along with digital signal processing (DSP) blocks on the same chip enabling even more powerful systems with very sophisticated signal processing algorithms to be fabricated in a smaller area.

In the following section, we review the common sensor technologies used for advanced driver assistance systems and investigate the advantages and disadvantages of each type.

1.1. Sensor technology overview

There are four leading sensor technologies used today for automotive applications. Each technology has advantages that make it appealing. These four technologies are:

1. **Cameras:** Cameras are perhaps the most familiar and widely used sensors in ADAS. They capture visual information through images or video, allowing the system to recognize and understand the environment. Cameras provide valuable data for tasks such as lane departure warnings, traffic sign recognition, pedestrian detection, and object classification. They specialize in identifying and tracking objects based on their visual appearance and characteristics. Drawbacks of this technology are the ambient light and weather conditions dependence. However, they are masters of textures interpretation.

2. **Radar:** Radar systems utilize radio waves to detect objects and measure their distance, speed, and angle. Radars are robust to adverse weather conditions such as rain and snow [4], providing reliable long-range detection capabilities. It can be installed behind the front bumper fascia. Radar sensors excel in detecting and tracking moving objects, making them essential for situational awareness in autonomous driving.
3. **LiDAR:** LiDAR, Light Detection and Ranging, sensors employ laser beams to measure distances and create a detailed 3D map of the environment. By measuring the time it takes for laser pulses to bounce back from objects, lidar systems can generate high-resolution point cloud data. This data enables precise localization, object detection, and mapping of the surrounding environment. LiDAR sensors are valuable for tasks like autonomous mapping, object detection and tracking, and path planning. They provide accurate depth perception and can detect static and dynamic objects. Unlike radar, lidar cannot operate under all weather conditions.
4. **Ultrasonic Sensors:** Ultrasonic sensors use sound waves to detect objects and measure distances. They are typically used for short-range detection and provide close-range proximity sensing capabilities. Ultrasonic sensors are commonly employed in parking assistance systems to detect obstacles and assist with parking maneuvers. They are also helpful for low-speed collision avoidance and pedestrian detection at close distances.

Sensors do not directly depend on each other but can be combined to give a complete image of the environment, known as the sensor fusion approach. In this approach, we can use each sensor technology's advantages and build a robust and reliable monitoring system. Radars are extensively used in automobiles compared to other sensors because of their reliability and ability to work under bad weather conditions.

1.2. Automotive radar applications

In real traffic situations, relying solely on a single radar is inadequate to ensure optimal safety. Consequently, automotive systems necessitate the integration of multiple radars, each serving distinct functions. These radars are typically classified into two categories based on their detectable distance: long-range radar (LRR) and short-range radar (SRR) [12]. By combining these radar types, vehicles can effectively perceive their surroundings and respond to various objects and obstacles at different distances, enhancing security and collision avoidance capabilities.

1.2.1. Long-Range radar (LRR)

As its name implies, Long-Range Radar is specifically designed to detect objects located at a significant distance from the vehicle (detection range 10m-250m). LRR systems operate at higher frequencies, typically within the 76 to 77 GHz range. This high-frequency range enables high resolution and accuracy when detecting and tracking objects. Normally positioned at the front of the vehicle, behind the front grille or bumper, LRR radar sensors offer an extensive detection range.

The primary objective of LRR is to provide long-range perception capabilities, particularly in support of ADAS. One notable application is adaptive cruise control (ACC), a system that automatically adjusts the vehicle's speed to maintain a safe distance from the car ahead. LRR aids in precisely measuring the distance and relative speed of the preceding vehicle, allowing the ACC system to respond accordingly.

Furthermore, LRR is critical in forward collision warning and autonomous emergency braking systems.

By detecting the presence and velocity of vehicles or obstacles in front of the vehicle, LRR can issue early warnings to the driver or activate automatic emergency braking to prevent or mitigate collisions. This capability enhances overall safety by providing proactive measures to avoid potential accidents.

1.2.2. Short-Range radar (SRR)

Short-Range Radar detection is designed to detect objects close to the vehicle (detection range 0.15m-30m) and typically operates within a frequency range of 77 to 81 GHz. SRR sensors are positioned at various locations on the vehicle, including the front, rear, and sides.

SRR finds widespread application in multiple scenarios, such as blind spot detection and rear cross-traffic alert systems. By actively monitoring the vehicle's sides and rear, SRR sensors excel at detecting the presence of vehicles in blind spot areas or approaching from the sides while the vehicle is in reverse. This valuable information is then utilized to provide drivers with visual or audible warnings, enhancing their awareness and significantly reducing the risk of accidents.

Additionally, SRR plays a crucial role in parking assistance systems. SRR sensors enable drivers to perform more accurate parking maneuvers by detecting nearby objects and measuring their distance. They provide proximity alerts or steering guidance, facilitating safer and more efficient parking experiences.

1.3. Fundamentals of Radar Systems

The working principle of radar systems is relatively simple. The transmitter sends electromagnetic energy from either a single antenna or an antenna array, and the surrounding objects intercept the radiation and reflect some of the electromagnetic energy depending on their material properties. The receiver utilizes one or more receive channels to capture the reflected power, and by analyzing the received signals in the digital domain, we can determine the targets' range, velocity, and relative positions in space. Signal processing algorithms used to extract this information from the received signal will be discussed later in the following chapters.

Radar systems are typically categorized into two main categories [12], continuous-wave (CW) and pulsed radars, based on the types of waveforms radiated by the transmitter. In pulsed radars, a short-duration pulse is emitted periodically. Between every two successive pulses, the receiver waits for the reflected pulse. The time between the transmitted and received pulses determines the target range. At the same time, the velocity is estimated from the frequency shift between the two pulses, utilizing the Doppler effect. On the other hand, a CW radar radiates electromagnetic energy continuously. The receiver compares the received signal with a replica of the transmitted waveform and extracts the targets' properties. CW radar is generally categorized into two types: unmodulated continuous wave and modulated continuous wave.

The carrier is transmitted without phase or frequency modulation in an unmodulated CW radar, and this limits the detection capabilities of the system as it depends solely on the Doppler shift in the received signal, which means that stationary objects will be undetectable. We can only detect the velocity of moving objects and not their range. Modulated CW radar has two types: frequency-modulated (FM) and phase-modulated (PM) continuous wave radars. Modulating the carrier allows the system to estimate both the velocity and range of targets simultaneously.

FMCW radars have different waveform variations and different methods of changing the carrier frequency with time, such as triangular waveform, fast chirp ramp sequence waveform, MFSK waveform, and many more. Each waveform utilizes a particular signal-processing algorithm to extract the targets' range,

velocity, and relative position. The most commonly used waveform is the chirp ramp sequence waveform, which will be discussed along with other waveforms later in the following chapters.

CW radar systems are dominant in automotive applications, especially FMCW radar systems, due to their advantages over pulsed radar systems. CW radar features low peak transmit power and is simpler than pulsed radar.

1.4. Literature review

1.4.1. historical background

The utilization of millimeter-wave-based radar technology in automobiles was first reported in the early 1970s when researchers and engineers started exploring its potential for enhancing vehicle safety through collision avoidance. During this time, the development of radar-based applications and prototypes began. However, until 1999, the integration-unfriendly nature of the technology, along with its large size and high cost, posed significant challenges and prevented radar products from entering the consumer market. The early radar systems did not fulfill the demanding criteria of the automotive sector concerning size, effectiveness, and affordability.

In 1998, a significant breakthrough occurred with the introduction of the first generation of automotive radar sensors operating in the 77 GHz frequency band. Daimler introduced the pioneering "Distronic" system as an optional feature in their S-Class vehicles. The Distronic system combined radar sensors with adaptive cruise control (ACC) functionality, allowing the vehicle to automatically regulate its speed and maintain a safe distance from the vehicle ahead.

Following Daimler's lead, other automotive manufacturers, including Jaguar, Nissan, and BMW, introduced radar-based systems in their vehicle lineups in the coming years. By 2003, radar systems had become prevalent in the upper-class segments of most major car manufacturers. Initially, these systems served as a comfort function, enhancing driving convenience and reducing driver fatigue.

As technology advanced and production costs decreased, radar-based driver assistance systems became more affordable and accessible. This led to their integration into mid-range and even entry-level vehicles, expanding their availability beyond luxury cars. Today, radar-based driver assistance systems have become commonplace, offering many features and applications.

With the increasing focus on vehicle safety and the drive towards autonomous driving, the number of radar sensors integrated into new vehicles has been steadily growing. This proliferation allows for a wide range of driver assistance functions that enhance comfort and safety. Radar technology continues to evolve, with ongoing advancements in sensor design, signal processing, and integration with other sensor modalities, contributing to developing more advanced and sophisticated automotive radar systems.

1.4.2. Frequency band selection

Two primary frequency bands, namely the 24 GHz and 77 GHz bands, are commonly used in automotive radar systems. Other frequencies (e.g., below 10 GHz or above 100 GHz) have been explored, but they currently have no practical applications [5]. There is ongoing competition between the 24 GHz and 77 GHz bands, and from an engineering standpoint, the 77 GHz band presents more challenges than the 24 GHz band. However, the 77 GHz band offers more significant opportunities for achieving higher-performance radar sensors.

- Comparison between the 24 GHz and 77 GHz bands:

The antenna aperture determines the size of an automotive radar sensor, which is crucial for flexible integration into vehicles. Operating at 77 GHz allows for a smaller antenna size, meeting beamwidth requirements and enabling better angular resolution with a smaller sensor size. In contrast, sensors using 24 GHz require an aperture approximately three times larger to achieve the same level of performance.

Another advantage of the 77 GHz band is the potential introduction of short-range sensors with high spatial resolution. These sensors would have a total bandwidth of up to 4 GHz, corresponding to a fractional bandwidth of only about 5% at 77 GHz, compared to nearly 17% at 24 GHz. This makes the design of antennas and wavelength-dependent components easier for 77 GHz systems.

Furthermore, the combination of high transmit power and high bandwidth is not allowed at 24 GHz but is available at 77 GHz. This enables long-range operation and increased distance separability.

On the other hand, at 24 GHz, it is possible to achieve better performance (power consumption, noise figure, performance margins) using the same technology node or by utilizing a different, potentially lower-cost technology node. Additionally, sensors operating at 77 GHz are typically much smaller, resulting in cost savings related to volume and weight.

- Frequency regulation

In addition to technical requirements, frequency regulation plays a significant role in determining the operation parameters of radar sensors. Regulatory bodies such as the FCC in the US and CEPT in Europe govern frequency regulations. Different countries may have their regulations, sometimes conflicting with neighboring countries. Two frequency bands have emerged for automotive radar sensors in the millimeter-wave range: the 76-77 GHz band, available worldwide, and the neighboring 77-81 GHz band, introduced in Europe as a replacement for ultrawideband (UWB) automotive radar sensors in the 24 GHz band. The 77-81 GHz band provides a larger bandwidth of 4 GHz compared to 1 GHz in the 24 GHz band [5].

1.4.3. Technology

The millimeter-wave front end of automotive radar sensors undergoes significant changes with each new generation, driven by advancements in semiconductor technology, packaging technology, and antenna systems. The primary goals for the technology node selection are to achieve higher performance, increased integration, and lower power consumption.

The first commercial 76 GHz automotive radar for passenger cars occurred in 1998/99 by Mercedes Benz. Initially, it was developed by Macom in the USA and utilized a GUNN oscillator and a microstrip receiver circuit. Subsequently, GaAs MMICs (Monolithic Microwave Integrated Circuits) were incorporated, improving the radar's functionality [7].

In 2010, silicon-based Silicon Germanium (SiGe) technology became more competitive in speed and cost, and a transition to silicon began. This shift allowed for higher integration densities, added functionality at a reduced cost, and facilitated mass production. Complete radar transceivers, including digital components, became achievable for the first time. This transition greatly simplified the implementation of the millimeter-wave section of the radar sensor itself.

The next anticipated technological transition is the CMOS (Complementary Metal-Oxide-Semiconductor) technology. Modern CMOS technologies offer high integration densities for digital circuits and can provide satisfactory RF performance. Transistor speed, a crucial aspect for CMOS feasibility in automotive radar at 77 GHz, has shown promising results in various CMOS processes, such as 45 nm down to 22 nm nodes, rivaling SiGe-BiCMOS technology. The primary driver for transitioning to CMOS is integration. In the coming years, new high-volume applications for radar sensors may emerge, justifying the substantial development costs associated with a CMOS solution. However, it is worth noting that the design costs for modern CMOS processes are expected to increase significantly due to circuit complexity and the need for advanced design tools.

Initial implementations of highly integrated 77 GHz radar transceivers in pure CMOS, using 90 nm and 65 nm nodes, have been published [5,9,10], demonstrating the technical feasibility of such systems. However, there are still challenges to overcome before a genuine automotive-grade radar sensor can be realized using CMOS technology. Achieving the required output power, dynamic range, and linearity and addressing the temperature range requirements, particularly for the power amplifier, remain areas of concern for CMOS technology [11]. So, for now, SiGe technology, in its BiCMOS form, is expected to be the dominant semiconductor technology for automotive millimeter-wave radar in the next few years.

References

- [1]. Yeap S.B., Qing X., Chen Z.N. ‘77-GHz dual-layer transmit-array for automotive radar applications’. *IEEE Transactions on Antennas and Propagation*. 2015, vol. 63(6), pp. 2833–7.
- [2]. Lee J., Li Y.-A., Hung M.-H., Huang S.-J. ‘A fully-integrated 77-GHz FMCW radar transceiver in 65nm CMOs technology’. *IEEE Journal of Solid-State Circuits*. 2010, vol. 45(12), pp. 2746–56.
- [3]. Luo T.N., Wu C.H.E., Chen Y.J.E. ‘A 77-GHz CMOs automotive radar transceiver with anti-interference function’. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2013, vol. 60(12), pp. 3247–55.
- [4]. B. -J. Jang, I. T. Jung, K. Soo Chong, H. K. Sung and S. Lim, "Rainfall Observation System using Automotive Radar Sensor for ADAS based Driving Safety," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 2019, pp. 305-310, doi: 10.1109/ITSC.2019.8917080.
- [5]. Hasch, J. et al. (2012) ‘Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band’, *IEEE Transactions on Microwave Theory and Techniques*, 60(3), pp. 845–860. doi:10.1109/tmtt.2011.2178427.
- [6]. Waldschmidt, C., Hasch, J. and Menzel, W. (2021) ‘Automotive radar — from first efforts to Future Systems’, *IEEE Journal of Microwaves*, 1(1), pp. 135–148. doi:10.1109/jmw.2020.3033616.
- [7]. M. Camiade, “Overview on GaAs MMICs for automotive radar,” in Proc. 24th Annu. Tech. Dig. Gallium Arsenide Integr. Circuit Symp., 2002, pp. 29–32
- [8]. T. Mitomo, N. Ono, H. Hoshino, Y. Yoshihara, O. Watanabe, and I. Seto, “A 77 GHz 90 nm CMOS transceiver for FMCW radar applications,” *IEEE J. Solid-State Circuits*, vol. 45, no. 4, pp. 928–937, Apr. 2010.
- [9]. Y. Kawano, T. Suzuki, M. Sato, T. Hirose, and K. Joshin, “A 77 GHz transceiver in 90 nm CMOS,” in Proc. IEEE Solid-State Circuits Conf., Feb. 2009, pp. 310, 311a–311.
- [10]. Y.-A. Li, M.-H. Hung, S.-J. Huang, and J. Lee, “A fully integrated 77 GHz FMCW radar system in 65 nm CMOS,” in Proc. IEEE Solid-State Circuits Conf., Feb. 2010, pp. 216–217.
- [11]. J.-J. Lin, K.-H. To, D. Hammock, B. Knappenberger, M. Majerus, and W. Huang, “Power amplifier for 77-GHz automotive radar in 90-nm LP CMOS technology,” *IEEE Microw. Wireless Compon. Lett.*, vol. 20, no. 5, pp. 292–294, May 2010.
- [12]. Gamba, J. (2020) in *Radar Signal Processing for autonomous driving*. Singapore: Springer.

2

Signal Detection and Processing

In the previous chapter, we briefly introduced the types of radar systems regarding the used waveform, such as pulsed radar and continuous-wave radar. In this chapter, the working principle of continuous-wave radar systems is discussed in detail, explaining how the required parameters (i.e., velocity and range) are estimated. This chapter's first section will discuss mathematical tools for all continuous-wave radar systems. The second section discusses the technical details of different continuous-wave radar systems.

2.1. Mathematical Tools

2.1.1. Sampling Theorem

Digital signal processing plays a vital role in estimating the required parameters from the down-converted signal in radar systems. An analog-to-digital converter samples and quantizes this signal to be processed in the digital domain. This is where the sampling theorem, also known as the Nyquist-Shannon sampling theorem, comes into play. It is one of the main concepts in digital signal processing as it explains how sampling alerts the properties of the input analog signal.

The sampling process is done by converting a continuous-time signal to a discrete-time signal by taking “samples” of the continuous-time signal at discrete time instants. Thus, if an input $x(t)$ is sampled, the output will equal $x(nT_s)$ where T_s is called the sampling interval.

The theorem states that to sample a continuous-time signal without any loss of information, the sampling rate must be at least twice the highest frequency component present in the signal. If the signal has a bandwidth of B , the sampling rate should be $2B$ or higher ($f_s > 2B$) to avoid information loss and aliasing effects.

Aliasing is the distortion that happens to the signal when sampled at a low rate. Sampling the signal in the time domain causes periodicity in the frequency domain, so the signal spectrum is infinitely repeated in the frequency domain. Suppose double the signal's bandwidth is larger than the repetition period. In that case, the repeated spectrums will overlap, causing aliasing, which means that the original signal cannot be regenerated from the sampled signal (information loss).

2.1.2. Discrete Fourier Transform (DFT)

Frequency estimation is essential for CW radar systems, and DFT is one of the main algorithms used for frequency estimation. The DFT is a mathematical technique used to analyze and convert discrete time domain signals to the frequency domain, and the significant advantage of the DFT is that it is discrete in both the time domain and frequency domain, making it suitable for digital processing. The effectiveness of using DFT relies not only on the frequency content of signals but also on strategic sampling techniques.

The DFT enables us to express a discrete-time signal by combining various complex exponential functions with different frequencies. This decomposition process provides us with significant knowledge about the spectral attributes of the signal, such as identifying both the magnitudes and phases associated with each frequency component in the signal. And this is very helpful for CW radar where the down-converted signal is usually a single tone at a certain intermediate frequency (IF) frequency which has the information of the speed and range of the target, so applying DFT to this signal the power at this frequency will be the highest so to measure this frequency we should measure the frequencies with power above some certain threshold.

The DFT processes a finite sequence of discrete samples, usually obtained by sampling continuous-time signals. The main purpose of the DFT is to convert a sequence of N samples into an equivalent sequence of complex numbers, referred to as the frequency-domain representation or DFT spectrum.

The DFT of a sampled signal $x(n)$ is given by:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi n K}{N}}, k = 0, 1, \dots, N-1, \quad 2.1$$

Where n and k are the sampling time and frequency indices, respectively, and N is the number of samples.

The discretization in the frequency domain in DFT implies periodicity in the time domain, so DFT assumes the time domain signal to be periodic and infinitely repeated, and as the real-time signals are time-limited, the DFT considers these signals to be infinitely repeated with period N , so for a discrete-time signal with L samples applied to DFT, L should be less than or equal N avoid time domain aliasing ($N \geq L$).

As shown in Eqn. 2.1, the DFT consists of addition and multiplication operations so that it is easily implemented, which is why it is so common in digital signal processing, practically the DFT is computed efficiently using the Fast Fourier transform (FFT) algorithm, which reduces the computational complexity from order to, the FFT is used in each of the following CW detection techniques.

Frequency resolution of the DFT

For any two frequencies f_1 and f_2 to be distinguishable in the frequency domain after applying DFT, the difference between these frequencies must be greater than or equal $(1/L)f_s$ which equals $(1/LT_s)$, where T_s is the sampling time, and L is the number of samples, so the resolution bandwidth will equal $(1/T)$, where $T = LT_s$, which is the total time of the time domain signal applied to the DFT. The value $(1/L)f_s$ is known as the resolution bandwidth or bin width, and it represents the smallest frequency difference that can be resolved by the DFT.

Equivalent Noise Bandwidth ENBW

The signals applied to DFT are time-limited, but the DFT considers these signals to be infinitely repeated, and this may cause discontinuities at the edges of these periodic extensions, which will cause spectral leakage, which means that the energy contained in one bin will leak into other bins, the spectral leakage will act as noise.

To reduce the spectral leakage, windowing functions are used; these functions are applied to the input data sequence as a multiplicative weighting which smoothes the discontinuities at the edges of the periodic extensions of the input sequence (windowed data are smoothly brought to zero at the boundaries) which will reduce the spectral leakage. As the DFT of a sampled signal $x(n)$ is given by Eqn. 2.1. The outputs of the DFT can be viewed as passing the signal $x(n)$ through a bank of N complex filters and taking the first sample from the output of each filter. The impulse response of the k^{th} filter can be written as:

$$h_k(n) = \begin{cases} e^{-\frac{j2\pi kn}{N}} & 0 \leq n < N \\ 0 & , \text{otherwise} \end{cases} \quad 2.2$$

which in the frequency domain represents sinc functions centered at $f_k = k/NT$.

When applying different windowing functions $w(n)$, then the transform of the windowed samples will be given by:

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\frac{2\pi nk}{N}}, k = 0, 1, \dots, N - 1 \quad 2.3$$

This is equivalent to shaping the response of the filter with this windowing function, so the new impulse responses of the filters are given by:

$$h[n] = \begin{cases} w(n)e^{-\frac{j2\pi kn}{N}} & 0 \leq n < N \\ 0 & , \text{otherwise} \end{cases} \quad 2.4$$

which in the frequency domain translates to having the frequency response of the windowing function as the shape of the filter centered at $f_k = k/NT$.

So according to this, having different windowing functions results in having different responses of the filter in the frequency domain, which changes the amount of accumulated noise in each bin in the DFT.

The equivalent noise bandwidth is the bandwidth of a fictitious brick-wall filter that gives the same

integrated noise as that produced for a certain windowing technique which is illustrated in Fig. 2.1: Illustrating equivalent noise bandwidth., and it is calculated as follows:

$$P_{\text{window}} = \int_{-f}^f |W(f)|^2 df = B_{\text{enbw}} \times |W(f_0)|^2. \quad 2.5$$

Where P_{window} is the actual integrated noise power, and B_{enbw} is the equivalent noise bandwidth.

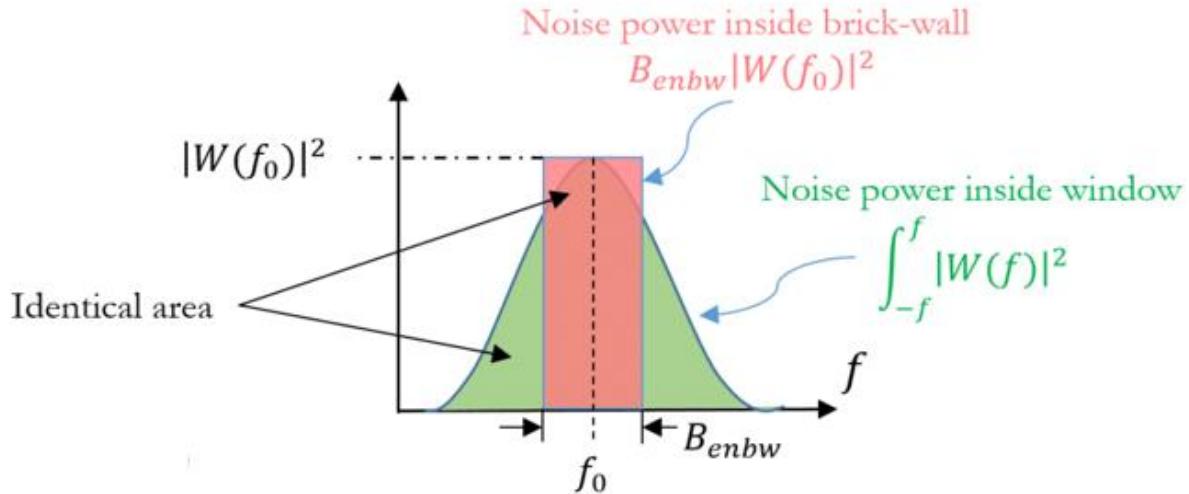


Fig. 2.1: Illustrating equivalent noise bandwidth.

By using Parseval's theorem:

$$B_{\text{enbw}} = \frac{\int_{-f}^f |W(f)|^2 df}{|W(f_0)|^2} = \frac{\sum |w_k(n)|^2}{|W(f_0)|^2} \quad 2.6$$

There are different windowing functions; one is the Rectangular Window, where the input data sequence is multiplied by one, which is equivalent to using the input sequence as it is.

For the rectangular window function $\sum |w_k(n)|^2 = N$ where N is the number of input data samples. In the frequency domain, this window is equivalent to a sinc function where $|W(f_0)| = N$, so for the rectangular window function, the equivalent noise bandwidth is given by:

$$B_{\text{enbw}} = \frac{\sum |w_k(n)|^2}{|W(f_0)|^2} = \frac{N}{N^2} = \frac{1}{N}, \quad 2.7$$

This is equal to the DFT bin width.

Multi-dimensional Discrete Fourier Transforms

The DFT can be extended to multi-dimensional signals [1]; for radar systems, the two-dimensional DFT is the basis for the range-Doppler estimation for chirp sequence continuous FMCW radar, which will be explained later in this chapter. The 2-dimensional Discrete Fourier Transform (DFT) is a mathematical transformation used to analyze and process two-dimensional signals in the frequency domain.

For a 2-dimensional signal $f(n, m)$ of size $N \times M$ the 2D-DFT is defined as follows:

$$F(k, l) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f(n, m) e^{-j2\pi(\frac{k}{M}m + \frac{l}{N}n)}, \quad k = 0, 1, \dots, M-1, l = 0, 1, \dots, N-1 \quad 2.8$$

Where k and l are the sampling frequency indices while n and m are the sampling time indices.

The 2D FFT decomposes the signal into its constituent frequency components. It calculates these components' complex amplitudes and phases, representing the frequency content. The algorithm divides the signal into smaller blocks (bins), performs 1D FFT along each row and column, and combines the results to obtain the frequency domain representation.

2.2. Continuous Wave Radar

As mentioned earlier, continuous wave radar transmits a known signal continuously, and then this signal reflects off an object and is received continuously by the radar. In this part, the different modulation types for continuous wave radar are discussed in detail.

2.2.1. Unmodulated continuous wave radar

The transmitted signal is an unmodulated single tone with known frequency. If this signal is reflected from a moving target, then according to the Doppler effect, the frequency of the received signal increases if the target is moving toward the Radar and decreases if the target is moving away from the radar.

The frequency shift f_d resulting due to Doppler effect is calculated as follows [2]:

$$f_d = \frac{2f_0}{c} v_r \quad 2.9$$

Where v_r is the relative velocity of the moving target, f_0 is the center frequency of the transmitted signal, and c is the speed of light.

This frequency shift is calculated by mixing the transmitted and received waveforms and applying this mixing product to a low pass filter (LPF) followed by an ADC, then performing the FFT algorithm. By calculating the frequency shift, the relative velocity of a moving target can be calculated. The direction of this moving target cannot be estimated. This is because positive and negative Doppler shifts result in the same beat frequency measured after mixing the transmitted and received signals. To measure the direction of the target, in-phase and quadrature-phase (I/Q) signals should be used.

Using I/Q signals, the measured beat frequency will be the same for both signals. The difference will be

in the phase of the quadrature signal, which will differ according to the sign of the frequency shift, which will cause the quadrature to lead or lag the in-phase signal accordingly.

Using this unmodulated signal, only the velocity of the moving target can be calculated, but not its distance, as the distance will cause only a constant delay (phase shift) to the signal. This phase shift will not be useful due to the periodic nature of the phase shift as the phase shift of 1 period will equal the phase shift of 2 periods and so on, so this phase shift information will not be useful.

2.2.2. A frequency-modulated continuous wave radar

In its simplest form, the frequency of the transmitted signal increases linearly with time for a certain duration, T_c , before repeating, as shown in Fig. 2.1. The linear variation with time is usually referred to as a chirp.

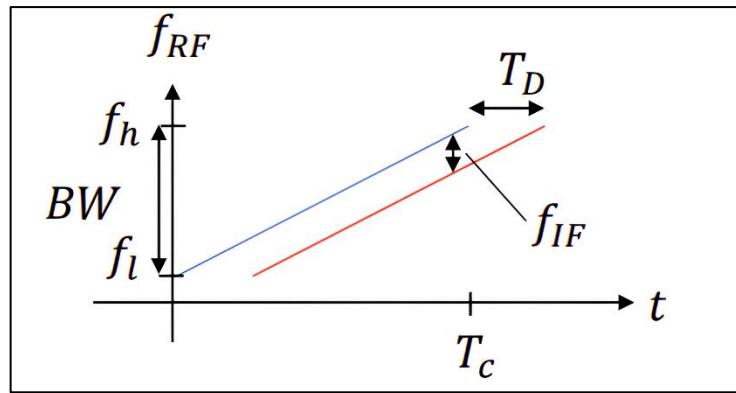


Fig. 2.1: FMCW transmit and receive signals.

For one chirp, the transmit frequency will be given by: $f(t) = f_l + (BW/T_c)t$, where f_l is the minimum transmitted frequency, BW is the bandwidth of the transmitted signal and T_c is the chirp duration.

For a stationary target, the reflected wave will consist of the transmitted waveform delayed by the two-way propagation delay as shown in Fig. 2.1. And the value of this delay will be given by:

$$T_{delay} = \frac{2R}{c}, \quad 2.10$$

Where R is the one-way distance and c is the speed of light.

Due to this delay, there will be a frequency shift between the transmitted and received signals. From Fig. 2.1, the phase shift will be calculated as follows:

$$f_{IF} = \frac{BW}{T_c} * T_{delay} = \frac{BW}{T_c} * \frac{2R}{c} \quad 2.11$$

If the target is moving, the Doppler shift will be added to f_{IF} and the total frequency shift becomes:

$$f_{IF} = \frac{BW}{T_c} * \frac{2R}{c} + \frac{2f_0}{c} v_r \quad 2.12$$

So, using FMCW only, we cannot measure target range R and radial velocity unambiguously. We need more (linearly independent) measurements.

2.2.3. Triangular FMCW chirp signal

This type will introduce more information that will help in decoupling the range and velocity measurements by using a triangular frequency modulation with up-chirp and down-chirp, as shown in Fig. 2.2.

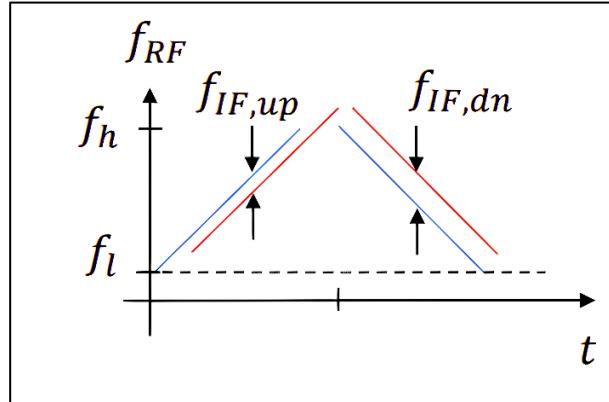


Fig. 2.2: Triangular FMCW transmit and receive signals.

After mixing, the two beat frequencies will be produced a low frequency and a high frequency, where:

$$f_{up} = \frac{BW}{T_c} * \frac{4R}{c} + \frac{2f_0}{c} v_r \quad 2.13$$

$$f_{down} = -\frac{BW}{T_c} * \frac{4R}{c} + \frac{2f_0}{c} v_r \quad 2.14$$

Then the range and velocity of the target can be obtained as follows:

$$R = \left(\frac{c T_c}{4B} \right) \left(\frac{f_{up} + f_{down}}{2} \right) \quad 2.15$$

$$v = \left(\frac{c}{2f_0} \right) \left(\frac{f_{up} - f_{down}}{2} \right) \quad 2.16$$

But this is for one target. If there are two targets, this method will not be enough as there will be 4 solutions,

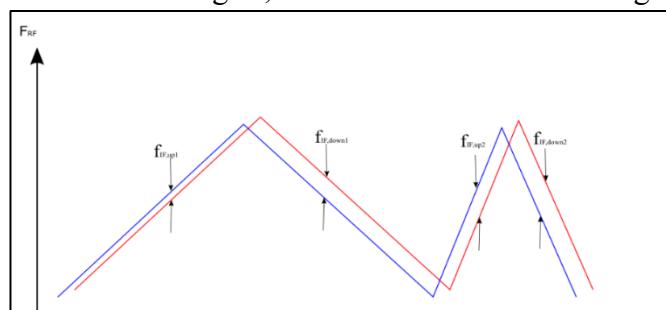


Fig. 2.3: Triangular FMCW with 2 different slopes.

two of which are real objects, and the other two are ghost targets that do not exist. So, we need more information to obtain the actual targets. One solution is to use 2 triangular modulations, each with a different slope, as shown in Fig. 2.3. This will lead to 2 solutions only, but what if we have 3 targets or more this will require more information and the solution of changing the slope will be very complex [1].

2.2.4. Multiple frequency shift keying waveform (MFSK)

Multiple Frequency Shift Keying waveform is used in automotive radar systems to enable the concurrent estimation of range and Doppler [3], even in scenarios involving multiple targets, while effectively avoiding the issue of ghost targets. The MFSK waveform is a combination of two linear FMCW waveforms with a fixed frequency offset. The two waveforms are combined in an interwind way, as shown in Fig. 2.4, but unlike the regular FMCW, the frequency is swept in discrete steps.

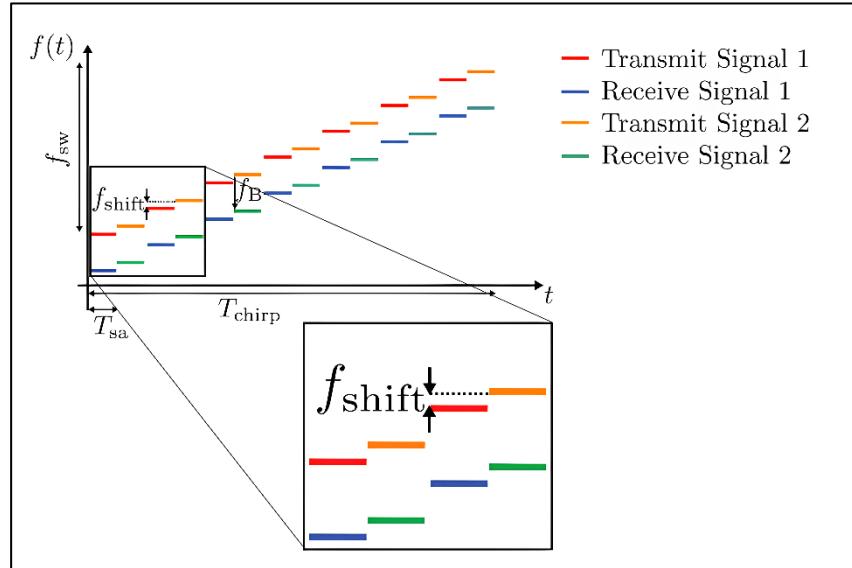


Fig. 2.4: MFSK transmit and receive signals.

The received signal is down-converted by the corresponding transmit frequency. This will result in two intertwined baseband signals:

$$s_1(t) = \exp(j2\pi(f_b \cdot t + \phi_1)) \quad 2.17$$

$$s_2(t) = \exp(j2\pi(f_b \cdot t + \phi_2)) \quad 2.18$$

The frequency shift between the two interwind signals is chosen very small compared to the carrier frequency such that the two signals will have the same beat frequency, but the constant phase term for each signal will be different. To extract the beat frequency and phase information, the baseband signals $s_1(t)$ and $s_2(t)$ are sampled separately by the sampling interval T_{sa} , one sample per frequency step, then the produced samples for each signal are applied to 1D-FFT separately as follows:

$$S_1(m) = \sum_{k=0}^{K-1} s_1(k) \cdot \exp(-j2\pi \frac{k \cdot m}{K}) \quad 2.19$$

$$S_2(m) = \sum_{k=0}^{K-1} s_2(k) \cdot \exp(-j2\pi \frac{k \cdot m}{K}) \quad 2.20$$

Where K is the number of samples in one chirp, k is the sample index of the time discrete baseband signal, and m is the beat frequency index. Each of these discrete frequency domain signals consists of K complex values indicating the amplitude and the phase of the signal at a given beat frequency. This beat frequency should be the same for the two signals but the phase will be different:

$$\phi_1(m) = \arg(S_1(m)) \quad 2.21$$

$$\phi_2(m) = \arg(S_2(m)) \quad 2.22$$

And using this beat frequency and the phase difference, the range and speed can be estimated [5]:

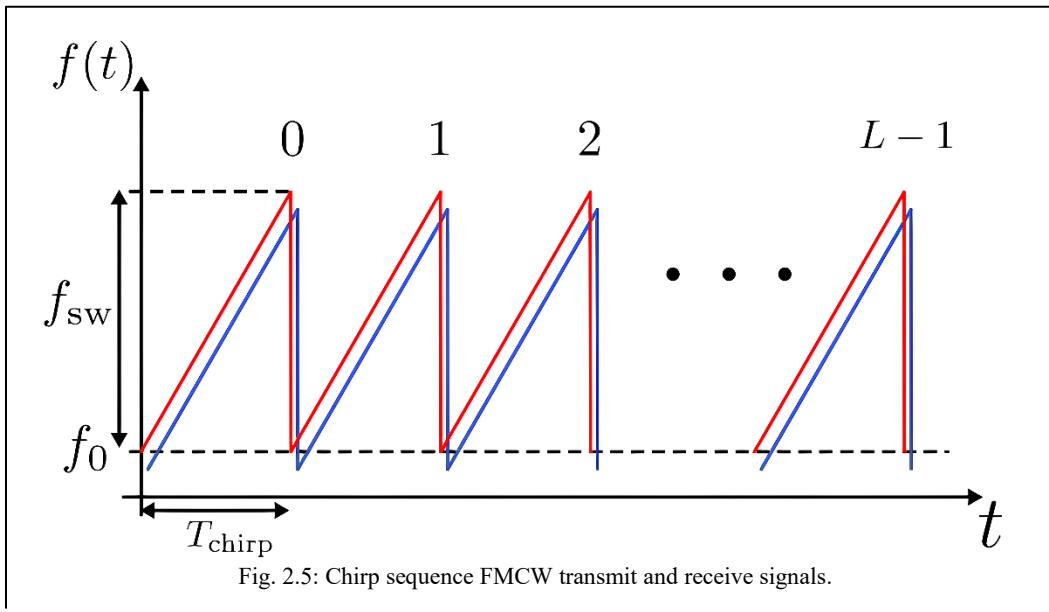
$$f_b = \frac{BW}{T_c} * \frac{2R}{c} + \frac{2f_0}{c} v_r \quad 2.23$$

$$\Delta\phi(m) = -\frac{4\pi T_s v_r}{\lambda} + \frac{4\pi f_{offset} R}{c} \quad 2.24$$

So using the MFSK waveform, the velocity and range can be obtained in an unambiguous way for multiple targets, but as this technique depends on phase measurements, hence it will suffer from low estimation accuracy.

2.2.5. Chirp sequence FMCW

To avoid ghost targets, we use a sawtooth waveform instead of a triangular waveform, and instead of sending one chirp, send L chirps, each with a small chirp duration, T_c , as shown in Fig. 2.5, to be able to resolve the range and velocity for multiple targets [4].



For this type, the transmitted signal is expressed as follows:

$$S_T(t) = S_T(ts, n) = \cos\left(2\pi\left(f_c + \frac{B}{2T}ts\right)ts + \phi_0\right), t = nT_c + ts, 0 < ts < T_c. \quad 2.25$$

Where B is the bandwidth, T_c is the chirp duration and t_s is the time from the start of the $(n+1)$ th chirp. The received echo signal is an attenuated and delayed version of the transmitted one and can be expressed as follows:

$$S_R(ts, n) = \alpha \cos\left(2\pi\left(f_c + \frac{B}{2T}(ts - \tau)\right)(ts - \tau) + \phi_0\right). \quad 2.26$$

Where α is the attenuation and τ is the transmission delay. For a moving object, the transmission can be expressed as follows:

$$\tau = \frac{2(R + v_r t)}{c} = \frac{2(R + v_r ts + v_r T_c n)}{c}. \quad 2.27$$

Substituting in Eqn. 2.26 then the received after mixing with the transmitted signal and filtering with a low pass filter:

$$S_b(ts, n) = \cos\left(2\pi\left(\frac{B}{2T}\left(\frac{2R}{c}\right)ts + \frac{2f_c v T}{c} \cdot n + \frac{2f_c R}{c}\right)\right) \quad 2.28$$

If this baseband signal is sampled by an ADC with a sampling rate of f_s , then $t_s = m/f_s$ where m is the sampling index, hence that sampled signal will be given by:

$$S_b(m, n) = \cos\left(2\pi\left(\frac{B}{2T}\left(\frac{2R}{c}\right)\frac{m}{f_s} + \frac{2f_c v T}{c} \cdot n + \frac{2f_c R}{c}\right)\right). \quad 2.29$$

From Eqn. 2.29, there is the term multiplied by m depends only on the range and represents the sampled

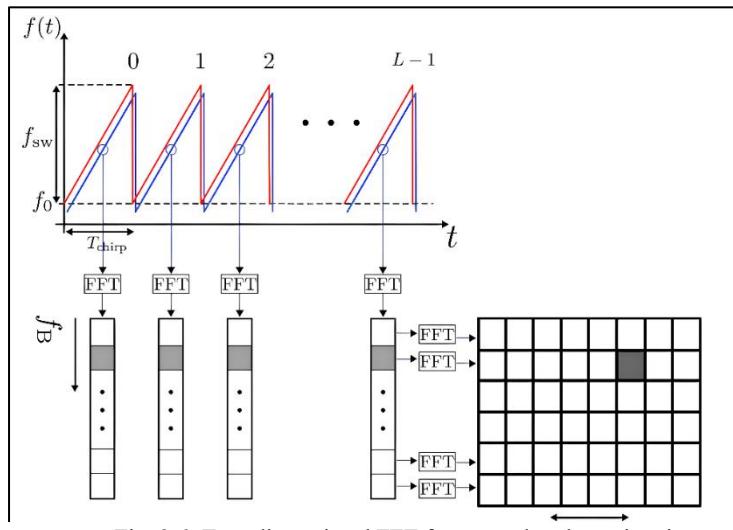


Fig. 2.6: Two-dimensional FFT for range doppler estimation.

signal in one chirp (fast time), and the term multiplied by n depends only on the velocity and it is constant for each chirp and varies from chirp to chirp (slow time), so by applying 2-D FFT on $S_b(m, n)$ then range and velocity can be estimated. By taking the FFT along the fast time for each chirp, the range can be estimated, then the output of this FFT is then applied to another FFT along the slow time so the velocity can be estimated. This is shown in Fig. 2.6.

This type also only needs one receive channel (in-phase component). For this type, the chirp duration T_c is chosen very small such that the measured beat frequency will be dominated by the range and less influenced by the velocity such that the beat frequency will always be negative, so there is no need for the Q-phase component to resolve the sign of the beat frequency which reduces hardware complexity [3].

2.2.6. Range and velocity resolution for FMCW radar systems

Two important parameters for radar systems are range and velocity resolution. Range resolution is the minimum distance δR such that a target at a distance R from the radar and another at $R + \delta R$ appear as two distinguishable targets for the radar. Similarly, the velocity resolution is the minimum allowed difference in velocities of two targets that would make these two targets to be distinguishable.

As the resolution bandwidth equals $1/T_c$, where T_c is the chirp duration, so for two frequency shifts to be distinguishable, the difference between them must be greater than $1/T_c$. This will limit the ability to differentiate between 2 targets' ranges or speeds as follows:

$$\Delta f_{Range} = \frac{BW}{T_c} * \frac{2\Delta R}{c} = \frac{1}{T_c} \rightarrow \Delta R = \frac{c}{2BW} \quad 2.30$$

$$\Delta f_d = \frac{2f_0}{c} \Delta v_r = \frac{1}{T_c} \rightarrow \Delta v_r = \frac{c}{2f_0 T_c} = \frac{\lambda}{2T_c} \quad 2.31$$

But for chirp sequence FMCW, the velocity resolution is given by:

$$\Delta v_r = \frac{c}{2f_0 n T_c} = \frac{\lambda}{2n T_c} \quad 2.32$$

Where n is the number of chirps (as the velocity is estimated by applying FFT along the slow time).

References

- [1]. J. Gamba, in *Radar Signal Processing for autonomous driving*, Singapore: Springer, 2020
- [2]. “Radar Systems - Doppler effect,” Online Courses and eBooks Library,
https://www.tutorialspoint.com/radar_systems/radar_systems_doppler_effect.htm.
- [3]. M. Kronauge and H. Rohling, “New chirp sequence radar waveform,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 4, pp. 2870–2877, 2014. doi:10.1109/taes.2014.120813
- [4]. T. Ma *et al.*, “A CMOS 76–81GHz 2-TX 3-RX FMCW radar transceiver based on mixed-mode PLL Chirp Generator,” *IEEE Journal of Solid-State Circuits*, vol. 55, no. 2, pp. 233–248, 2020. doi:10.1109/jssc.2019.2950184
- [5]. “Simultaneous range and speed estimation using MFSK waveform,” Simultaneous Range and Speed Estimation Using MFSK Waveform - MATLAB & Simulink,
https://www.mathworks.com/help/radar/ug/simultaneous-range-and-speed-estimation-using-mfsk-waveform.html?s_eid=PSM_15028.

3

DIRECTION OF ARRIVAL ESTIMATION

We have seen how automotive radar systems can simultaneously determine targets' range and velocity through various modulation schemes and signal-processing algorithms. However, these two properties are not, by any means, sufficient for the automotive radar to be able to perform its overall functionality and safety in autonomous driving and advanced driving assistance systems (ADAS). The system also needs to be able to determine the direction of an object, the direction of arrival (DOA), both horizontally (azimuth angle) and vertically (elevation angle). Accurate DOA estimation is of paramount importance in automotive radar systems. It enables precise object detection, localization, and tracking, enhances situational awareness, and provides critical input for decision-making algorithms in autonomous driving and ADAS. By leveraging the information obtained from DOA estimation, radar systems contribute to safer, more efficient, and reliable transportation, making significant strides towards the realization of fully autonomous vehicles and reducing the occurrence of accidents on the road.

The fact that there will be multiple and an unknown number of objects in the radar's field of view, causing an unknown number of received signals impinging on the receiver simultaneously, often complicates the DOA estimation. Additionally, the received signals are corrupted by additive noise and clutter, unwanted echoes, which adds even more complexity to the process.

DOA estimation relies fundamentally on the utilization of multiple antenna elements (antenna arrays), so in this chapter, we will first quickly revisit the fundamentals of antenna arrays, especially uniform linear arrays (ULA), then discuss some of the used DOA estimation algorithms and have a look about MIMO radar concepts and virtual antenna arrays.

3.1. Antenna Arrays Fundamentals

The main reason why antenna arrays become very instrumental in estimating the DOA is the fact that we can change the array radiation pattern by either changing the elements' configuration in space or the excitation phase and amplitude of individual elements. This concept is known as Beamforming and can be done in analog or digital domains. The difference between both methods will be discussed in the next section.

To understand how this works, consider N isotropic antenna elements distributed in a certain configuration in space, as shown in Fig. 3.1, with \mathbf{r}_n being the position vector to each element. For simplicity, assume an incident linearly polarized plane wave with a certain wave number $k = 2\pi/\lambda$.

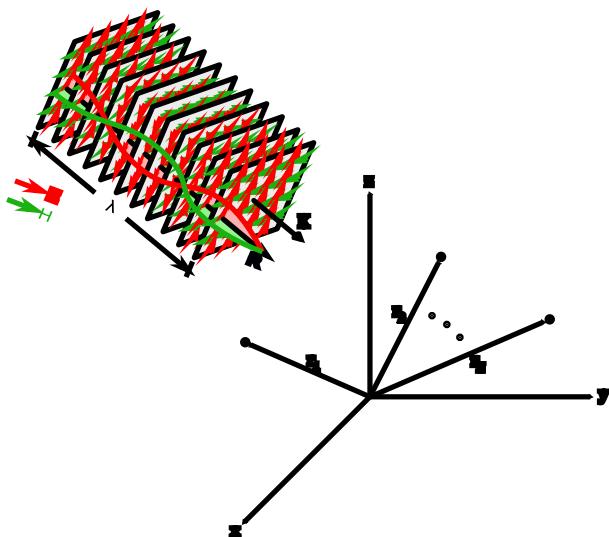


Fig. 3.1: A linearly polarized plane wave incident on an antenna array

Under these assumptions, it follows that the electric field has the same amplitude of oscillation everywhere in space, but the phase is position dependent. The field phasor at any point in space can be given by the following equation. The details of the foregoing discussion are found in [1].

$$E(\mathbf{r}) = E_0 e^{-j\mathbf{k} \cdot \mathbf{r}} \quad 3.1$$

Where $\mathbf{k} = k\hat{\mathbf{u}}$ is the wavevector with $\hat{\mathbf{u}}$ being a unit vector in the direction of incidence, and \mathbf{r} is the position vector to the point.

Since the elements are distributed at different locations in space, elements receive signals with different phases. If X_n is the signal received from the n^{th} element, then we have:

$$X_n = x_0 e^{-j\mathbf{k} \cdot \mathbf{r}_n} \quad 3.2$$

Multiplying the signal received from the n^{th} element by a complex coefficient w_n , adjusting the amplitude and phase of the received signal from that element, then the total output signal of the array Y can be written as:

$$Y = \sum_{n=1}^N w_n * x_0 e^{-j\mathbf{k} \cdot \mathbf{r}_n} \quad 3.3$$

Eqn. 3.3 shows that the output Y depends on the complex weights w_n , the wave direction of incidence, and the configuration of the elements in space. If the weights w_n are equal to 1, the output signal still depends on the wave direction of incidence, which means that the array has spatial filtering properties even though all of its elements are isotropic. In other words, the overall array is directive, and this directivity can be manipulated through complex weights, which is the main idea behind Beamforming.

Since antennas are reciprocal, the same idea holds when using the array for transmission rather than reception; by changing the excitation phase and amplitude of individual elements, we can control the direction of array transmission.

3.1.1. Array Factor

The principle of pattern multiplication for arrays of identical elements, shown in Fig. 3.2, states that the far-zone radiated field of the array is equal to the product of the field of a single element placed at a selected reference point, usually the origin, and a factor that depends on the array geometry called the array factor AF. That is [2],

$$E_{total} = E_{element} * \overbrace{\sum_{i=1}^N I_n e^{j k \hat{r} \cdot r_n}}^{AF(\hat{r})} \quad 3.4$$

where I_n is the complex excitation coefficient for the n^{th} element, \mathbf{r}_n is the position vector of the n^{th} element, and \hat{r} is the direction unit vector, a function in θ and ϕ . This expression is derived for the transmitting mode of the array. A similar expression is derived for the receiving mode but with a negative sign in the exponential factors, as can be shown from Eqn. 3.3 . Once again, we notice that the array factor depends on the complex excitation coefficient of each element, which means that we can control the array radiation pattern just by changing the excitation of individual elements.

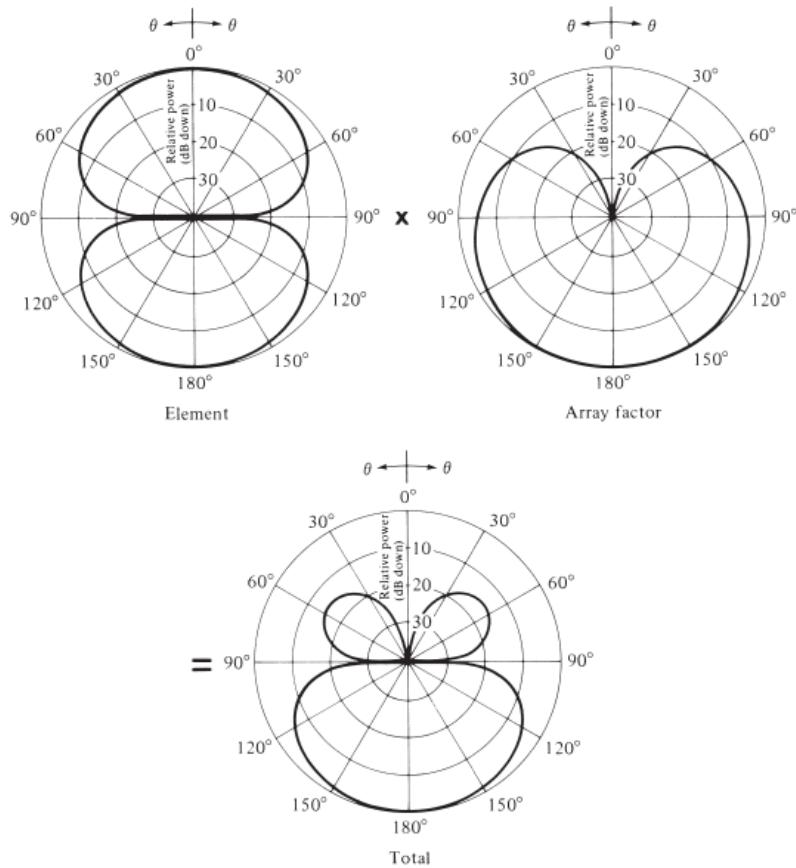


Fig. 3.2: The principle of pattern multiplication for an array of identical elements

3.1.2. Uniform Linear Array (ULA)

Consider a linear antenna array along the z axis consisting of N elements with spacing distance d between every two elements, as shown in Fig. 3.3.

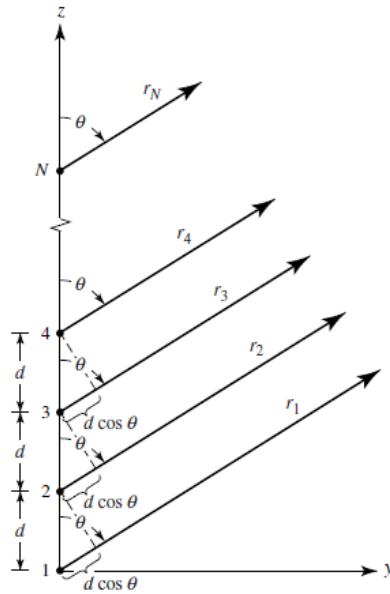


Fig. 3.3: An N -element linear antenna array along the Z -axis

Assuming that all elements have identical amplitudes but each succeeding element has β progressive phase lead current excitation relative to the preceding one, then we can write the array factor as follows,

$$AF(\theta) = \sum_{i=1}^N e^{j(n-1)\beta} e^{jk(n-1)d\cos(\theta)} = e^{j(N-1)\psi} \left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{\psi}{2}\right)} \right] \quad 3.5$$

Where $\psi = kdcos(\theta) + \beta$. Usually, the absolute of the array factor normalized to its maximum value is plotted, which is given by

$$(AF(\theta))_n = \frac{1}{N} \left[\frac{\sin\left(\frac{N}{2}\psi\right)}{\sin\left(\frac{\psi}{2}\right)} \right] \quad 3.6$$

Due to symmetry, it is predictable to have the array factor independent of the angle ϕ , which means that the pattern is always symmetric about the z -axis. When $\beta = 0$, we will also have symmetry about the axis normal to the array. The angles at which the previous expression attains its maximum value are called the grating lobe angle directions. They are essential to know because they represent the main directions of the array transmission. These angles are given by the following expression.

$$\theta_m = \cos^{-1} \left((\pm 2n\pi - \beta) \frac{\lambda}{2\pi d} \right), n = 0, 1, 2, \dots \quad 3.7$$

The angles at which the AF expression drops to zero are called the null angles, directions where the array doesn't radiate at all, and are given by the following expression.

$$\theta_n = \cos^{-1} \left(\left(\pm \frac{2n\pi}{N} - \beta \right) \frac{\lambda}{2\pi d} \right), n = 0, 1, 2, \dots \quad 3.8$$

$n \neq N, 2N, 3N, \dots$

It is desirable to have only one main beam in the visible observation angle range $(-\pi, \pi)$. To ensure that, we need the previous equation to have a solution only for $n = 0$. For other values of n , the argument of the arccosine function should have an absolute value greater than one. The condition for only one main beam in the visible range then becomes,

$$\left| (\pm 2n\pi - \beta) \frac{\lambda}{2\pi d} \right| > 1 \text{ for } n = 1, 2, 3, \dots \quad 3.9$$

In the case of $\beta = 0$, the previous condition reduces to,

$$d < n\lambda \text{ for } n = 1, 2, 3, \dots \quad 3.10$$

That is, when there is no progressive phase shift between the elements, the separation distance can't be a multiple of the wavelength; otherwise, we will have more than one main beam in the visible angle range. Fig. 3.4 shows the radiation pattern of a linear array along the z-axis with $\beta = 0$ and $d = \lambda$.

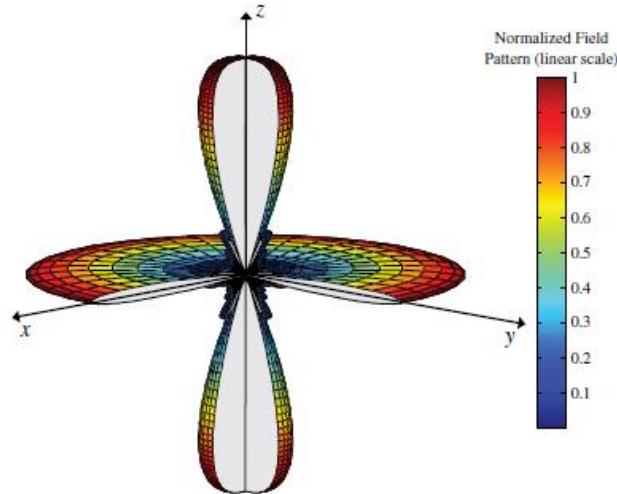


Fig. 3.4: The radiation pattern of a linear array along the z-axis with $\beta = 0$ and $d = \lambda$.

It is obvious how the radiation pattern has two main lobes, one at $\theta = 0$ and the other at $\theta = 90^\circ$.

If the condition in Eqn. 3.9 is satisfied, the main beam direction is a function in the progressive phase shift β and is obtained by putting $n = 0$ in Eqn. 3.8.

$$\theta_m = \cos^{-1} \left(\frac{\lambda\beta}{2\pi d} \right), n = 0, 1, 2, \dots \quad 3.11$$

If $\beta = 0 \rightarrow \theta_m = 90^\circ$, which means that the linear array has a maximum radiation direction normal to the array axis. This is known as a broadside array. The following Figure shows a broadside linear array along the z-axis with $\beta = 0$ and $d = \lambda/4$.

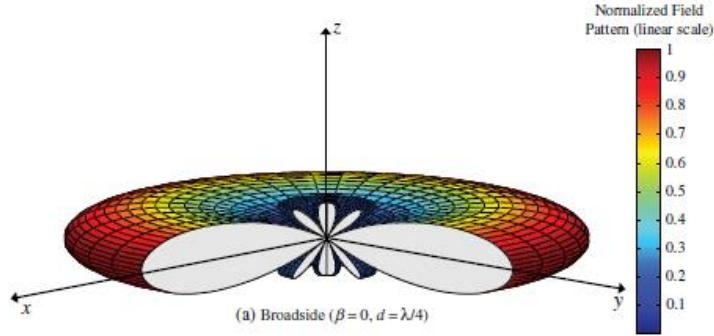


Fig. 3.5: A broadside linear array along the z-axis with $\beta = 0$ and $d = \lambda/4$.

Another vital thing to notice is that the main lobe beamwidth depends on the separation distance and the number of array elements through the following equation.

$$\Delta\theta \cong \frac{2\lambda}{Nd} \quad 3.12$$

If we want a narrower main beam, we can either increase the number of array elements or the separation distance. Increasing the number of elements is a more viable option because the separation distance is restricted. After all, we don't need more than one main beam, as illustrated before. Having a narrower beam width means better angle resolution for the DOA estimation.

Fig. 3.6 shows the array factor for two linear arrays with different numbers of elements, both having main beams along their axes. It is clear that the array with more elements is more directive but has more sidelobes than the other.

Fig. 3.7 shows how the direction of the main beam is being adjusted by changing the value of β .

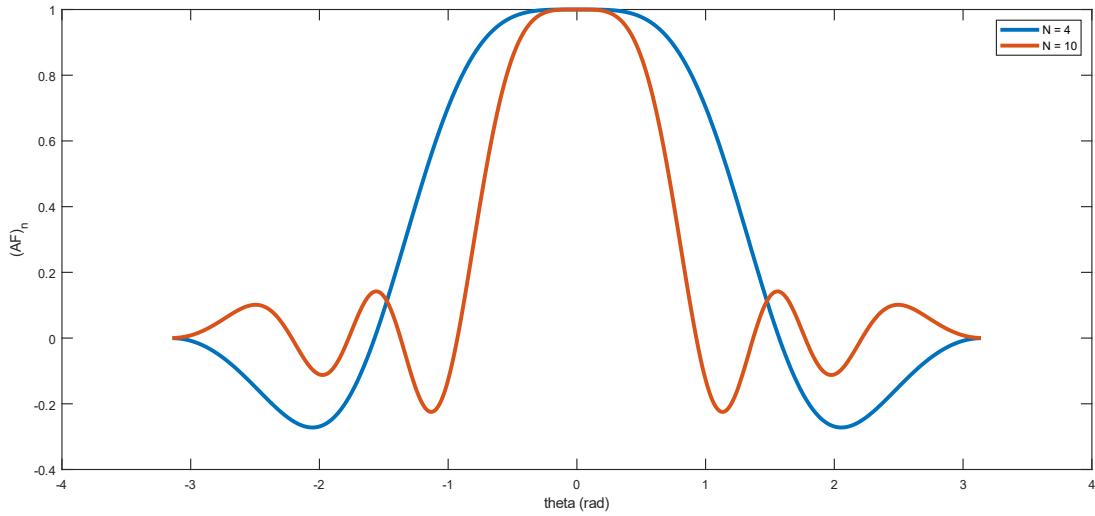


Fig. 3.6: The array factor for two linear arrays with $N = 4$ and $N = 10$, both having main beams along their axes.

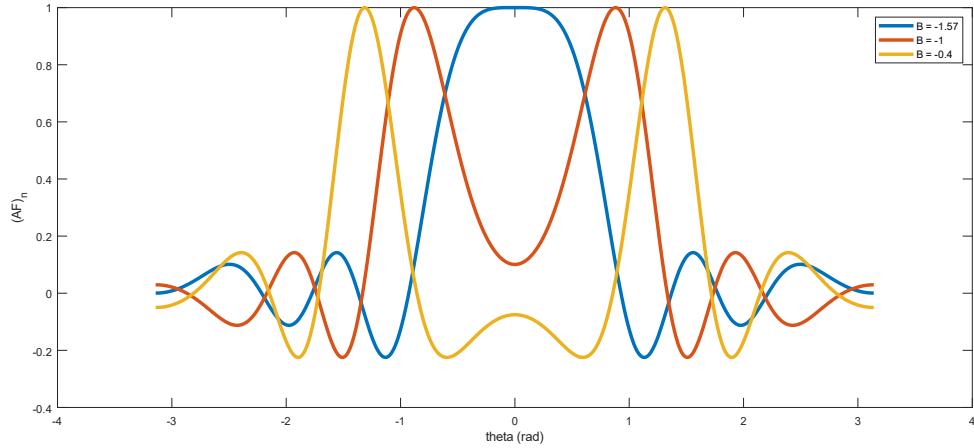


Fig. 3.7: adjusting the main beam direction by changing the value of β .

3.1.3. Linear Array in Rx mode

Consider a linear array of two antenna elements with spacing d meters, and there is an incident plane wave with an angle θ measured from the normal axis, as shown in Fig. 3.8.

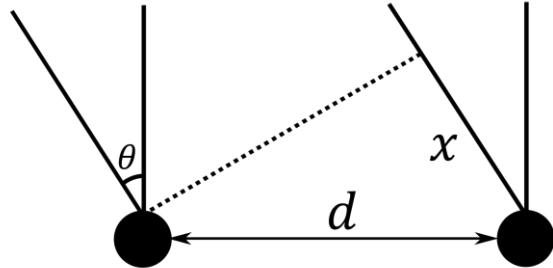


Fig. 3.8: two antenna elements separated by a distance d .

The wave has to travel an additional distance of $x = d\sin(\theta)$ to reach the second antenna, which causes a phase difference between the two elements given by,

$$\Delta\phi = \omega * \Delta t = \omega * \frac{x}{c} = \frac{2\pi d}{\lambda} \sin(\theta) \quad 3.13$$

For the general case when we have N elements in the array, the phase difference at the n^{th} antenna with reference to the first element is given by,

$$\Delta\phi = \frac{2\pi(n - 1)d}{\lambda} \sin(\theta) \quad 3.14$$

In practice, phase measurements are restricted to the range $(-\pi, \pi)$ because phase shifts that are multiples of $\pm 2\pi$ outside that range are indistinguishable from shifts that are within that range. This is referred to as phase wrapping. For the N -antenna array, the phase-wrap angle can be calculated through the following expression.

$$\frac{2\pi(N - 1)d}{\lambda} \sin(\theta) = \pi \rightarrow \theta_{wrap} = \sin^{-1}\left(\frac{\lambda}{2(N - 1)d}\right) \quad 3.15$$

Angles of arrival beyond the phase wrap angle will cause a wrong direction of arrival estimation because it will cause a phase shift beyond the interval $(-\pi, \pi)$, which is indistinguishable from the phase shift caused by a smaller angle of arrival. The phase wrap angle can be increased by decreasing the number of elements in the array, but this will, in return, cause less angular resolution because the main beam width will increase, as discussed before.

3.1.4. Analog VS digital beamforming

Beamforming can be done either in the analog or digital domains. Analog beamforming uses only one receive chain and a phased antenna array. A phased array means that the signal from each antenna experiences different gain and phase shifts using an analog phase shifter and a variable amplifier in the path of each antenna. These signals are then summed together using a power combiner and fed to the single receive chain. As explained in previous sections, the array will have an electronically steerable main beam that can be directed by choosing the suitable gain and phase for each antenna element.

On the other hand, digital beamforming utilizes a receiver chain and an analog-to-digital converter (ADC) for each antenna element. Phase and gain variation is done in the digital domain by multiplying the sampled signals by the suitable complex coefficients. The main advantage of digital beamforming is the ability to process signals from multiple virtual beams simultaneously. To form a beam in a specific direction, we multiply the sampled signals by certain coefficients, and if we want a beam in another direction, we can use the same sampled signals and just change the multiplication coefficients. The difference between the used hardware for analog and digital beamforming is shown in Fig. 3.9 and Fig. 3.10, respectively.

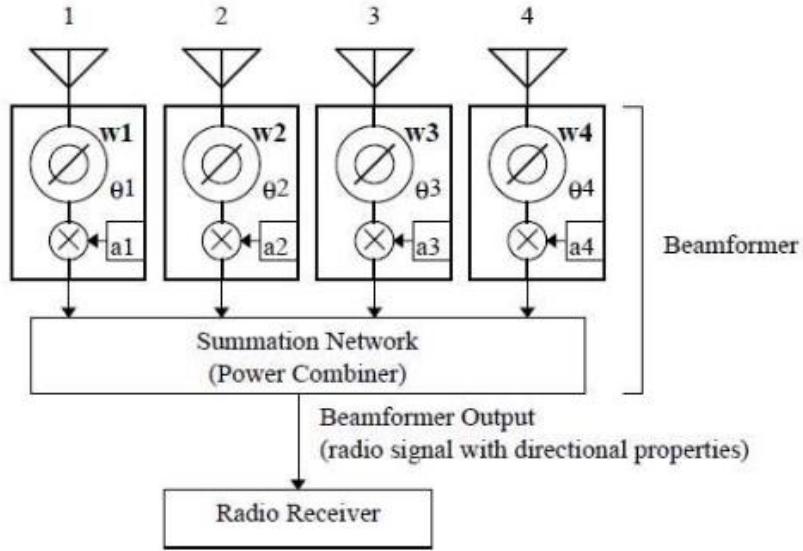


Fig. 3.9: Analog Beamformer.

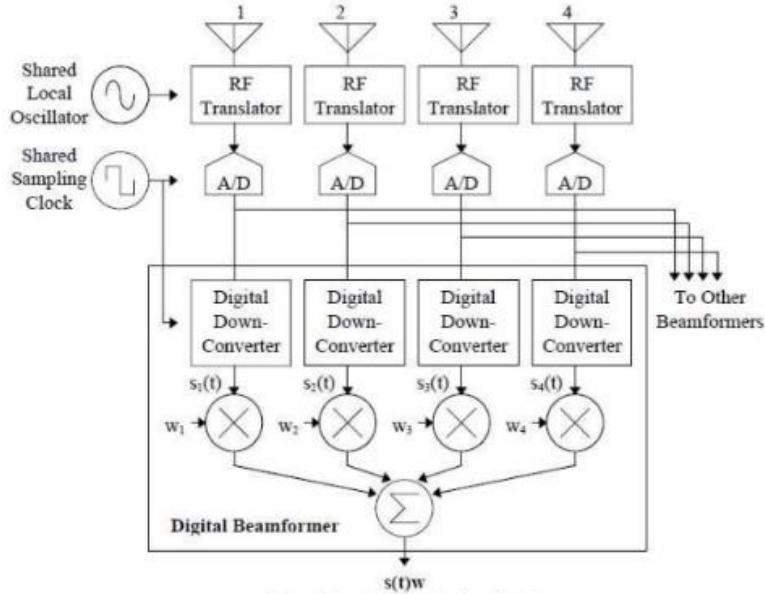


Fig. 3.10: Digital Beamformer

3.2. DOA Estimation Methods

This section focuses mainly on the methods used for one-dimensional DOA estimation. However, we will briefly discuss how these ideas are extended for the more practical case of 2D DOA estimation. Almost all of the 1D DOA estimation methods utilize a linear antenna array and process the signal captured from it in a certain way to extract the angles of the targets; that's why we need to derive a mathematical model for the signal from a linear antenna array before we discuss these algorithms.

3.2.1. Signal Model

Consider a linear antenna array of M elements onto which signals from K targets are received. the i^{th} target is at an angle θ_i measured from the normal to the array axis. Fig. 3.11 shows the incident signal from one of these targets. The received signal from the i^{th} element is $x_i(t)$.

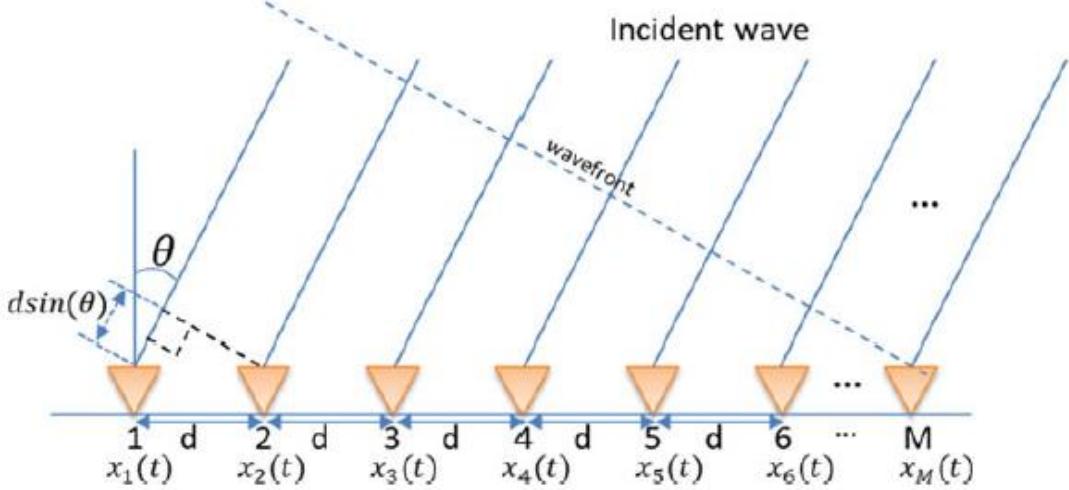


Fig. 3.11: A linear antenna array of M elements with an incident wave with angle θ .

If there is only one target at an angle θ and the first element receives a signal $S(t)$, then the i^{th} element receives a signal given by [Section 3.1.3]

$$x_i(t) = e^{-j \frac{2\pi d(i-1) \sin(\theta)}{\lambda}} * S(t) \quad 3.16$$

Which can be rewritten in the matrix notation as follows

$$\mathbf{X}(t) = \mathbf{a}(\theta)S(t) \quad 3.17$$

Where $\mathbf{X}(t) = [x_1(t), \dots, x_M(t)]^T$ is an $M \times 1$ received signals vector, $S(t)$ is the received signal from the first antenna and $\mathbf{a}(\theta)$ is the $M \times 1$ steering vector, which is given by

$$\mathbf{a}(\theta) = \left[1, e^{-j \frac{2\pi d \sin(\theta)}{\lambda}}, \dots, e^{-j \frac{2\pi d(M-1) \sin(\theta)}{\lambda}} \right]^T \quad 3.18$$

Now for the general case when there are K targets, the received signals vector can be expressed as follows.

$$\mathbf{X}(t) = \mathbf{A}(\theta)\mathbf{S}(t) + \mathbf{N}(t) \quad 3.19$$

Where $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$ is the $M \times K$ steering matrix whose columns are the steering vectors for different targets, $\mathbf{S}(t) = [S_1(t), \dots, S_K(t)]^T$ is the $K \times 1$ source signal vectors as received by the first element and $\mathbf{N}(t) = [n_1(t), \dots, n_M(t)]^T$ is the $M \times 1$ noise vector, assuming there is an AWGN added to the received signal by each antenna.

If the signal from each antenna element $x_i(t)$ is multiplied by some complex coefficient w_i^* , where the

asterisk (*) denotes complex conjugation, then the weighted linear combination of sensor outputs can be written as follows:

$$y(t) = \sum_{i=1}^M w_i^* x_i(t) = [w_1^*, \dots, w_M^*] [x_1(t), \dots, x_M(t)]^T = \mathbf{w}^H \mathbf{X} \quad 3.20$$

Where H denotes the Hermitian transpose, transposing with conjugation of each element. The output power, $P(\mathbf{w})$, of the signal $y(t)$ can be expressed as follows:

$$P(\mathbf{w}) = E[|y(t)^2|] = E[yy^H] = \mathbf{w}^H E[\mathbf{X}\mathbf{X}^H] \mathbf{w} = \mathbf{w}^H \mathbf{R} \mathbf{w} \quad 3.21$$

Where $\mathbf{R} = E[\mathbf{X}\mathbf{X}^H]$ is the $M \times M$ input signal covariance matrix whose elements represent the auto and cross-correlations between different signals from different antenna elements. An important thing to notice here is that we have expressed signals from different elements, x_i , and the total output signal, y , in terms of the continuous-time variable t but in practice, these signals are sampled for further processing, meaning that x_i and y are generally complex vectors in the case of using I/Q downconverters.

3.2.2. Digital Beamforming (DBF) methods

Different DBF algorithms rely on the same basic principle; they choose a certain weights vector \mathbf{w} that maximizes the received signal power $P(\mathbf{w})$ in a certain direction, and if power peaks are detected from that direction, a target is detected at that specific angle. There are multiple DBF algorithms, some of which are to be discussed.

Bartlett Beamformer

Bartlett Beamformer [3] chooses the weight vector as follows:

$$\mathbf{w}^H = \mathbf{a}^H(\theta) \rightarrow P(\theta) = \mathbf{a}^H(\theta) \mathbf{R} \mathbf{a}(\theta) \quad 3.22$$

This method has low computational complexity as it only requires the estimation of the input signal covariance matrix \mathbf{R} . The locations of the targets can be easily found by finding the angles at which the previous expression is maximized. The choice of that specific weight vector is straightforward when we rewrite Eqn. 3.19 as follows:

$$\mathbf{X}(t) = \mathbf{A}(\theta) \mathbf{S}(t) + \mathbf{N}(t) = \mathbf{a}(\theta_1) S_1(t) + \dots + \mathbf{a}(\theta_k) S_k(t) + \mathbf{N}(t) \quad 3.23$$

It is obvious that if we want to maximize the signal from the direction of the i^{th} target, we just multiply the received signals vector $\mathbf{X}(t)$ by the conjugate transpose of its steering vector $\mathbf{a}^H(\theta_i)$.

Another choice of the optimum weight vector [4], which has no advantages over the previous method, is:

$$\mathbf{w} = \frac{\mathbf{a}(\theta)}{\sqrt{\mathbf{a}^H(\theta) \mathbf{a}(\theta)}} \quad 3.24$$

Which gives the following expression of the received signal power as a function of the angle θ :

$$P(\theta) = \frac{\mathbf{a}^H(\theta)\mathbf{R}\mathbf{a}(\theta)}{\mathbf{a}^H(\theta)\mathbf{a}(\theta)} = \frac{\mathbf{a}^H(\theta)\mathbf{R}\mathbf{a}(\theta)}{M} \quad 3.25$$

This method of DOA estimation suffers from low angle resolution, unable to resolve closely spaced targets. However, it can be used as a first step to narrow the target range for high-resolution methods. The resolution of this method is limited by the 3-dB beamwidth of the main array beam, which is inversely proportional to the number of array elements.

Capon Beamformer

Capon Beamformer aims to maintain constant gain for signals arriving from a specific direction while giving smaller weight to noise [4]. The weights are chosen to minimize the noise by minimizing the total output power of the array while preserving the desired signal in a certain direction. The optimization problem can be formulated as follows:

$$\min_{\mathbf{w}}(P(\mathbf{w})) \text{ subject to } \mathbf{w}^H \mathbf{a}(\theta) = 1 \quad 3.26$$

The solution to this optimization problem gives the optimum weight vector.

$$\mathbf{w}_{opt} = \frac{\mathbf{R}^{-1} \mathbf{a}(\theta)}{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta)} \quad 3.27$$

Substituting with the optimum weight vector into Eq. 3.21 gives the following expression for the capon power spectrum.

$$P_{Capon}(\theta) = \frac{1}{\mathbf{a}^H(\theta) \mathbf{R}^{-1} \mathbf{a}(\theta)} \quad 3.28$$

Capon Beamformer achieves better performance when it comes to resolving closely separated objects than the Bartlett Beamformer. Additionally, the main computation in this method is the determination of the inverse covariance matrix, which makes it very attractive for situations when high resolution is required without so much computational overhead. Both methods don't require prior knowledge of the number of signal sources and can estimate the DOA and power of the signals simultaneously. Fig. 3.12 and Fig. 3.13 show the difference between Bartlett and Capon Beamformers in resolving two objects at $\pm 10^\circ$ using a five-element ULA antenna, respectively.

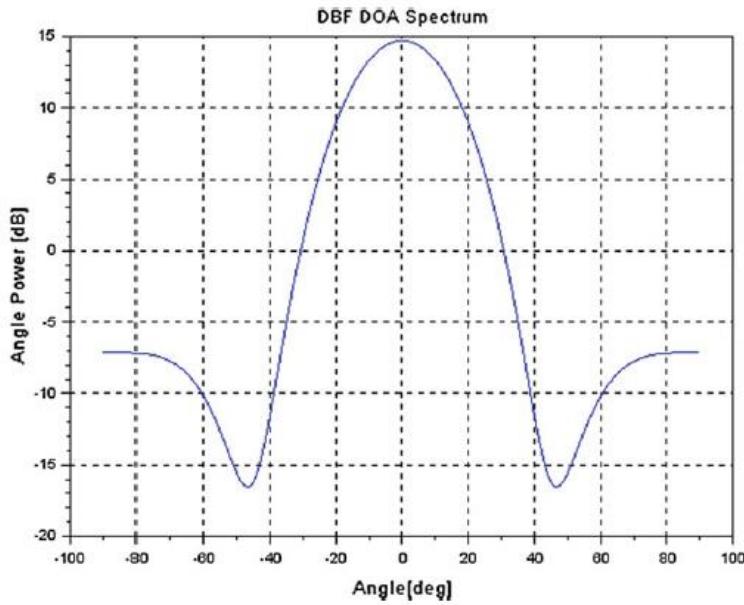


Fig. 3.12: Bartlett Beamformer spectrum with two objects at $\pm 10^\circ$ using a five-element ULA antenna.

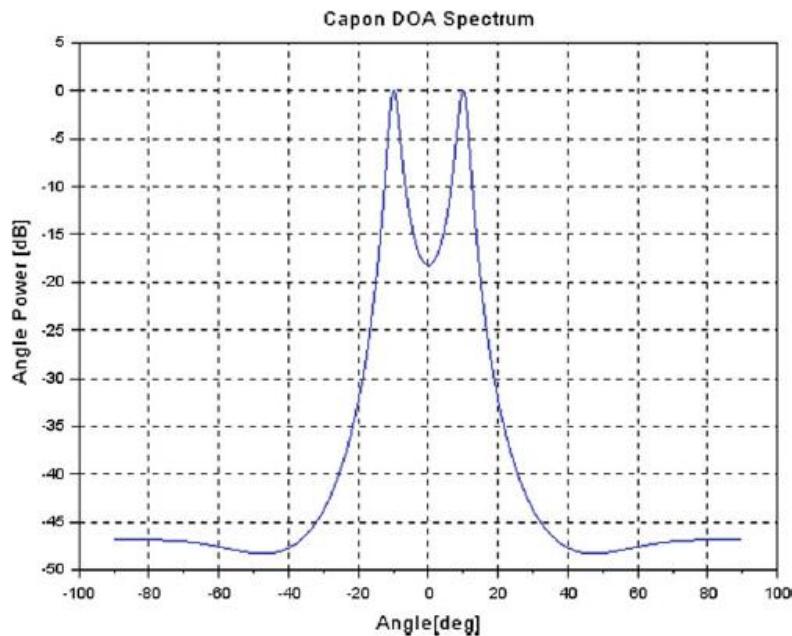


Fig. 3.13: Capon Beamformer spectrum with two objects at $\pm 10^\circ$ using a five-element ULA antenna.

It is clear how the Capon method is capable of resolving such close targets while the ordinary DBF method detects both targets as a single peak.

Angle FFT method

To understand how this method works, let's take the chirp sequence FMCW radar as an example. For the sake of generality, assume that we are transmitting a complex signal and using I/Q demodulators at the receivers. As discussed in the previous chapter, if the chirp duration is short enough, both the transmitter and the receiver are simplified, and we only transmit and receive real signals [5]. The complex transmitted signal can be written as [6]

$$S_T(t) = \exp \left\{ j2\pi \left(f_c + \frac{B}{2T} t \right) t \right\}, \quad 0 < t < T \quad 3.29$$

Where f_c is the carrier frequency, B is the signal bandwidth, and T is the chirp duration. The received signal from the first antenna element is a superposition of all the signals reflected off the K targets and is conjugately mixed with the transmitted signal to produce a low-frequency beat signal. The complex beat signal is sampled with a sampling frequency f_s . N Samples over the fast time and for P successive chirps are put into columns to form the two-dimensional $N \times P$ complex samples matrix as shown in Fig. 3.14.

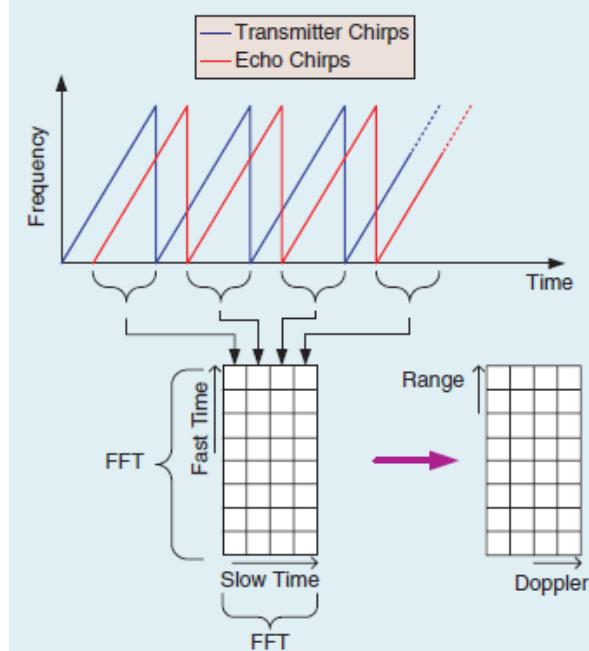


Fig. 3.14: Forming the Range-doppler map from multiple chirps data

The signal received from the first antenna element, after downconversion and sampling, is given by [6]

$$x_1(n, p) = \sum_{i=1}^K S_i(n, p) = \sum_{i=1}^K \alpha_i \exp \left\{ j2\pi \left[\left(\frac{2BR_i}{TC} + \frac{2v_i f_c}{C} \right) \cdot \frac{n}{f_s} + \frac{2v_i f_c T}{C} p + \frac{2f_c R_i}{C} \right] \right\} \quad 3.30$$

Where R_i and v_i are the range and velocity of the i^{th} target, respectively. Taking the 2D FFT for $x_1(n, p)$ gives another $N \times P$ complex matrix with K peaks at different bins. A peak at the bin (k_i, l_i) corresponds to a certain target and can be used to estimate its range and velocity as follows:

$$\left(\frac{2BR_i}{TC} + \frac{2v_i f_c}{C} \right) \cdot \frac{1}{f_s} = \frac{k_i}{N} \quad \text{and} \quad \frac{2v_i f_c T}{C} = \frac{l_i}{P} \quad 3.31$$

The 2D FFT resulting matrix is called a Range-Doppler map because it maps range and velocity to targets. Now to estimate the angles of these targets, we use signals captured by other antenna elements. The signal received from the m^{th} antenna element can be written as [Section 3.2.1]

$$x_m(n, p) = \sum_{i=1}^K S_i(n, p) e^{\frac{-j2\pi(m-1)dsin(\theta_i)}{\lambda}}, m = 0, 1, \dots, M \quad 3.32$$

Eqn. 3.32 shows that if we perform the 2D FFT on the signal captured by any element, the resulting 2D FFT matrix will have peaks at the exact same bins corresponding to the same range-velocity pairs. To estimate the angle of the i^{th} target, we just collect the complex values corresponding to the target peak in the 2D FFT matrix across different channels to form one complex vector, \mathbf{v}_i , then take the FFT to that vector. This process is shown in Fig. 3.15.

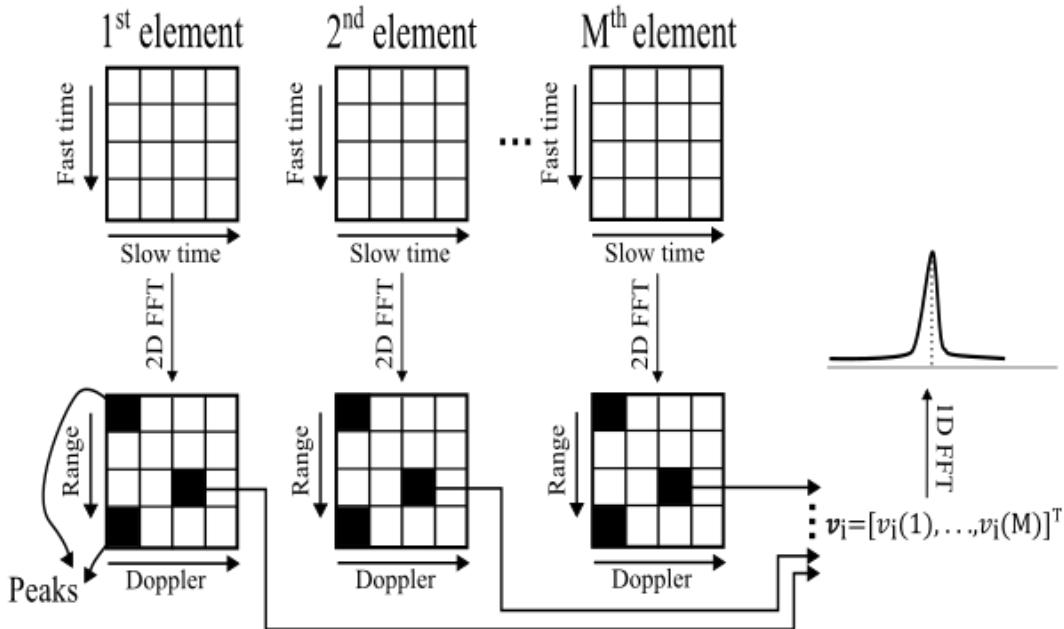


Fig. 3.15: Angle FFT process

The main idea here is that the elements of the complex vector \mathbf{v}_i will have a progressive phase shift from one element to the other, and this phase shift depends on the angle of that specific target that caused the peak in the first place. This is because the signal received from a certain target has a progressive phase shift that depends on the angle of arrival from one channel to the other. The elements of \mathbf{v}_i can be written as follows:

$$\mathbf{v}_i(m) = v_0 e^{-j2\pi \left(\frac{dsin(\theta_i)}{\lambda} \right) m}, m = 0, 1, \dots, M - 1 \quad 3.33$$

Where v_0 is a complex coefficient. It is now obvious that the vector \mathbf{v}_i consists of a single discrete angular frequency $\omega = 2\pi(ds\sin(\theta_i)/\lambda)$, which means that if we take the FFT for that vector, a single peak should be observed at the bin corresponding to that angular frequency. If the peak is observed at the bin number k , then the angle of the target can be estimated as follows:

$$\frac{2\pi}{M}k = 2\pi \left(\frac{dsin(\theta_i)}{\lambda} \right) \rightarrow \theta_i = \sin^{-1} \left(k \frac{\lambda}{Md} \right) \quad 3.34$$

The main advantage of this method is that we can pair the estimated angles with their corresponding targets in the Range-Doppler domain. This method is like the Capon and Bartlett methods in the sense that it doesn't require prior knowledge of the number of source targets. One of the main demerits of this method is the low angle resolution.

Another angle estimation method that uses the same process as the angle FFT method is found in [3]. In this method, the vector \mathbf{v}_i is used to create the correlation matrix \mathbf{R}_i as follows

$$\mathbf{R}_i = \frac{1}{M} \mathbf{v}_i \mathbf{v}_i^H \quad 3.35$$

We can then use this correlation matrix in any Beamforming method that depends on the correlation matrix, such as Capon or Bartlett, as discussed before.

Angle FFT Resolution

If two targets are at the same distance and moving with the same velocity relative to the radar, then they will correspond to the same peak in the 2D FFT matrix. This time when we take the 1D FFT of the complex vector \mathbf{v} corresponding to that peak, we get two peaks instead of one, indicating two objects at two different angles. The minimum angle between the two objects that allows the distinction between the two peaks in the angle FFT is known as the angle resolution. Since the two objects give the same peak in the 2D FFT matrix, then the elements of the corresponding vector can be formulated as

$$\mathbf{v}(m) = v_0 e^{-j2\pi \left(\frac{dsin(\theta_1)}{\lambda} \right)m} + v_1 e^{-j2\pi \left(\frac{dsin(\theta_2)}{\lambda} \right)m}, m = 0, 1, \dots, M - 1 \quad 3.36$$

Where θ_1 and θ_2 are the angles of the first and second targets, respectively, v_0 and v_1 are just complex coefficients. This time, the vector consists of two discrete angular frequencies $\omega_1 = 2\pi(ds \sin(\theta_1)/\lambda)$ and $\omega_2 = 2\pi(ds \sin(\theta_2)/\lambda)$. If the two objects are separated by an angle $\Delta\theta$, then the following condition needs to be satisfied for the two peaks in the angle FFT to be distinguishable [7].

$$\Delta\omega = \frac{2\pi d}{\lambda} (\sin(\theta + \Delta\theta) - \sin(\theta)) > \frac{2\pi}{M} \quad 3.37$$

If $\Delta\theta$ is small enough, we can use the derivative approximation $\sin(\theta + \Delta\theta) - \sin(\theta) \approx \cos(\theta)\Delta\theta$. The above condition then reduces to

$$\frac{2\pi d}{\lambda} \cos(\theta) \Delta\theta > \frac{2\pi}{M} \rightarrow \Delta\theta > \frac{\lambda}{Md \cos(\theta)} \rightarrow \theta_{res} = \frac{\lambda}{Md \cos(\theta)} \text{ (in rads)} \quad 3.38$$

An interesting thing to notice here is that the angle resolution depends on the angle θ . Two objects at larger values of θ are allowed to come closer to each other. A typical value for the separation between the elements is $d = \lambda/2$ which gives a resolution of $2/M$ at $\theta = 0$. This means that for a linear antenna array

of four elements, the angular resolution is about 29° which can be as large as the entire field of view of the long-range radar.

3.2.3. Multiple Signal Classifier (MUSIC)

MUSIC algorithm is completely different than the DBF methods as it takes the subspace approach to estimate the DOA [8, 9]. It utilizes some linear algebra and matrix manipulation concepts to estimate the DOA with very high angular resolution. The main concept used by this algorithm is the eigen-decomposition of matrices which can be found in [10]. The main idea behind this algorithm is discussed without going into much detail. We know from [section 3.2.1] that the array covariance matrix is given by $\mathbf{R} = E[\mathbf{XX}^H]$. Assuming that the signal and noise are uncorrelated, \mathbf{R} can be separated into signal and noise components as follows:

$$\begin{aligned}\mathbf{R} &= E[(\mathbf{AS} + \mathbf{N})(\mathbf{AS} + \mathbf{N})^H] = \mathbf{AE}[\mathbf{SS}^H]\mathbf{A}^H + E[\mathbf{NN}^H] \\ &= \mathbf{AR}_s\mathbf{A}^H + \sigma^2 \mathbf{I}_{M \times M}\end{aligned}\quad 3.39$$

Where \mathbf{R}_s is the $K \times K$ unobserved source signal covariance matrix and σ^2 is the AWGN variance. \mathbf{R} and \mathbf{R}_s are both Hermitian matrices, stay the same after taking the Hermitian operator, which implies that they have real eigenvalues whose sign determines the definiteness of these matrices, whether they are positive, semi-positive, or negative definite matrices. If there is no noise, then $\mathbf{R}_{M \times M} = \mathbf{AR}_s\mathbf{A}^H$. If the matrix \mathbf{R}_s is a non-singular matrix, $\text{Rank}(\mathbf{R}_s) = K$, then $\text{Rank}(\mathbf{AR}_s\mathbf{A}^H) = K$ because the matrix $\mathbf{A}_{M \times K}$ is assumed to be a Vandemonde matrix. This means that the matrix \mathbf{R} is not full rank having K positive real eigenvalues and $M - K$ zero eigenvalues, a semi-positive definite matrix. Now considering the noise, \mathbf{R} becomes a full-rank matrix with M eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ that are the sum of the eigenvalues of $\mathbf{AR}_s\mathbf{A}^H$ and σ^2 . The largest K eigenvalues and their corresponding eigenvectors belong to the signal subspace, while the rest $M - K$ belong to the noise subspace. Let the vector \mathbf{v}_i be the corresponding eigenvector to the eigenvalue λ_i , then we have:

$$\mathbf{R}\mathbf{v}_i = (\mathbf{AR}_s\mathbf{A}^H + \sigma^2 \mathbf{I}_{M \times M})\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, 2, \dots, M \quad 3.40$$

If we consider only the last $M - K$ eigenvalues and vector that belong to the noise subspace, then we have:

$$\mathbf{R}\mathbf{v}_i = (\mathbf{AR}_s\mathbf{A}^H + \sigma^2 \mathbf{I}_{M \times M})\mathbf{v}_i = \sigma^2 \mathbf{v}_i, i = K + 1, \dots, M \quad 3.41$$

After some simplification, we can reach the following result:

$$\mathbf{A}(\theta)^H \mathbf{v}_i = 0 \rightarrow (\mathbf{a}(\theta_k)^H) \mathbf{v}_i = 0, k = 1, 2, \dots, K, i = K + 1, K + 2, \dots, M \quad 3.42$$

That is, if we have any eigenvector that belongs to the noise subspace and a steering vector of one of the targets, then they have to satisfy the relation in Eqn. 3.42. If the steering vector is evaluated at any other angle that is not one of the targets' angles, then it will not satisfy the same relation with the noise eigenvectors. This is the main idea of the MUSIC algorithm. We use the noise eigenvectors to see whether the steering vector at a certain angle is normal to them or not. If yes, then there is a target at that angle.

The steps of the algorithm are as follows:

We first find the eigenvalues of the array covariance matrix \mathbf{R} and their corresponding eigenvectors. The eigenvalues are sorted by size. The larger K values belong to the signal space, and the rest $M - K$ belong to the noise space. The noise eigenvectors are used to construct a noise matrix $\mathbf{U}_N = [\mathbf{v}_{K+1}, \dots, \mathbf{v}_M]$. The MUSIC pseudo-spectrum is defined as

$$P_{MUSIC}(\theta) = \frac{1}{\|\mathbf{U}_N^H \mathbf{a}(\theta)\|^2} = \frac{1}{\mathbf{a}(\theta)^H \mathbf{U}_N \mathbf{U}_N^H \mathbf{a}(\theta)} \quad 3.43$$

The denominator is the inner product of the signal vector and the noise matrix. The angle at which the steering vector $\mathbf{a}(\theta)$ is normal to every column of the noise matrix, the denominator is ideally zero, and a peak is detected in the pseudo-spectrum, indicating a possible target at that angle.

The main advantage of the MUSIC algorithm is the high angular resolution, as it can reliably resolve targets with an angular separation of as low as 1° from severely degraded received signals with very few spurious peaks in the pseudo-spectrum. MUSIC can not estimate the source power like DBF methods do because it simply does not direct the main beam of the array in a certain direction and detect the received signal power. For power estimation, DBF methods can be used in conjunction with MUSIC. A main drawback of this algorithm is that it requires prior knowledge of the size of the signal space. It also requires higher computational complexity compared to DBF methods. Fig. 3.16 shows the MUSIC pseudo-spectrum when resolving two objects at $\pm 10^\circ$ using a five-element ULA antenna. It is clear how the two objects appear as very sharp peaks in the spectrum, indicating the high-resolution capabilities of the algorithm.

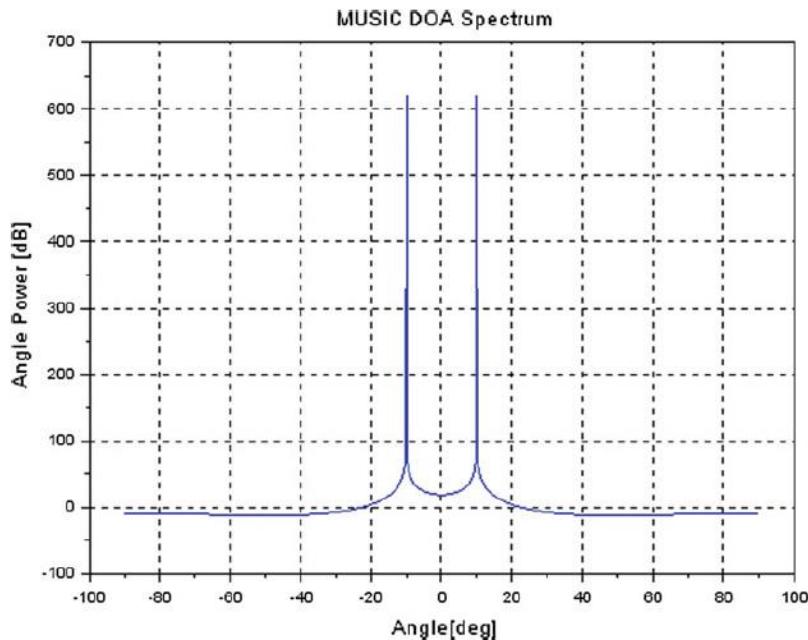


Fig. 3.16: the MUSIC algorithm pseudo-spectrum with two objects at $\pm 10^\circ$ using a five-element ULA antenna.

Variants of the MUSIC algorithm and other DOA estimation methods can be found in [4].

3.2.4. Multi-dimensional DOA algorithms

We have mainly focused on algorithms for one-dimensional DOA estimation. However, in automotive applications, it is necessary to estimate the angle of a target in both azimuth and elevation (two-dimensional DOA estimation). This is because, in many applications, the vehicle needs to estimate the height of on-road objects and road infrastructures, such as pedestrians, other vehicles, and bridges. There are two approaches for 2D DOA estimation. The first is to use the discussed algorithms in this chapter to perform 1D DOA estimation separately in both the azimuth and elevation directions, then do a pairing between the obtained angles. The main advantage of this approach is that it does not need more complicated algorithms for 2D DOA estimation. Instead, it uses well-established 1D algorithms. A drawback is that we have to perform pairing between the obtained angles in azimuth and elevation, which is not trivial and can lead to wrong position estimation.

The second option is to obtain the data from a 2D antenna array and perform 2D DOA estimation algorithms such as 2D DBF or 2D MUSIC. This method allows the direct extraction of both the azimuth and elevation angles of each target without pairing at the cost of increased computation complexity. The increased computation makes it challenging to use such algorithms for real-time applications.

References

- [1]. Pozar, D.M. (2012) ‘ELECTROMAGNETIC THEORY’, in *Microwave engineering*. 4th edn. Hoboken, NJ: Wiley.
- [2]. Balanis, C.A. (2016) ‘Arrays: Linear, Planar, and Circular’, in **Antenna theory: Analysis and design**. 4th edn. Hoboken, NJ: Wiley.
- [3]. Lee, S.; Kwon, S.-Y.; Kim, B.-J.; Lim, H.-S.; Lee, J.-E. Dual-Mode Radar Sensor for Indoor Environment Mapping. *Sensors* 2021, 21, 2469. <https://doi.org/10.3390/s21072469>
- [4]. Gamba, J. (2020) ‘Direction of Arrival (DOA) Estimation’, in *Radar Signal Processing for autonomous driving*. Singapore: Springer.
- [5]. Kronauge, M. and Rohling, H. (2014) ‘New chirp sequence radar waveform’, *IEEE Transactions on Aerospace and Electronic Systems*, 50(4), pp. 2870–2877. doi:10.1109/taes.2014.120813.
- [6]. Patole, S.M. et al. (2017) ‘Automotive radars: A review of Signal Processing Techniques’, *IEEE Signal Processing Magazine*, 34(2), pp. 22–35. doi:10.1109/msp.2016.2628914.
- [7]. *MIMO radar (rev. A) - texas instruments india*. Available at <https://www.ti.com/lit/an/swra554a/swra554a.pdf> (Accessed: 12 July 2023).
- [8]. Schmidt, R. (1986) ‘Multiple emitter location and signal parameter estimation’, *IEEE Transactions on Antennas and Propagation*, 34(3), pp. 276–280. doi:10.1109/tap.1986.1143830.
- [9]. Tang, H. *DOA estimation based on music algorithm - diva*. Available at: <https://www.diva-portal.org/smash/get/diva2:724272/FULLTEXT01.pdf> (Accessed: 12 July 2023).
- [10]. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore, MD (1996)

4

Interference and Interference Mitigation

The importance of sensors that are being used in vehicles had a big growth with the increasing development in self-driving technology. Recent achievements and improvements in the semiconductor industry made it possible to have mass production at a low cost for single-chip automotive radars. This, in turn, increased the spread of using radars in automotive applications. Currently, automotive radar is one of the main elements for Adaptive Driver Assistance Systems (ADAS), that's because of its good ability for detection in day, night, and different weather conditions.

Radar systems are mainly used to enhance situational awareness and road safety for vehicles; these systems share a frequency spectrum from 76 to 81 GHz. Due to the increased usage of radars, and thus a higher number of radar-equipped vehicles, this arises the problem of sensing unwanted radar signals generated by other radars (which is known as interference). This problem affects the functionality of the radar system negatively as it decreases the sensing capability. This problem may occur in congested traffic areas where we have many existing radar systems operating at the same time. Interference that occurs between multiple radar units may increase the system noise floor and, in turn, degrades the signal-to-noise ratio of the received signal, which reduce the capability of the operating radar to detect the target object. Another interference that is generated by another radar system may be strong enough to cause ghosts or false targets. Another type of interference is self-interference which may be generated from the mutual coupling between the transmitter and the receiver.

Thus, it's very important to mitigate the interference effect or at least avoid and reduce its effect because it decreases the resolution of the radar and may not be able to track or detect objects with small Radar Cross Sections (RCSs). This may introduce very dangerous situations, especially in fully autonomous vehicles where no intervention from humans exists, the dependence on the sensors becomes very high, and thus the failure of the sensors is completely unacceptable.

In this chapter, different types of automotive radar interference are introduced, their characteristics, as well as their effects on the overall operating radar system. Another thing to be reviewed is the different current techniques used for interference mitigation.

4.1. Automotive radar interference

Mitigation of interference has been a crucial issue for automotive radar and has gotten remarkable attention recently to improve the operation and the existence of multiple radar systems concurrently. Communication and joint radar represent the next stage of automotive radars that aim to combine sensing and communication functionalities on the same chip.

The standards for testing short-range radar devices are quite clear. However, there is no obligation or regulation on the modulation scheme or the structure of the waveform for the transmitted signal. Thus,

there are various approaches for mitigation of the interference and compensating its effects. As their various parameters are taken into account while designing a real-time automotive radar, so there is no clear approach that is considered to be the best for interference mitigation, and using one single technique won't be enough to ensure solving the problem.

4.1.1. Signal model for radar and the interference

Various signal models are transmitted in the automotive radar. In this case, Frequency-modulated Continuous-waveform (FMCW) is considered to deal with. The transmitted signal of a continuous waveform for the radar (S_t) can be described as

$$S_{t_{CW}}(t) = \text{rect}\left(\frac{t - \frac{T_p}{2}}{T_p}\right) e^{j2\pi f_c t}, \quad 4.1$$

where $\text{rect}()$ is the rectangular function that represents the chirp, T_p is the chirp duration of the transmitted signal, and f_c is the carrier frequency. Thus, the transmitted signal for FMCW will be

$$S_{t_{FMCW}}(t) = \text{rect}\left(\frac{t - \frac{T_p}{2}}{T_p}\right) e^{j2\pi(f_c t + \frac{1}{2}kt^2)}, \quad 4.2$$

where $k = BW/T_p$ is the chirp rate or the rate of the change in the modulated frequency across the chirp duration and is defined as the ratio of the bandwidth of the transmitted signal BW and the chirp duration T_p . The frequency band of the FMCW signals ranges from $[f_c - BW/2, f_c + BW/2]$. We can also notice from the previous equations that the FMCW signal becomes the CW signal when the chirp rate equals zero ($k = 0$).

Now the transmitted signal propagates in space and returns to the antenna at the receiver. Now it has new power and becomes a delayed version of the transmitted signal.

$$S_{r_{target}} = \text{rect}\left(\frac{t - \frac{T_p}{2} - \tau}{T_p}\right) \sqrt{P_{target}} e^{j2\pi(f_c(t-\tau) + \frac{1}{2}k(t-\tau)^2)}, \quad 4.3$$

where τ is the time delay taken for the signal to propagate between the radar and the target. Generally, the received signal contains a target signal, jammers, and an interference signal. So, to simplify the analysis, it's considered to have a single target with a single interference source. The interference signal can be described as

$$S_{r_{int}} = \text{rect}\left(\frac{t - \frac{T_{p_{int}}}{2} - \tau_{int}}{T_{p_{int}}}\right) \sqrt{P_{int}} e^{j2\pi(f_{c_{int}}(t-\tau_{int}) + \frac{1}{2}k_{int}(t-\tau_{int})^2)}, \quad 4.4$$

where τ_{int} is the time delay taken for the interference signal to propagate between the interference source and the victim radar, $f_{c_{int}}$ is the interference carrier frequency, k_{int} is the interference chirp rate, and $T_{p_{int}}$ is the chirp duration of the interfering radar.

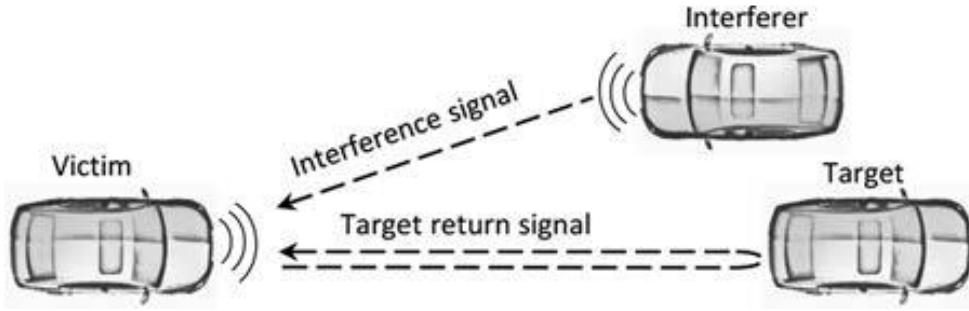


Fig. 4.1: The target signal suffers from a two-way round-trip propagation delay, while the interference signal suffers from only a one-way propagation delay.

From Figure 4.1, one can notice that the target and interference signals suffer from different delays, which should be considered in calculating the received power.

The target return signal can be described as

$$P_{target} = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R_{target}^4} \quad 4.5$$

And the interference signal can be described as

$$P_{int} = \frac{P_{t,int} G_{t,int} G_r \lambda_{int}^2}{(4\pi)^2 R_{int}^2} \quad 4.6$$

where P_t and $P_{t,int}$ are the victim and interference transmitted power, respectively, G_t and $G_{t,int}$ are the transmitting antenna gain and interference transmitting antenna gain, respectively, G_r is the receiving antenna gain, λ and λ_{int} are the wavelength of the transmitted signal and the interference signal, respectively, σ is the radar cross-section of the target object, and R_{target} and R_{int} is the range for the target and interference objects, respectively. Now from EQN and EQN, it's noticeable that the target power is inversely proportional to the target range as $P_{target} \propto (R_{target})^{-4}$, while the interference power is inversely proportional to the interference range as $P_{int} \propto (R_{int})^{-2}$, this may result in receiving interference power that would be higher than the target power. Now the signal-to-interference ratio (SIR) should be calculated to measure the possibility of detecting the target signal

$$SIR_{receiver} = \frac{P_{target}}{P_{int}} = \frac{\sigma}{4\pi} \frac{P_t}{P_{int}} \frac{G_t}{G_{t,int}} \frac{\lambda^2}{\lambda_{int}^2} \frac{R_{int}^2}{R_{target}^4} \quad 4.7$$

After the processing of the received signal, SIR will be given as

$$SIR_{proc} = \frac{G_p}{G_{p,int}} SIR_{receiver} \quad 4.8$$

where G_p is the processing gain for the target signal, and $G_{p,int}$ is the processing gain for the interference signal. We may also assume that the victim and interference radar are of the same kind (having the same parameters), then SIR will be reduced to

$$SIR_{proc} \approx \frac{G_p}{G_{p_{int}}} \frac{\sigma}{4\pi R^2} \quad 4.9$$

Thus, for good detection, SIR should be high. So, the waveform and radar parameters should be chosen to have $G_p > G_{p_{int}}$.

4.1.2. Characteristics of the automotive radar interference

Before going through the different characteristics of automotive radar interference, we should first classify the types of interference.

4.1.2.1. Types of the automotive radar interference

We may characterize the interference signal according to various criteria. Based on the interference source, we may classify the automotive radar interference into three sub-groups (induced by the self, cross, and other sources). Also, based on synchronization, we may have two sub-groups (synchronous and asynchronous interference).

The self and/or cross interference is usually caused by the radar system itself, and it is a synchronous interference. This type can be mitigated through the radar system design because its characteristics are well known. As an example, the mutual coupling between the transmitter and the receiver may be mitigated using an adaptive cancellation system to subtract the interference signal from knowing the transmitted waveform.

The other type of interference is the one generated from other sources (especially other radar systems). This type occurs on having multiple radar systems operating at the same time in the same frequency band. This results in mutual interference between each other. Unlike the past two types, this type of interference is usually asynchronous and has dynamic characteristics. That is because the parameters or the operation of the interferer radar may be changed at any time, and we cannot predict or control this change. That results in a continuous change in the interference characteristics as the interferer radar changes. For this, we may assume to have two FMCW automotive radars operating in the same vicinity, as shown in Figure 4.1. According to the chirp rate of the two operating radars, we may have a narrowband interference or a wideband interference, as shown in Figure 4.2.

Having different chirp rates for the two radars generates multiple beat frequencies during the duration of the interference, as noted in Figure 4.2a. This spreads out in a large frequency band. The duration of the interference is mainly determined by the low-pass filter (LPF) in the radar system. This wideband interference increases the noise floor of the victim radar and degrades the SNR of the received signal. On the other hand, having the same chirp rates results in having a single beat frequency. It has a narrow spread in the spectrum. This may result in ghost targets on the victim's radar. This is also called narrowband interference.

If a specific frequency is considered for the received interference signal, one may assume that the received interference signal (4.4) is mixed with the complex conjugate of the transmitted reference signal (4.2) in the analog domain, which can be expressed as:

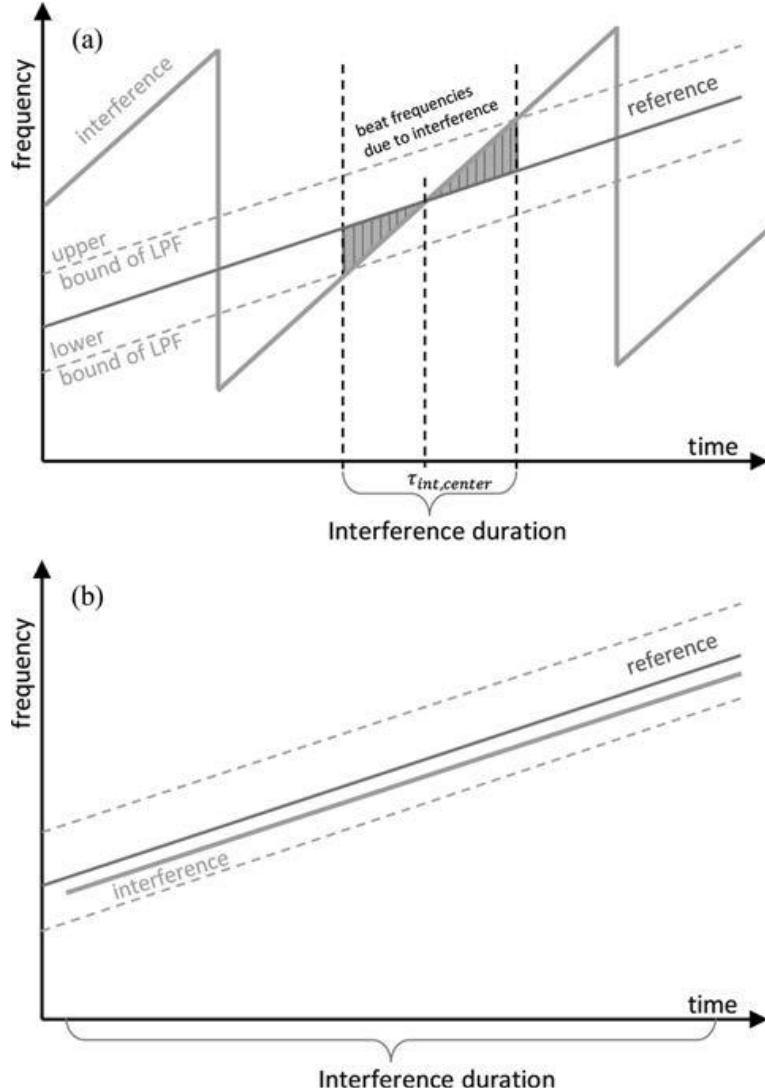


Fig. 4.2: Wideband interference is shown in (a), while narrowband interference is shown in (b).

$$y_{int} = S_{t_{FMCW}}^* S_{r_{int}} = \text{rect}\left(\frac{t - \frac{T_p}{2} - \tau_{int}}{T_p}\right) \sqrt{P_{int}} e^{-j2\pi(f_{c_{int}}\tau_{int} + (f_c - f_{c_{int}} + k_{int}\tau_{int}) + t^2(\frac{1}{2}k - \frac{1}{2}k_{int}) - \frac{1}{2}k_{int}\tau_{int}^2)} \quad 4.10$$

This process is called mixing or de-chirping. Usually, the beat frequencies that result from de-chirping or the down-conversion process are less than the bandwidth, so we use analog LPF to filter out any unwanted frequencies. On having the same chirp rate, the resulting beat frequency will cause a ghost target and won't be affected by the LPF bounds, as illustrated in Figure 4.2b. On the other hand, having different chirp rates results in multiple beat frequencies. Some of these frequencies pass in the duration of the interference range, and others are removed by the LPF bounds, as illustrated in Figure 4.2a. To define the limits of interference duration, we may assume to have a perfect Brick wall filter with a cut-off frequency of f_{LPF} , this can be described as

$$H(f) = \text{rect}\left(\frac{f}{f_{LPF}}\right) \quad 4.11$$

The phase of the received interference signal will be

$$\phi_{int}(t) = -2\pi \left(f_{c_{int}}\tau_{int} + (f_c - f_{c_{int}} + k_{int}\tau_{int}) + t^2 \left(\frac{1}{2}k - \frac{1}{2}k_{int} \right) - \frac{1}{2}k_{int}\tau_{int}^2 \right) \quad 4.12$$

Thus, we can calculate the received interference frequency as

$$f_{int}(t) = \frac{1}{2\pi} \frac{d}{dt} \phi_{int}(t) = t(k_{int} - k) + (f_{c_{int}} - f_c) - k_{int}\tau_{int} \quad 4.13$$

According to the previous equation (4.13), we may notice that the instantaneous frequency of the interference signal consists of three terms: the difference between chirp rates at a given time, the difference between the carrier frequency, and the corresponding beat frequencies of the interference.

The difference in the chirp rates describes the beat frequencies generated and their spread in the frequency domain. As mentioned before, that part of the interference signal is removed by the LPF, and the left pass to the ADC is digitized; then, we can describe these frequencies.

$$-f_{LPF} \leq (t(k_{int} - k) + (f_{c_{int}} - f_c) - k_{int}\tau_{int}) \leq f_{LPF} \quad 4.14$$

It's also noticeable that the interference is symmetric around the reference signal within the LPF bounds, thus the duration of the interference, T_{int} , can be given by

$$T_{int} \leq \left| \frac{2f_{LPF}}{k_{int} - k} \right| \quad 4.15$$

From this, one should now be able to find the number of samples that are affected by the interference as

$$N_{int} \leq T_{int}f_s \quad 4.16$$

As previously discussed, it's also been shown in Figure 4.2a that we have a symmetric FMCW interference around a specific time instance, $t_{int,center}$. From that, one can compute the final interference signal that is resulted from the LPF to be

$$y_{int,LPF}(t) = \text{rect}\left(\frac{t - \tau_{int,center}}{T_{int}}\right) y_{int}(t) \quad 4.17$$

The time instance of the center of the interference can be computed by equating the time when both frequencies are equal.

$$\tau_{int,center} = \frac{-(f_{c_{int}} - f_c) + k_{int}\tau_{int}}{k_{int} - k} \quad 4.18$$

Now, considering all the previous equations, one can compute the post-LPF received signal that contains both the interference and the target signals to be defined as

$$y_L = \begin{cases} y_{target} + y_{int} & t \leq \left| \frac{f_{LPF} - (f_{c_{int}} - f_c) + k_{int}\tau_{int}}{k_{int} - k} \right| \\ y_{target} & \text{otherwise} \end{cases} \quad 4.19$$

Where y_L is the received target signal after the LPF.

4.1.2.2. Duration of the Interference

To have a better understanding of the duration of interference, we examine a controlled experiment that is discussed briefly in [REF]. In this experiment, two controlled MIMO radar systems were used, one as a victim and the other as the interference source. Both of the radars were operating with the same parameters except for the chirp rate, and now we have a wideband interference.

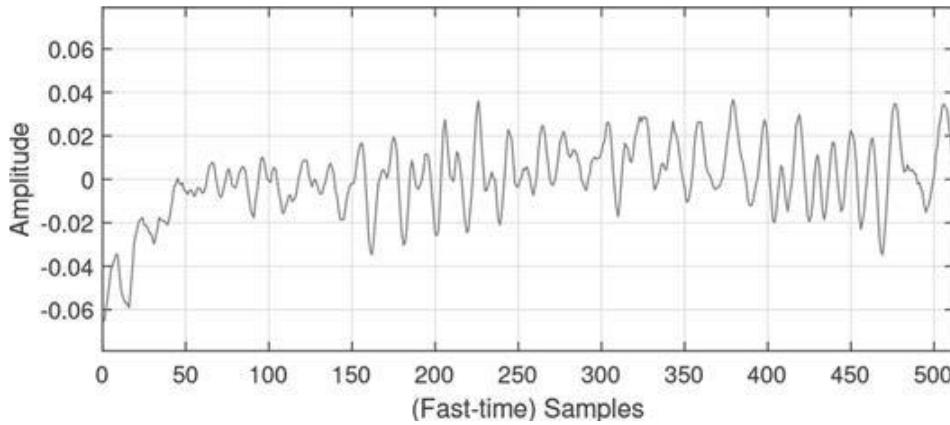


Fig. 4.3: Beat signal of the collected signal by victim radar.

The two radars were set to operate in MIMO mode, and the real part of the collected beat signal was displayed. Figure 4.3 shows collected signal samples when no interference occurs, while Figure 4.4 shows one pulse of the collected beat signal where the interference has an effect.

One cannot observe the characteristics we previously discussed for the interference nor the beat signal of the target due to the clutter used around the radar system. But we may be able to observe the high interference power that appeared between samples 230 to 330.

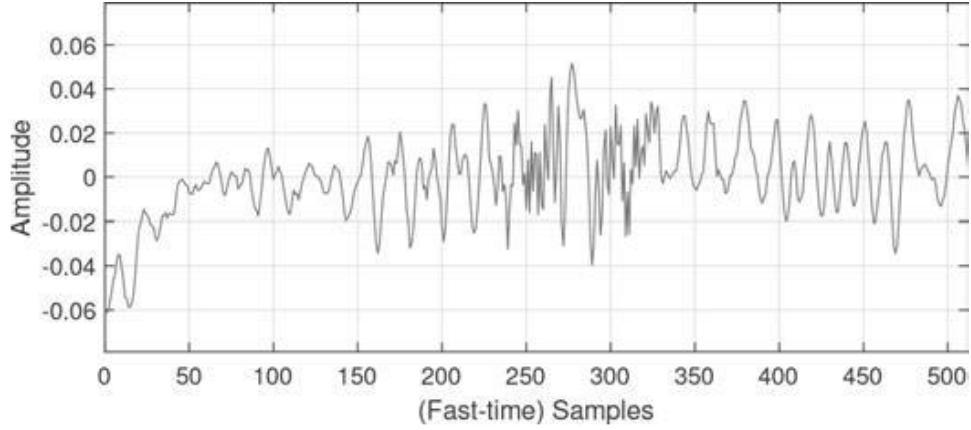


Fig. 4.4: Beat signal of the collected signal by the victim in the presence of interference.

For more investigation, we may use a High-pass Filter (HPF) as a ground clutter filter. This filter eliminates the frequency generated by the stationary clutter objects and has more focus on the target moving object. Now, it's clearly observable the characteristics of the moving target and the interference signals.

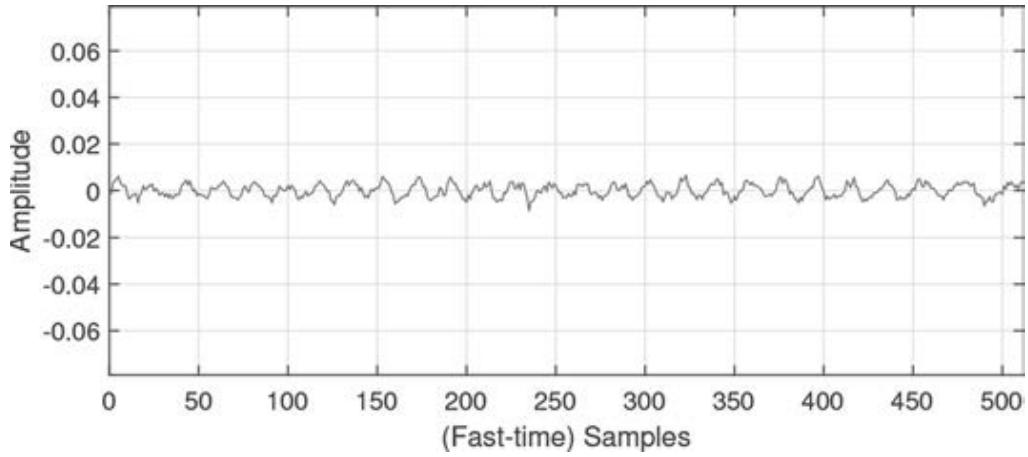


Fig. 4.5: Beat signal after using HPF from the collected signal by victim radar.

As shown in Figure 4.5, using the HPF filter eliminated ground clutter resulting from stationary objects (because stationary objects result in beat frequencies that are close to zero), and now our signal is a single tone according to (EQN). While Figure 4.6 shows the collected beat signal in the presence of interference. Now, it's clear to observe the beat signal of the moving target as well as the wideband interference. From the difference between the chirp rates, the duration of interference can be calculated (EQN). Thus, the number of samples to be affected can be calculated too (EQN).

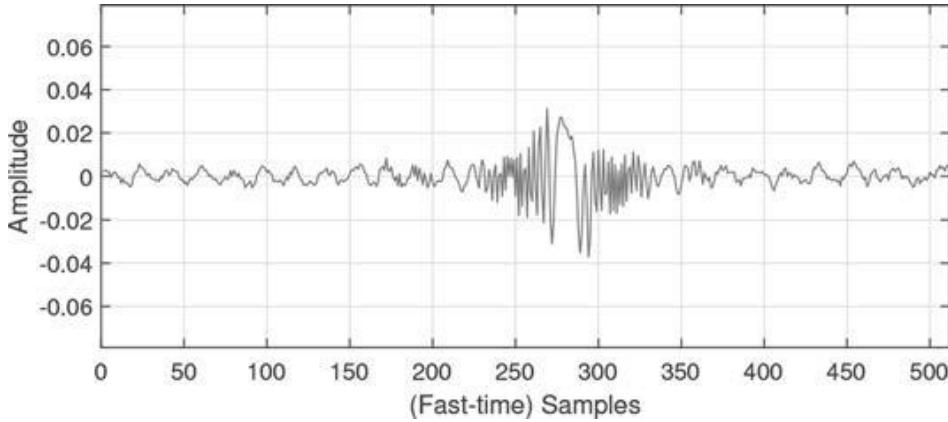


Fig. 4.6: Beat signal after applying the HPF for the collected signal in the presence of interference.

It should also be noted that the previous derivations for the duration of interference (4.16) assumed having an ideal LPF (it has a rectangular response in the frequency domain). Practically, ideal LPF cannot be realized using analog filters and end up having non-idealities in the frequency domain. The non-idealities will affect the duration of the interference resulting in more samples being affected by the interference than expected from the theory.

4.1.2.3. Interference Effects in range profiles

The range profile of the target object can be calculated by applying spectral estimation methods. It's noticeable that the frequency domain is directly proportional to the range domain. Recalling that $f_{beat} = k\tau$, where $\tau = 2R/c$, then the range profile can be estimated from the frequency spectrum analysis as follows

$$r = \frac{cf}{2k} = \frac{cfT_p}{2BW} \quad 4.20$$

Thus, the effect of the interference on the range domain is different according to the type, duration, and power of the interference. If the interferer and victim radar both have the same chirp rate ($k_{int} = k$), then the power of the interferer appears at a specific bin in the range domain. This bin now represents a ghost target because it reflects a power from an object that doesn't actually exist in this range. It's also hard to differentiate between the ghost and real targets in the range domain. If the interferer and victim radar doesn't have the same chirp rate ($k_{int} \neq k$), then the power of the interference starts to spread on more range bins creating a wideband interference. This interference raises the noise floor of the system in the range domain.

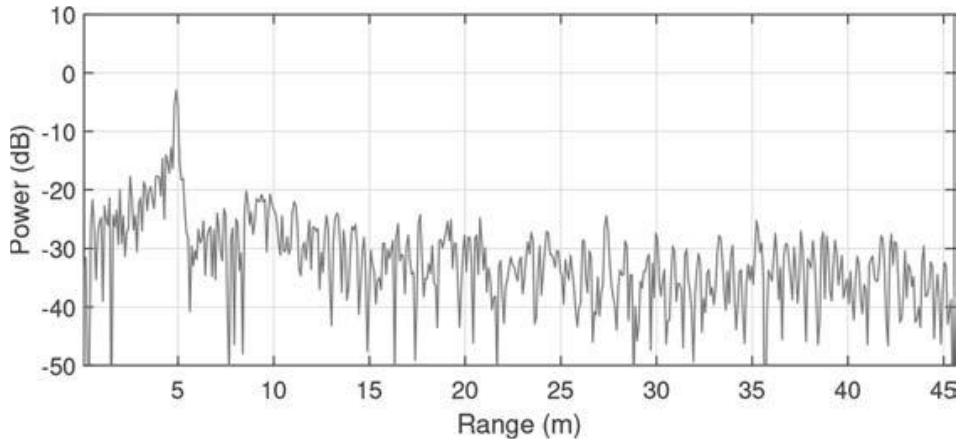


Fig. 4.7: Range profile of a single pulse from the collected data by the victim radar with no interference exists. The target object exists at 5m.

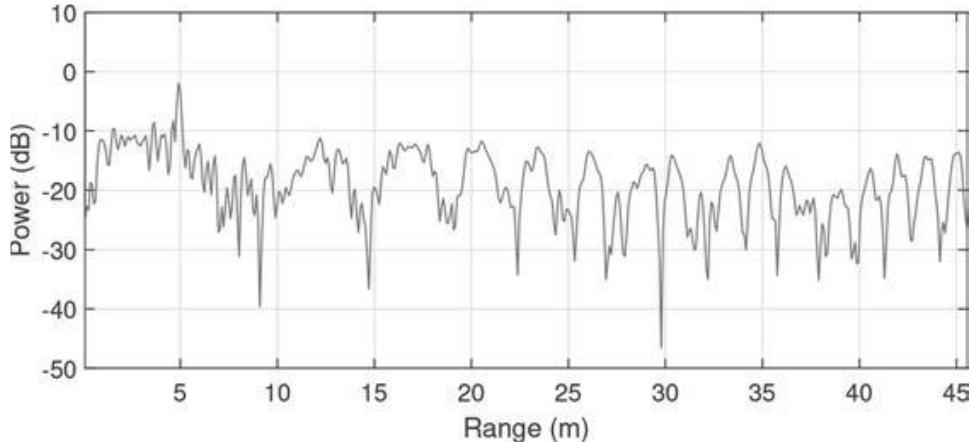


Fig. 4.8: Range profile of a single pulse from the collected data by the victim radar in the presence of interference.

It's clear to notice in Figure 4.7 the range profile of the target object by the victim radar, where the target occupies the bin 5m away from the radar. Figure 4.8 illustrates the same but in the presence of interference. The target object stays in the same bin but the noise floor increases.

4.1.2.4. Summary

Automotive radar interference has many degrees of freedom, so it's not easy to visualize or present all the possible cases. In this section, FMCW-to-FMCW interference is presented for a single interference source. This is a single case among all the different cases. The effect of this case was discussed to give an intuition to the reader about the interference. To be more general, one should use the generalized radar equation and should investigate it in different scenarios. This is more covered and validated by experiments in the literature. Interested readers may refer to References [REF] for different interference scenarios.

4.2. Interference Mitigation

Automotive radar interference became very challenging when the number of radar users increased. That's because as the number of users increases, the allocated spectrum for transmission stays the same, so the rate of interference increases. Despite having many different approaches for mitigation that have been

developed, none of them is efficient enough to handle all the interference cases. Most of the mitigation and avoidance techniques depend on the orthogonal transmission and reception in frequency or time domain. Thus, it requires improvements in the radar hardware to allow different waveforms, which ensure orthogonality.

Interference mitigation techniques can be classified into two groups. The first group aims to handle interference effects after it happens, and the second group aims to avoid interference by adapting radar parameters. Though, one should note that none of these groups ensures the mitigation of interference in congested areas. In this case, advanced methods that are based on communication protocols are essential. There are many mitigation techniques available in the literature. Some of them are presented in this chapter.

4.2.1. Detection of interference

To have a good choice for the mitigation technique, one should first specify the type of interference. So, it's necessary to detect the interference in the received signal first, which will help in choosing a mitigation technique. Detecting the interference is usually easy when the interference power is large. One should note that it's not a preferable case because the interference effect is worse. Having a high interference input power, that is higher than the maximum allowable power by the front end, saturates the receiver and results in a clipped signal at the ADC. The clipped signal samples are now distorted. They may not be recovered if the number of samples is too large compared to the total number of samples.

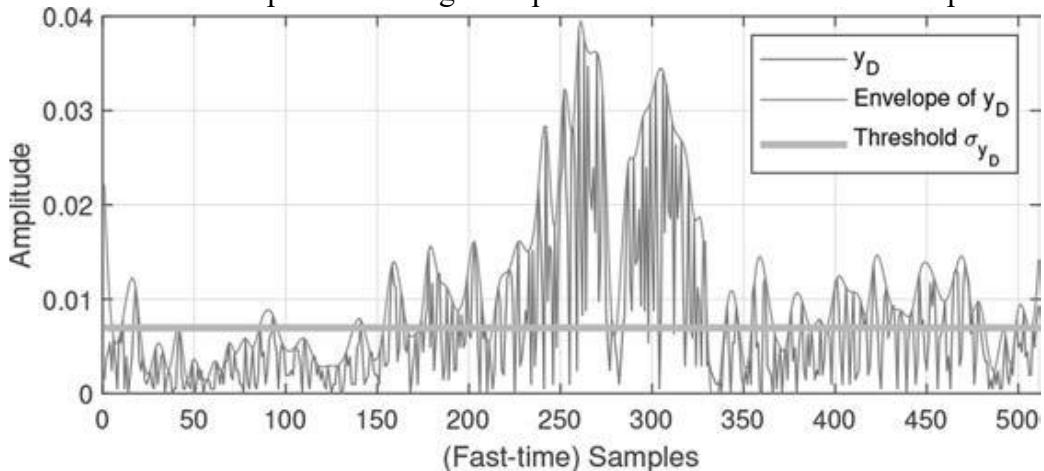


Fig. 4.9: Interference detection where a detector is applied to the received signal.

On the other hand, an interference power that is low enough not to saturate the receiver but still higher than the target signal is harder to detect than in the first case. In this case, a detector with a fixed threshold may be applied to detect the change in the signal power, thus the interference too. This threshold can be computed on the value of the standard deviation. Also, a smoothing filter may be added to cluster the output signal. These two modifications are clear in Figure 4.9.

The detector only is simple, and sometimes it becomes effective for some applications. The success of this detector depends on some parameters, such as interference, target, and clutter power. As shown in Figure 4.9, many target samples passed the threshold. So, it's hard to determine which samples are affected by the interference and which ones didn't. Thus, the fixed threshold is not sufficient for interference detection.

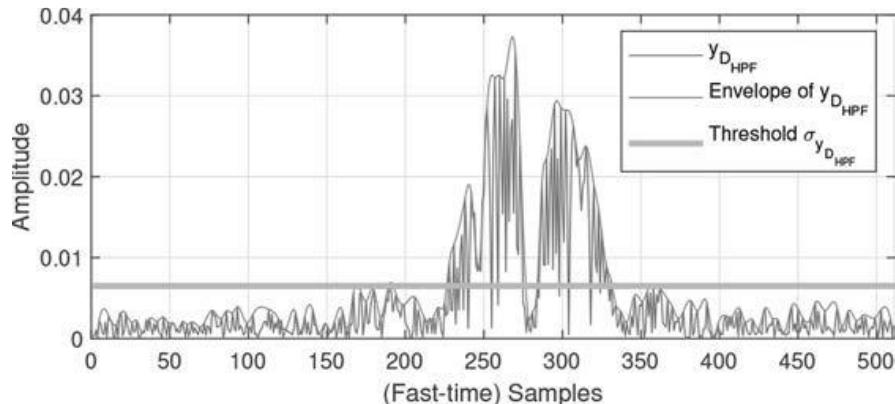


Fig. 4.10: Interference detection after HPF where a detector is applied to the received signal.

One approach to solving the previous problem is to have a dynamic threshold that is computed according to a specific algorithm. Instead, the approach presented is to use the simple detector on a high-pass filtered received signal as shown in Figure 4.10. As previously discussed, the HPF acts as a ground clutter to remove the power of the close target to have more focus on the target object. One can determine samples affected by interference (from 221 to 331) as there are only two target samples that passed the threshold (191 and 359). It's also clear that the detection of interference increased.

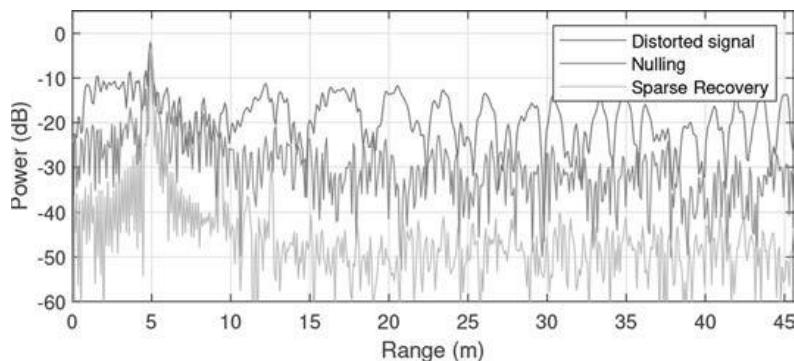
4.2.2. Interference Mitigation and Avoidance

The detection of interference results in choosing among different signal processing steps that aim to mitigate or avoid the interference. The radar system itself may alter the radar parameters to avoid interference effects in future frames. Currently, the approaches used for interference mitigation in the industry focus on different techniques to remove and repair distorted samples. These approaches are presented in this section.

4.2.2.1. Nulling and signal restoration

Nulling is a simple approach where the affected samples by interference are removed and replaced by zero. This process is done before range processing, so it reduces the noise floor generated by interference. Nulling is applicable in different domains, such as time or frequency domains. There's a drawback, as this nulling causes discontinuity in the received signal. This raises the sidelobes around strong power targets, which may block the detection of weak power targets. One approach for the raised sidelobes is to repair the deleted samples. There are different techniques for repairing deleted samples as interpolation, aggressive modeling [REF], and sparse recovery [REF].

Nulling method is effective in having a small number of samples affected by interference compared to the total number of samples. If the number of interfered samples is relatively large, then the nulling method



is not efficient in that case. Instead, the sparsity-based signal separation method is used to separate the beat signals from the interference by making use of knowing interference features [REF]. So, they don't need any interference detection as the nulling does.

Fig. 4.11: Illustration for the received signal in the presence of interference and the mitigation by nulling and sparsity-based method

As shown in Figure 4.11, the first method used was the Nulling method, where the detected interfered samples were removed. The second method was a sparsity-based signal separation method with no interference detection. Both methods are followed by FFT to generate the range profile of the received signal. As a comparison, the first method mitigates the interference but still suffers from high sidelobes that result from deleting the interfered samples. The sparsity-based separation method provides better interference mitigation. Although the sparsity-based separation method has a better response, it's currently not preferred due to its computational complexities.

4.2.2.2. Frequency altering and timing jitter

The frequency-hopping approach depends on having interference detection at the start. First, interference detection is applied for each pulse of the detection frame. If interference is detected, the interferer's center frequency is identified as above or below the victim radar's center frequency. According to this estimate, the center frequency of the victim radar is shifted up or down within the available frequency band [REF]. This method is efficient in having available bandwidth to hop within it. Short-range radars require large bandwidth of 2-4 GHz, which corresponds to 77-81 GHz [REF]. So, it's not possible to hop to another frequency because theirs is not enough available bandwidth. Instead, the transmitted pulse can be divided into smaller sub-pulses, each randomly hopping their center frequency within the band. Similarly, the time jitter is the same frequency altering but in the time domain. This can be applied by changing the temporal position of the transmitted signal.

These two methods can be well controlled on the victim's radar, but the transmission in the interferer radar cannot be controlled. So, it's not always efficient to ensure robust interference mitigation.

4.2.2.3. Cooperative Time and frequency scheduling

As the number of operating radar systems is increasing, the existence of all these radars becomes complex. Also, as the time and frequency resources are limited, so frequency and time altering cannot be left uncontrolled. One approach is the cooperative time and frequency scheduling method. This method mitigates interference by coordinating automotive radars through a specific communication standard. This way, each radar can decide its frequency or time scheme without interfering with other radars.

It's possible to have an external communication channel dedicated to communication between different radars. Though, it's preferable to have communication and sensing capabilities on the same chip. This, in turn, has more advantages in size, cost, power consumption, and robustness [REF]. This can be completely realized in the digital domain without having any change in the used hardware. This can be done by using a suitable waveform for sensing and communication [REF]. The communication approach requires synchronization, while the used hardware is not designed for either communication or synchronization. So, modifications to the hardware are under investigation to have reliable communication between different radar systems [REF].

4.2.2.4. Digital beamforming and null steering

Most of the current radar systems use MIMO technology to exploit the angle of arrival of the received signals. Beamforming is used to maximize the received power at a specific angle (angle of interest). It is also possible to reduce the power in a specific direction to mitigate interference, which is called null

steering. Figure 4.12 illustrates an example where beamforming is applied to focus on the target object. The response is shown in Figure 4.13.

Simple null steering can be done by calculating the steering vector in the interference direction and removing its response from the beamforming response. This requires knowing the exact direction of arrival for the interferer. This, in turn, increases the complexity and workload of the signal processor. One should also note that beamforming and null steering are efficient for interference mitigation in the spatial domain. Though, it may cause losses of the target return signal in the direction of interference radar. So, it won't be possible to detect weak power targets around the interference direction.

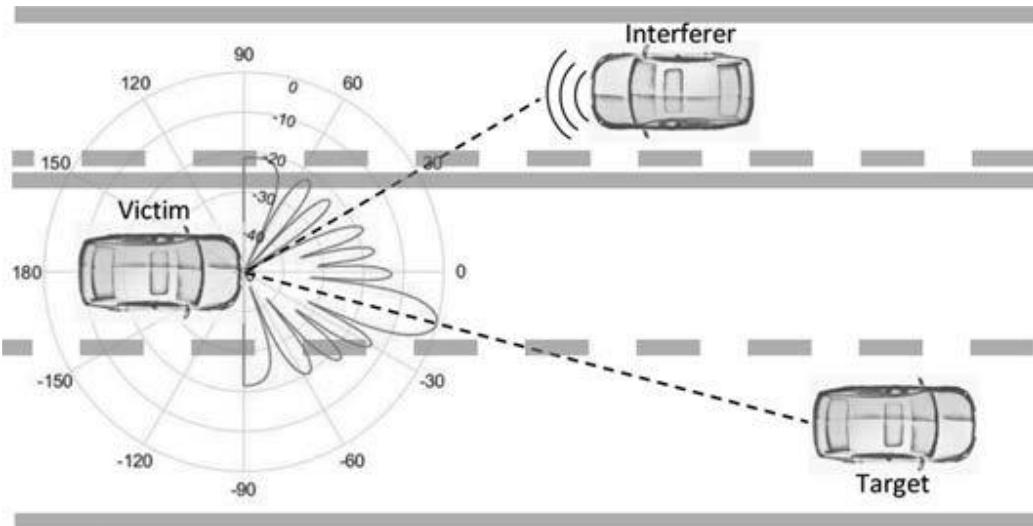


Fig. 4.12: Example scenario for beamforming and null steering.

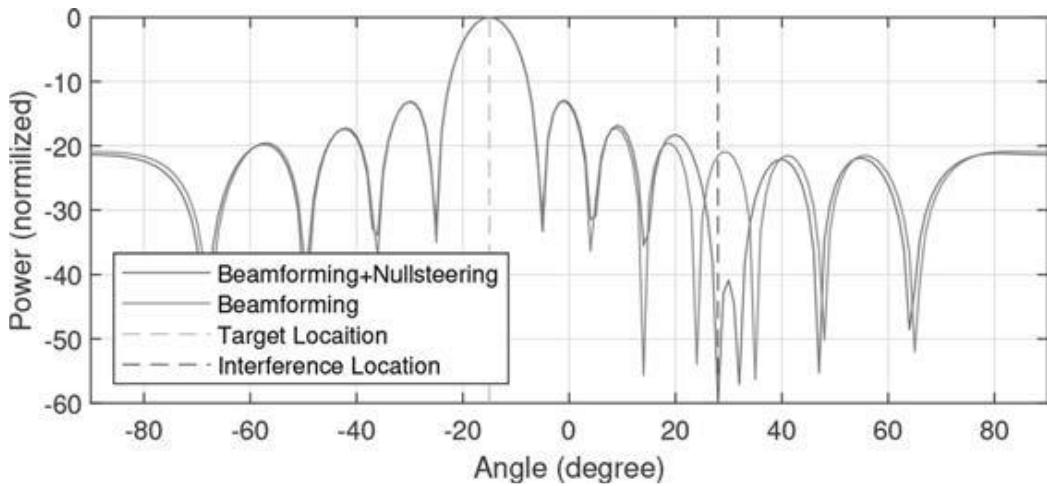


Fig. 4.13: Array response for the scenario in Figure 4.12. Target and interference are located at -15° and 28° , respectively.

4.2.3. Summary

This section gave an introduction to the interference detection problem and presented some of the approaches used for interference mitigation and avoidance. As previously discussed, automotive radar interference has many degrees of freedom. So, it's very hard to find a single mitigation technique to solve the addressed problem. The lack of standardization of automotive radar systems as well as the emerging hardware technology, makes interference control challenging. It's expected shortly to have a multi-functional single-chip for both sensing and communication. Multi-functional radar systems are expected to operate in cooperation instead of coexistence. This may finally solve the problem of interference.

References

- [1]. Peng, Z., Li, C. and Uysal, F. (eds.) (2022) *Modern Radar for automotive applications*. S.l.: INST OF ENGIN AND TECH.
- [2]. Sanka, S. (2017) *RADAR to RADAR Interference for 77GHz Automotive RADARs*. thesis.
- [3]. Series M. ‘Systems characteristics of automotive radars operating in the frequency band 76–81 GHz for intelligent transport systems applications’. Recommendation ITU-R, M; 2014. pp. 2057–1.
- [4]. Uysal F. ‘Synchronous and asynchronous radar interference mitigation’. IEEE Access: Practical Innovations, Open Solutions. 2019, vol. 7, pp. 5846–52.
- [5]. Kumbul U., Uysal F., Vaucher C.S., Yarovoy A. ‘Automotive radar interference study for different radar waveform types’. IET Radar, Sonar & Navigation. 2022, vol. 16(3), pp. 564–77.
- [6]. Uysal F., Sanka S. ‘Mitigation of automotive radar interference’. IEEE Radar Conference (RadarConf18); Oklahoma City, OK, 2018. pp. 0405–10.
- [7]. Norouzian F., Pirkani A., Hoare E., Cherniakov M., Gashinova M. ‘Phenomenology of automotive radar interference’. IET Radar, Sonar & Navigation. 2021, vol. 15(9), pp. 1045–60.
- [8]. Rameez M. Interference Mitigation Techniques in Fmcw Automotive Radars. [Phd. Thesis]. Blekinge Tekniska Högskola; 2020.
- [9]. echter J., Roos F., Rahman M, et al. “Automotive radar interference mitigation using a sparse sampling approach”. European Radar Conference; 2017. pp. 90–93.
- [10]. Correas-Serrano A., Gonzalez-Huici M.A. “Sparse reconstruction of chirplets for automotive FMCW radar interference mitigation”. IEEE MTT-S International Conference on Microwaves for Intelligent Mobility; 2019. pp. 1–4.
- [11]. ourdoux A., Parashar K., Bauduin M. ‘Phenomenology of mutual interference of FMCW and PMCW automotive radars’. IEEE Radar Conference (RadarConf17); Seattle, WA, 2017. pp. 1709–14.
- [12]. Bechter J., Sippel C., Waldschmidt C. ‘Bats-inspired frequency hopping for mitigation of interference between automotive radars’ in ICMIM’. 2016 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility; 2016. pp. 1–4.
- [13]. Aydogdu C., Keskin M.F., Carvajal G.K., et al. ‘Radar interference mitigation for automated driving: exploring proactive strategies’. IEEE Signal Processing Magazine. 2020, vol. 37(4), pp. 72–84.

- [14]. Ma D., Shlezinger N., Huang T., Liu Y., Eldar Y.C. ‘Joint radarcommunication strategies for autonomous vehicles: combining two key automotive technologies’. IEEE Signal Processing Magazine. 2020, vol. 37(4), pp. 85–97.
- [15]. Uysal F., Orru S. ‘Phase-coded FMCW automotive RADAR: application and challenges’. 2020 IEEE International RADAR Conference (RADAR); Washington, DC, USA, 2020. pp. 478–82.
- [16]. Lampel F., Uysal F., Tigrek F., et al. ‘System level synchronization of phase-coded FMCW automotive radars for radcom’. 2020 14th European Conference on Antennas and Propagation (EuCAP); Copenhagen, Denmark, 2020. pp. 1–5.

5

Detection Fundamentals

5.1. Probability of detection and false alarm detection

There are two possibilities for a radar measurement. The first one is that when there is no target so the measurement is only due to noise H_0 . The second is that when there is a target so this measurement is due to noise and a reflected signal from that target H_1 .

The samples in each frequency bin after performing the 2-D FFT can be expressed as a random variable y . This random variable has two different probability density functions (PDF) given each of the two different natures of the measurement (H_0 or H_1). $p_y(y|H_0)$ is the pdf of the sample, given that there is no target, so it represents the distribution of the noise while $p_y(y|H_1)$ is the pdf of the sample, given that there is a target, so it represents the distribution of the actual received signal with noise.

With the assumption that the noise of the system is white noise, then $p_y(y|H_0)$ is normally distributed, and by assuming a non-fluctuating target (has constant cross-section (RCS)), then $p_y(y|H_1)$ is also normally distributed as H_1 is a combination of the reflected signal and noise.

$$H_0 : y \sim N(a_n, \sigma_n^2) \quad 5.1$$

$$H_1 : y \sim N(a_s, \sigma_s^2) \quad 5.2$$

Where a_n , a_s are the means of H_0 and H_1 respectively, σ_n^2 and σ_s^2 are the variances of H_0 and H_1 respectively, so the pdfs are given by:

$$p_y(y|H_0) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{(y-a_n)^2}{2\sigma_n^2}} \quad 5.3$$

$$p_y(y|H_1) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{(y-a_s)^2}{2\sigma_s^2}} \quad 5.4$$

Probability of detection (P_D) is the probability that there is a target H_1 and the radar system detects that target. While the probability of false alarm detection (P_{FA}) is the probability that there is no target H_0 but the radar system detects a target.

For radar systems, the Neyman–Pearson criterion is commonly used where a threshold is chosen. Target

detection happens when the value of y is above the specified threshold. If the received y is below the threshold, no target is detected, even if there was actually a target. Thus, the probability of detection P_D and the probability of a false alarm P_{FA} are given by, see Fig. 5.1.

$$P_D = \int_{y=T}^{\infty} p_y(y|H_1) dy \quad 5.5$$

$$P_{FA} = \int_{y=T}^{\infty} p_y(y|H_0) dy \quad 5.6$$

If a certain threshold is specified, the probability of a false alarm remains constant, as it depends only on the noise distribution in the absence of targets. However, the probability of detection depends on the received signal SNR. Higher SNR will shift the red curve in Fig. 5.1 to the right causing a higher probability of detection for the same P_{FA} . This is because when the SNR is high, the received signal mean, the mean value of y , in the presence of a target is higher. If we need a certain P_D and P_{FA} , a threshold T is specified according to the required P_{FA} , and a certain minimum SNR is needed to ensure that the P_D will not go below the required value.

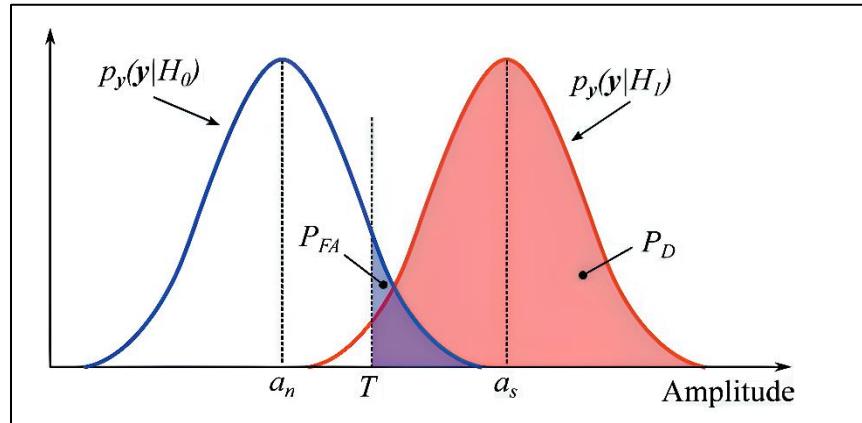


Fig. 5.1: PDFs of H_0 and H_1 with P_D and P_{FA} illustrated.

In the following section, we will discuss coherent and non-coherent integration as methods to improve the system performance either by improving the received SNR, which increases the P_D for the same P_{FA} , or by reducing the minimum required SNR.

5.2. Coherent and noncoherent integration

To improve the SNR for radar systems, integration techniques are commonly used. There are two types of integration, namely, coherent integration and non-coherent integration. Coherent integration means that the phase information is taken into consideration, while the non-coherent integration, only the magnitude is considered. Coherent integration achieves better SNR compared to non-coherent integration, but it requires more processing.

5.2.1. Coherent integration

Integration is done by taking multiple measurements. This may be done by taking the measurements from different receiving channels or repeating the measurements N times. For coherent integration, these data should be added in their complex form (with phase considered), and this phase is assumed to be the same for each of these measurements, which is not always true, so the measured data needs to be pre-processed to align the phase of these measurements which increases the complexity.

If a radar system received a signal $Ae^{i\phi} + W$ from a target where A and ϕ are the amplitude and phase of the signal, respectively, and W is the additive white Gaussian noise with variance σ^2 , then the SNR is given by:

$$SNR_1 = \frac{A^2}{\sigma^2} \quad 5.7$$

If N measurements are taken. Coherent integration (by adding these measurements with its phase) is applied as follows:

$$\sum_{n=0}^{N-1} (Ae^{i\phi} + W(n)) = NAe^{i\phi} + \sum_{n=0}^{N-1} W(n) \quad 5.8$$

As the signal is added coherently, the integrated power of these signals will equal $N^2 A^2$, and because the noise is uncorrelated, the noise will be added in the power domain so the total integrated noise power will equal $N\sigma^2$ so the integrated SNR will equal:

$$SNR_{integrated} = \frac{N^2 A^2}{N\sigma^2} = N \cdot SNR_1 \quad 5.9$$

So the coherent integration improves the SNR by multiplying it by the number of measurements, which is also known as the coherent integration gain.

5.2.2. Non-coherent integration

For non-coherent integration, the amplitude of the received signal is only considered, and the phase information is not used, so with the assumption of a nonfluctuating target with additive white Gaussian noise ω , the data to be processed in each measurement in the case where there is no target (H_0) is $y = |\omega|$, and in the case where there is a target (H_1) is $y = |Ae^{i\phi} + \omega|$.

Non-coherent integration changes the distributions of y whether there was a target (H_1) or not (H_0). It can be shown [1] that non-coherent integration doesn't improve the SNR. However, the improvement will be in the false alarm detection probability. After integration, the threshold required for achieving a certain P_{FA} decreases, as illustrated in Fig. 5.2. This reduces the minimum required SNR for a certain probability of detection. As shown in Fig. 5.3, for $N = 1$ to achieve a certain P_D it requires a high SNR while for a higher N the same P_D can be achieved with a smaller SNR.

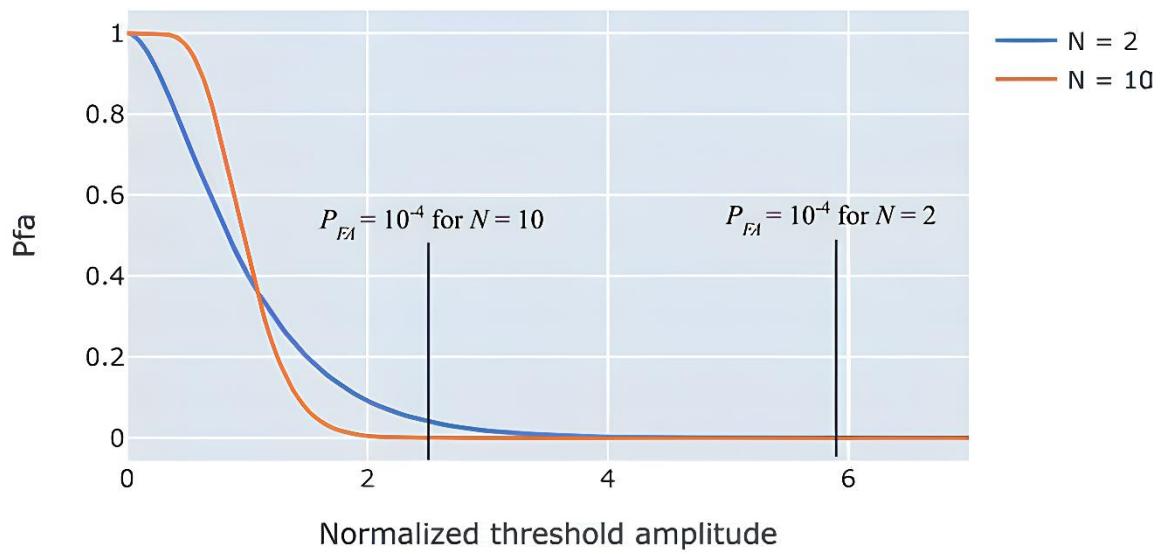


Fig. 5.2: P_{FA} versus the normalized threshold for different N values.

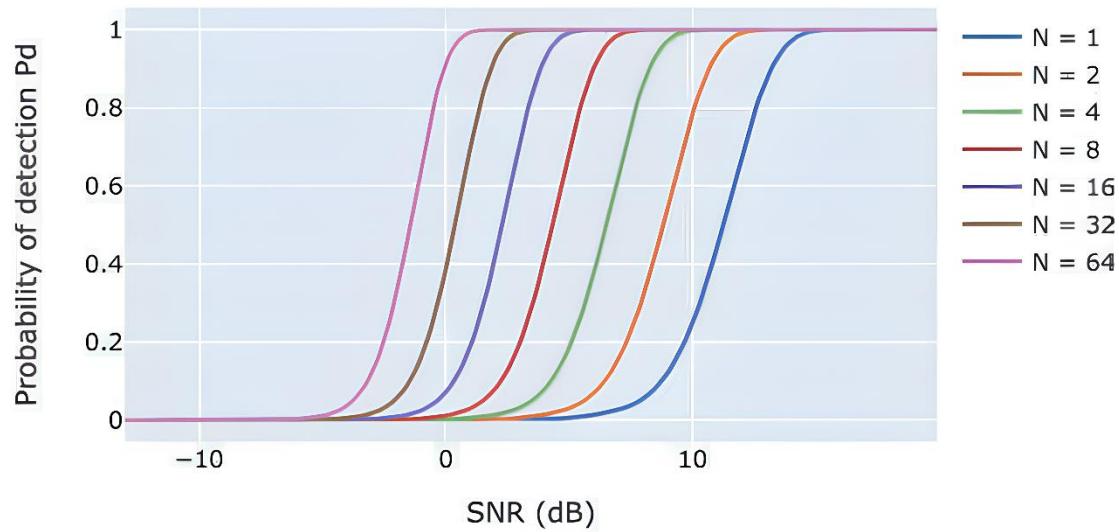


Fig. 5.3: P_D verses SNR for different N values.with $P_{FA} = 10^{-6}$

5.3. Peak detection

In the previous section, the threshold T , which corresponds to a certain P_{FA} is assumed to be constant, and that is a result of the assumption of having a constant and stable noise level which in reality is not very accurate due to the dependency of the noise on various factors such as environment, Temperature, and components aging. Also, for automotive radar, the radar is usually surrounded by other radars, which will cause interference. This interference will also raise the noise level, so a constant value for the threshold is not the best solution for real cases; instead, this threshold needs to be adaptive.

Constant false alarm rate (CFAR) detection is a technique used to approximately maintain a constant false alarm rate for realistic noise scenarios by adaptively changing the required threshold, and the threshold is calculated by estimating the real-time noise power. There are different algorithms based on CFAR, such as cell-averaging CFAR (CA-CFAR), Cell-Averaging Smallest-Of CFAR (CASO-CFAR), Cell-Averaging Greatest-Of CFAR (CAGO-CFAR), and ordered-statistic CFAR (OS-CFAR), choosing which algorithm to apply is usually a trade-off between accuracy and complexity.

5.3.1. CA-CFAR

CA-CFAR is a very simple CFAR algorithm, and it is widely used. It is based on calculating the average of real-time noise in the cells surrounding the cell under test (training cells) and calculating the threshold based on this average. The algorithm uses the sliding window technique to determine the average by averaging the noise power in the training cells while leaving some guard cells around the cell under test to avoid the effect of leakage from the tested cell to the surrounding ones causing wrong estimation for noise level. This window can be 1-D or 2-D, as shown in Fig. 5.4. The noise power for the training cells is calculated as follows:

$$P_n = \frac{1}{M} \sum_{m=1}^M y_m \quad 5.10$$

Where M is the number of training cells and y_m is the power of the m^{th} training cell, this power is used to estimate the threshold as follows:

$$T = \alpha P_n \quad 5.11$$

Where α is a scaling factor, and its value depends on the required probability of a false alarm P_{FA} .

As shown in Fig. 5.5, the CA-CFAR algorithm causes a high threshold near peak values which will cause masking of the small targets near the large ones. For this algorithm to be accurate, different targets need to be far enough from each other to avoid masking each other. The algorithm also assumes that the noise in the training cells is independent and identically distributed, which may not always be true. These limitations for CA-CFAR caused the development of other algorithms which are more complex but more accurate.

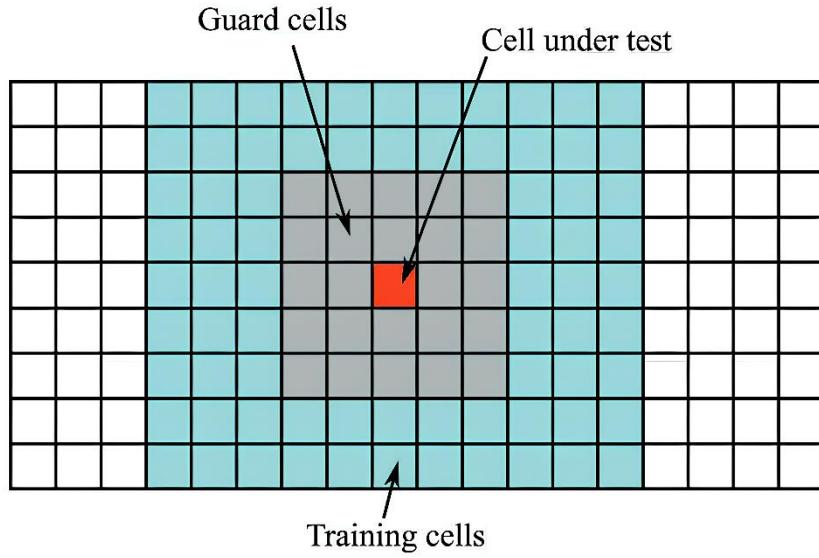


Fig. 5.4: 2-D CA-CFAR processing.

5.3.2. Cell-Averaging Greatest-Of CFAR (CAGO-CFAR)

CAGO-CFAR is an algorithm that is similar to CA-CFAR. In the 1-D case, the difference is that there are 2 averaging processes done; one is for the training data to the left of the cell under test, and the other is done for the data to the right. The highest value of these two averages is chosen as the noise power to be used in estimating the threshold. The advantage of this algorithm is the reduced complexity, and it may improve target detection in some cases. However, the problem of masked targets is not solved.

5.3.3. Cell-Averaging Smallest -Of CFAR (CASO-CFAR)

It is the same as CAGO-CFAR, but the difference is that the chosen value for the average is the smaller value. This may help in increasing the detection of masked targets. However, this may raise the probability of false alarm detection.

5.3.4. OS-CFAR

OS-CFAR is an algorithm that is more adaptive to varying noise conditions. It uses the same sliding window as CA-CFAR, but it uses order statistics to estimate the noise power as it ranks the collected noise power in the training cells ascendingly and chooses the n^{th} element from that ordered data. It may seem that this algorithm depends on one sample only. However, due to the ordering process, this algorithm depends on all data values. After calculating the n^{th} value $y(n)$ of the ordered list, this value is used to estimate the threshold as follows:

$$T = \alpha_{os}y(n) \quad 5.12$$

Where α_{os} is a scaling factor. It can be seen from Fig. 5.5 that the OS-CFAR doesn't have a high threshold near the peak value, which is an advantage for this algorithm, but the sorting process requires more processing, so it is more complex.

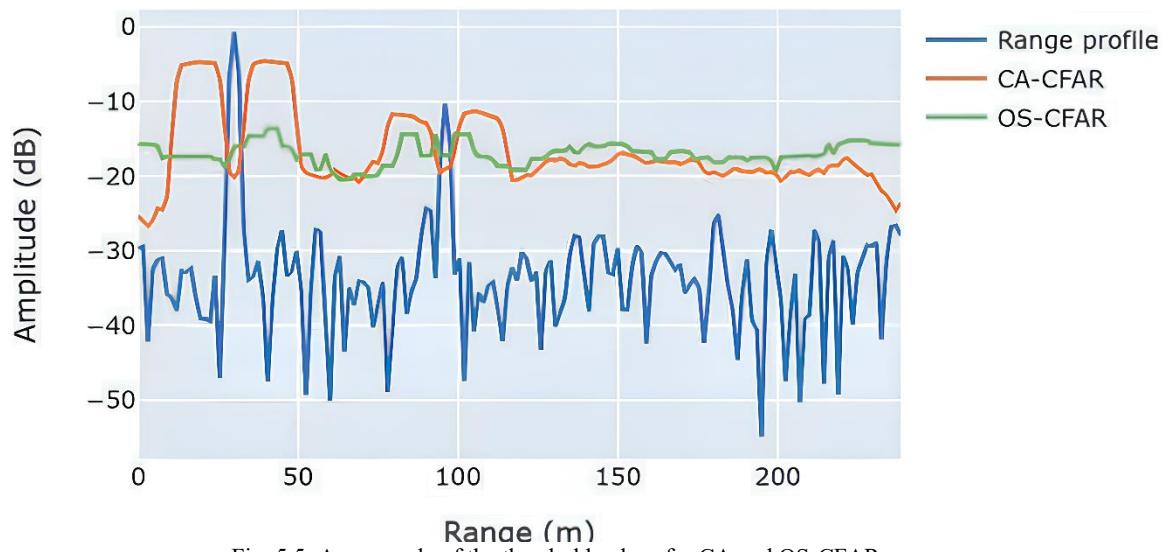


Fig. 5.5: An example of the threshold values for CA and OS-CFAR.

References

- [1] J. Gamba, *Radar Signal Processing for Autonomous Driving*. Singapore: Springer, 2020.
- [2] *Modern Radar for Automotive Applications*. S.l.: INST OF ENGIN AND TECH, 2022.

6

SYSTEM-LEVEL DESIGN

The radar system performance is determined by the false alarm rate and the probability of detection. As discussed in the previous chapter, there is a minimum value for the received signal SNR to guarantee a given P_{FA} and P_D . In this chapter, we will investigate some of the impurities in the receiver chain that cause SNR deterioration, namely, the electronic components' noise represented by the receiver noise figure (NF), the signal distortion due to the nonlinearity of different blocks, and the ADC quantization noise. Radar-to-radar interference and clutter also increase the system noise floor causing more SNR deterioration, but we will not focus on them. We will also discuss the different radar transceiver architectures, and at the end of this chapter, we will take a look at the different approaches along with some MATLAB simulations that can be used to derive the required specs of the system RF front-end.

6.1. Performance Parameters

In this section, we will discuss the main metrics used to quantify the performance of the RF systems, such as system-added noise (noise figure), linearity (IIP3, P1dB), receiver sensitivity, and dynamic range.

6.1.1. System-added Noise

The input signal to a receiver chain comes from the outside world with an added thermal noise floor. Each component in the receiver adds more noise which causes the signal SNR to deteriorate along the receive path. Thus, the total output SNR is less than the Rx input SNR. To quantify how a certain block in the chain deteriorates the signal-to-noise ratio, we define the noise factor F as follows:

$$F = \frac{SNR_i}{SNR_o} \quad 6.1$$

The noise factor is always greater than the unity. If the circuit is noiseless; adds no noise, then its noise factor is equal to one. The noise factor of a block is always evaluated for a certain source resistance as follows [1]:

$$F = \frac{SNR_i}{SNR_o} = 1 + \frac{\overline{v_{n,out}^2}}{4kT R_s A_0^2} \quad 6.2$$

Where $\overline{v_{n,out}^2}$ is the output referred noise power spectral density due to the circuit noise sources only, A_0^2 is the total voltage gain of the circuit from the input source to the output terminals, R_s is the source resistance, k is Boltzmann's constant, and T is the absolute noise temperature in kelvin. The noise factor

is rarely used in its scalar form; we use the dB quantity instead, known as the noise figure $NF = 10 \log_{10} F$.

For a cascade of linear gain blocks, the total noise factor F_t can be calculated using the Friss equation as follows

$$F_t = 1 + (F_1 - 1) + \frac{F_2 - 1}{A_{p1}} + \dots + \frac{F_m - 1}{A_{p1}A_{p2} \dots A_{p(m-1)}} \quad 6.3$$

Where F_i and A_{pi} are the noise factor and available power gain of the i^{th} stage, respectively. The available power gain is defined as the power that the stage delivers to a matched load divided by the available source power. In a matched system, the available power gain becomes just the power gain of the stage. Once we have the total noise figure of the receive chain, we can calculate the output SNR through Eqn. 6.4.

$$[\text{SNR}_{\text{out}}]_{dB} = [P_{in}]_{dBm} + 174 \left(\frac{dBm}{Hz} \right) - NF_{tot} - 10 \log_{10} BW \quad 6.4$$

Where $[P_{in}]_{dBm}$ is the input signal power to the receiver in dBm, NF_{tot} is the total cascaded noise figure of the receiver in dB, and BW is the noise bandwidth in hertz. In the case of designing a radio receiver, the noise bandwidth becomes the channel bandwidth of the radio standard to be implemented. However, when the receiver is intended for a radar system, the noise bandwidth is the equivalent noise bandwidth (ENBW) of the FFT, which is discussed in Chapter 2.

Another important metric that relates directly to the receiver-added noise is the receiver sensitivity. The sensitivity is defined as the minimum input signal power that can be detected with an acceptable output SNR. In other words, it is the input power that results in the minimum output signal-to-noise ratio $SNR_{o,min}$. The sensitivity is calculated by rearranging Eqn. 6.4 as follows.

$$[P_{Sen}]_{dBm} = [\text{SNR}_{o,min}]_{dB} - 174 \left(\frac{dBm}{Hz} \right) + NF_{tot} + 10 \log_{10} BW \quad 6.5$$

If the input power is less than the sensitivity level, then the output SNR will be less than the minimum acceptable value, and the target will not be detected. Eqn. 6.4 shows that the output SNR increases with increasing the input signal power in dBm, which means that at high input powers, the performance limiting factor is no longer the output SNR, and other performance metrics should be taken into consideration, such as linearity.

6.1.2. Linearity

An ideally linear amplifier has its output signal $v_{out}(t)$ linearly related to the input signal $v_{in}(t)$ with some gain factor α_1 . That is for a linear amplifier, we can write $v_{out}(t) = \alpha_1 v_{in}(t)$. However, in real amplifiers, the relation between the output and input is a lot more complicated, and we can model the system as a memoryless nonlinear system having higher order powers of the input as follows:

$$v_{out}(t) = \alpha_1 v_{in}(t) + \alpha_2 v_{in}^2(t) + \alpha_3 v_{in}^3(t) + \dots \quad 6.6$$

If the amplitude of the input signal is small enough, a small signal approximation is valid, and the higher-

order terms are neglected. In this case, the amplifier behaves as a linear system. When the input signal level starts to increase, the higher-order terms can no longer be ignored, and many interesting effects are observed, such as harmonic distortion, gain compression, gain desensitization, cross modulation, and intermodulation (third and second order). We will mainly focus on gain compression and intermodulation. Gain compression is characterized using the 1-dB compression point P_{1dB} while the third-order and second-order intercept points are used to quantify the intermodulation.

6.1.2.1. 1-dB compression point (P_{1dB})

For the memoryless non-linear system described by Eqn. 6.6, if the input signal is $v_{in}(t) = A\cos(\omega t)$, then after some simplification the output signal can be written as in Eqn. 6.7.

$$v_{out}(t) = \left(\frac{1}{2}\alpha_2A^2\right) + \left(\alpha_1 + \frac{3}{4}\alpha_3A^2\right)A\cos(\omega t) + \frac{1}{2}\alpha_2A^2\cos(2\omega t) + \frac{1}{4}\alpha_3A^3\cos(3\omega t) + \dots \quad 6.7$$

Therefore, we conclude that a single input tone generates harmonics at the fundamental frequency and its multiples which is known as harmonic distortion. Eqn. 6.7 also shows that the gain of the fundamental is equal to $\alpha_1 + (3/4)\alpha_3A^2$. Once again, if A is small enough, the second term is neglected, and the gain becomes α_1 . However, as A increases, the gain starts to compress, assuming that α_3 is negative. The level of the input at which the fundamental gain drops by 1 dB from its small signal value is defined as the 1-dB compression point. $A_{in,1dB}$ definition is shown in Fig. 6.1.

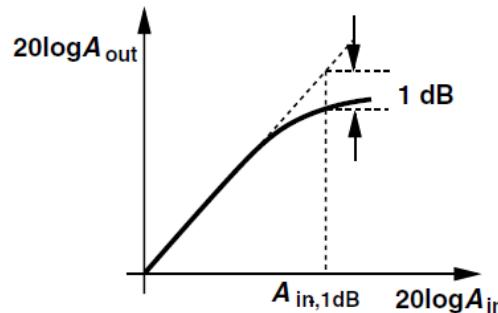


Fig. 6.1: 1-dB compression point definition.

The $A_{in,1dB}$ can be easily found through the following equation.

$$20 \log_{10} \left(\left| \alpha_1 + \frac{3}{4}\alpha_3A_{in,1dB}^2 \right| \right) = 20 \log_{10} |\alpha_1| - 1 \rightarrow A_{in,1dB} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|} \text{ (linear scale)} \quad 6.8$$

Usually, the power level of the input at which the compression happens IP_{1dB} is used instead of the input amplitude. The output-referred 1-dB compression point OP_{1dB} can be used. The OP_{1dB} and IP_{1dB} are related through the following equation:

$$OP_{1dB} = IP_{1dB} + G \quad 6.9$$

Where G is the linear power gain of the block in dB

6.1.2.2. Intermodulation

If two tones at frequencies ω_1 and $\omega_2 = \omega_1 + \delta$ with amplitude A are applied to the memoryless non-linear system described by Eqn. 6.6, it can be shown that the output spectrum will contain components at the fundamental frequencies and all possible combinations of the two frequencies. Components at frequencies $n\omega_1 \pm k\omega_2$ where n, k are non-zero integers are called intermodulation (IM) products. The third-order intermodulation products (IM3) are the two products at frequencies $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$ while second-order products (IM2) are located at $\omega_2 - \omega_1$ and $\omega_2 + \omega_1$.

If the two tones are close to each other, the IM3 will be close to the fundamental tones ω_1 and ω_2 causing some sort of in-band distortion, as shown in Fig. 6.2. If we are dealing with a modulated signal instead of the two tones, this distortion degrades the system performance if the signal-to-distortion ratio SDR is not acceptable.

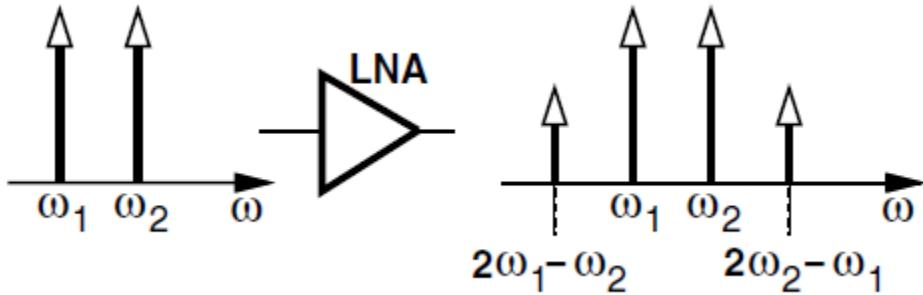


Fig. 6.2: third-order intermodulation products

The level of IM3 products depends on the input power level of the two tones and can be represented either by its absolute value in dBm or relative to the output power level of the fundamental tones in dBc. Since the IM3 level is not constant, it cannot be used to characterize the IM3 distortion; instead, we use the third-order intercept point (IP_3) as the metric for IM3 distortion measurement.

The input-referred third-order intercept point (IIP_3) is defined as the theoretical input power level at which the power of the IM3 products reaches that of the fundamental tones. As shown in Fig. 6.3, if the input power is small enough, the power level of the IM3 products increases by 3dB for every 1dB increase in the input power. At high input powers, both curves start to compress, but if the curves are extrapolated, they intercept at the IP_3 point. Just like the P_{1dB} , we can refer the IP_3 point either to the input or the output.

For a small enough input power, we can get the IM3 level in dBc through the following expression:

$$IM3_{dBc} = -2(IIP_3 - Pin_{dBm}) = -2(OIP_3 - Pout_{dBm}) = -SDR \quad 6.10$$

Eqn. 6.10 means that if two blocks have the same input power, the block with higher IIP_3 will cause less IM3 power level therefore, less IM3 distortion or, in other words, a higher signal-to-distortion ratio SDR.

In practice, the IIP_3 is measured by applying two tones with a small input power, smaller than the compression point of the block, then measuring the IM3 level in dBc in the output spectrum. Finally, we use Eqn. 6.10 to calculate the intercept point. IM2 products are usually less important because, in most applications, they will be located outside of the band of interest. However, similar concepts and equations are used for the IM2.

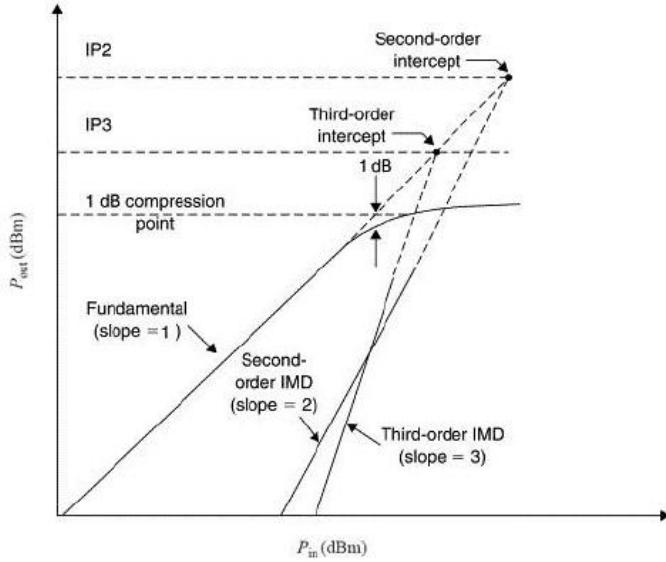


Fig. 6.3: illustration of IIP3 and IIP2

6.1.2.3. Cascaded Non-Linearity

Same as we did with the noise figure, we wish to find the total 1-dB compression point or third-order intercept point for a cascade of blocks. Equations 6.11 and 6.12 are used for that purpose where G_i is the power gain of the i^{th} block in the chain.

$$\frac{1}{P_{1dB,tot}} = \frac{1}{P_{1dB,1}} + \frac{G_1}{P_{1dB,2}} + \frac{G_1 G_2}{P_{1dB,3}} + \dots \quad 6.11$$

$$\frac{1}{IIP_{3,tot}} = \frac{1}{IIP_{3,1}} + \frac{G_1}{IIP_{3,2}} + \frac{G_1 G_2}{IIP_{3,3}} + \dots \quad 6.12$$

From these equations, we can conclude that the latter stages in the RF chain are the ones that dominate the overall linearity of the chain due to the increase in the power level of the signal as it flows through the chain. Knowing the overall IIP_3 of the chain, we can calculate the output signal to $IM3$ distortion ratio (SDR) through Eqn. 6.10.

6.1.3. Dynamic Range

Dynamic range (DR) is defined as the maximum input signal level that a receiver can tolerate divided by the minimum input level that it can detect, which is the receiver sensitivity by definition. The maximum input level is defined as the power level beyond which the receiver starts to suffer from severe linearity issues such that the signal is not detectable anymore due to the degraded SDR. DR is limited by the compression point at the upper end and noise at the lower end. Variable gain is necessary to reduce the gain for high input powers, increasing the maximum tolerable signal level, which increases the receiver DR in return. In the log scale, the DR becomes the difference between the maximum and minimum input powers in dBm, as shown in Fig. 6.4.

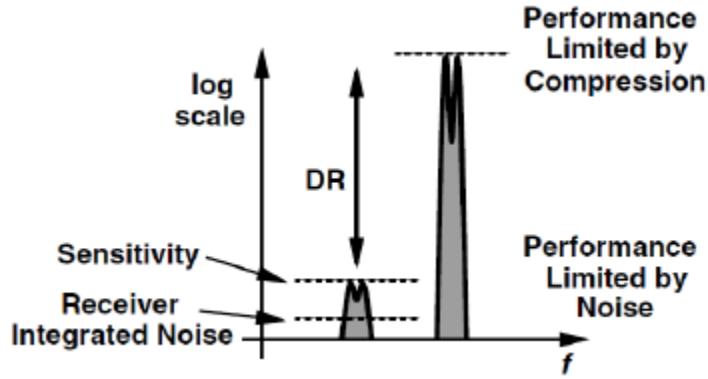


Fig. 6.4: Dynamic Range definition.

Another important definition is the spurious-free dynamic range (SFDR). For the SFDR, the maximum input level is the maximum level in a two-tone test for which the $IM3$ products level, when referred to the input, does not exceed the integrated noise of the receiver. The definition of the SFDR is shown in Fig. 6.5.

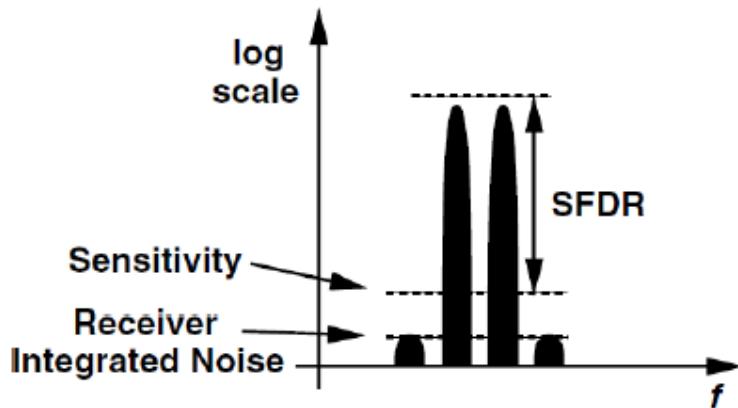


Fig. 6.5: Spurious-free dynamic range definition.

SFDR represents the maximum relative level of an interferer that can be tolerated by the receiver while producing an acceptable signal quality for a small input signal at the sensitivity level.

6.1.4. Quantization noise

At the end of any receiver chain, there is an analog-to-digital converter (ADC). The ADC samples the analog signal with a certain frequency. The sample value is mapped to one of the values from a set of discrete levels, quantization levels; then, each level is represented by a certain digital code, see Fig. 6.6. The analog value of the least significant bit (LSB) of the output digital code is equal to the width of the quantizer's steps. The quantization process is necessary as we cannot represent continuous values with a finite number of digital codes, and it adds some error to the system. The quantization error V_e is defined as the difference between the quantization level V_q and the sample's original value V_{in} . We now wish to derive a model for the quantization error to analyze the performance degradation due to the quantization process. The quantization process can be seen as adding the quantization error sequence to the input samples. That is, $V_q[n] = V_{in}[n] + V_e[n]$. Under certain conditions [2], $V_e[n]$ can be modeled as a random process whose samples are drawn from a certain distribution. This allows the use of a deterministic statistical model.

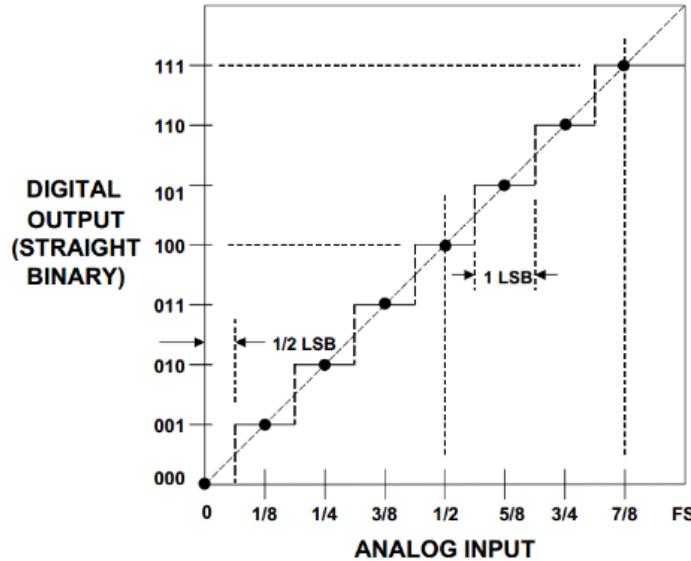


Fig. 6.6: the transfer function of an ideal unipolar three-bit ADC

Fig. 6.7 shows the quantization error sequence resulting from applying the discrete cosine signal $x[n] = \cos(n/10)$ to a uniform eight-bit (256 level) quantizer. The LSB of this quantizer is equal to $2/256$. The error signal, in this case, seems to be random, and the distribution of the error value can be found by constructing a histogram, as shown in Fig. 6.8.

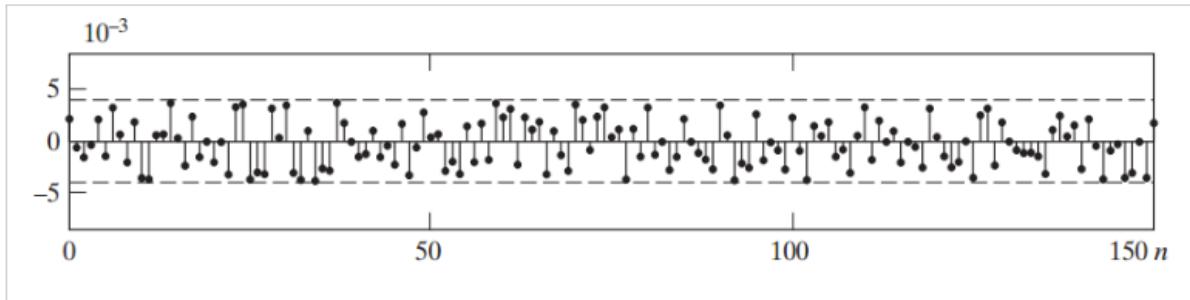


Fig. 6.7: the quantization error signal of $x[n] = \cos(n/10)$ when applied to an 8-bit quantizer.

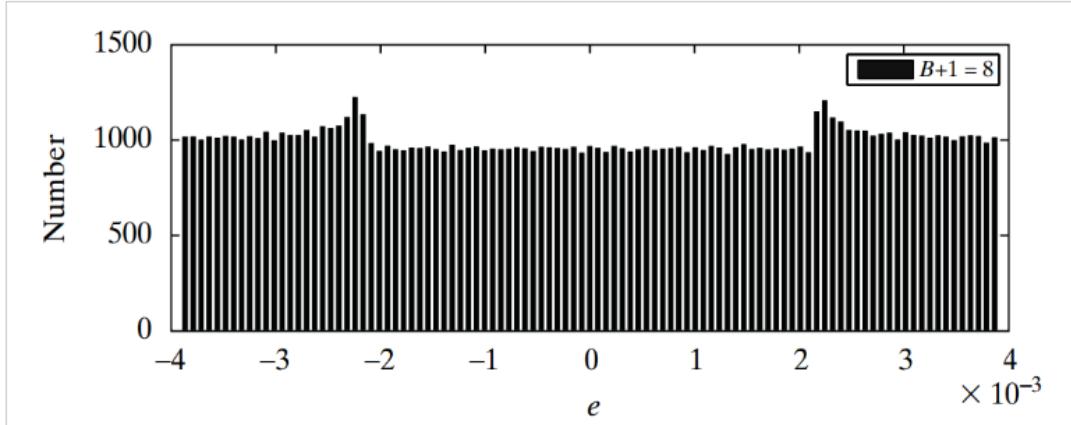


Fig. 6.8: The constructed histogram for the error value

Fig. 6.8 shows that the error value is a zero-mean uniformly distributed random variable with the following probability density function.

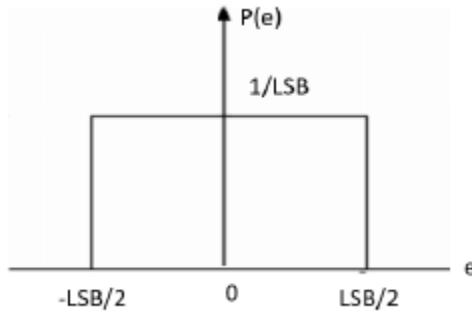


Fig. 6.9: PDF of the error value

We can now calculate the quantization noise total power (error variance) as follows:

$$P_{Qnoise} = \sigma_e^2 = E[e^2] - \overbrace{(E(e))^2}^{Zero} = \int_{-\infty}^{\infty} e^2 P(e) de = \int_{\frac{-LSB}{2}}^{\frac{LSB}{2}} \frac{e^2}{LSB} de = LSB^2/12 \quad 6.13$$

Since the quantization noise samples are uncorrelated, then the autocorrelation function (ACF) of the noise process is a delta function in the time domain. The power spectral density (PSD) of the noise process is the Fourier transform of the ACF. Since the Fourier transform of the delta is one, then the PSD of the noise will be frequency-independent, which means that the quantization noise is a white noise with a total power of $LSB^2/12$. The PSD of the quantization noise is shown in Fig. 6.10.

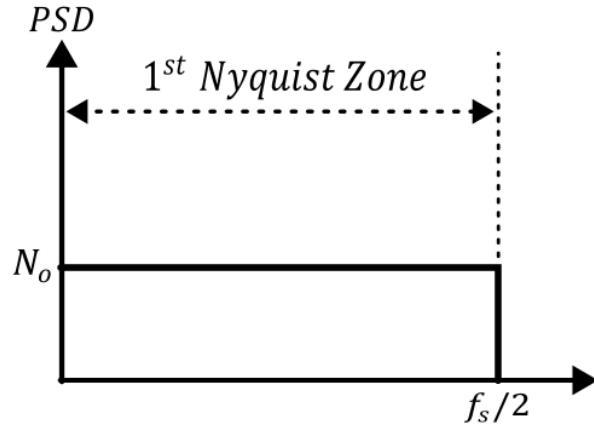


Fig. 6.10: quantization noise power spectral density

To find the single-sided spectrum value (N_o) of the noise PSD, the total noise power must be $LSB^2/12$ as calculated from Eqn. 6.13.

$$P_{Qnoise} = N_o \cdot \frac{f_s}{2} = \frac{LSB^2}{12} \rightarrow N_o = \frac{LSB^2}{6f_s} \quad 6.14$$

Where f_s is the ADC sampling frequency. In the case of a radar system, we are only interested in the integrated noise power within the FFT equivalent noise bandwidth (ENBW), the FFT resolution bandwidth when using a rectangular window [?]. The integrated quantization noise power is given by Eqn. 6.15.

$$P_{integrated_{Qnoise}}(W) = N_o \cdot ENBW = \frac{LSB^2}{6f_s} ENBW \quad 6.15$$

The signal-to-quantization noise ratio (SQNR) is then given by:

$$SQNR_{dB} = P_{in} + G_{dB} - 10 \log \left(\frac{\overbrace{\frac{LSB^2}{6f_s} ENBW * 10^3}^{P_{Qnoise}(mw)}}{} \right) \quad 6.16$$

In the following two sections, we will talk about the chosen architecture for the RF front end and do some MATLAB simulations to extract the required specifications on the RF blocks. The main idea behind all the simulations is that we want the target to be detectable. The target will be detectable if the received signal from it has an acceptable signal to noise, quantization noise, and intermodulation distortion ratio (SQNDR). The acceptable value of the SQNDR is determined by the required false-alarm rate P_{FA} and detection probability P_D [Chapter 0]. The total SQNDR for the signal received by one channel from a certain target is calculated as follows:

$$SQNDR_{dB} = -10 \log \left(10^{-\frac{SNR_{dB}}{10}} + 10^{-\frac{SDR_{dB}}{10}} + 10^{-\frac{SQNR_{dB}}{10}} \right), \quad 6.17$$

Where SNR_{dB} is the signal-to-noise ratio in dB as calculated from Eqn. 6.4, SDR_{dB} is the signal-to-intermodulation distortion ratio as calculated from Eqn. 6.10, and $SQNR_{dB}$ is the signal to quantization noise ratio as calculated from 6.16.

We will discuss two models. The first is the one-target model. In this model, only one target is assumed in the field of view of the radar, and we want that target to be detectable when it is within the detectable distance range of the radar. the second model is the two-object model. In this model, we assume two targets, one at the minimum distance while the other is at the maximum distance, and we want the far object to be detectable in the presence of the close one.

6.2. System architecture

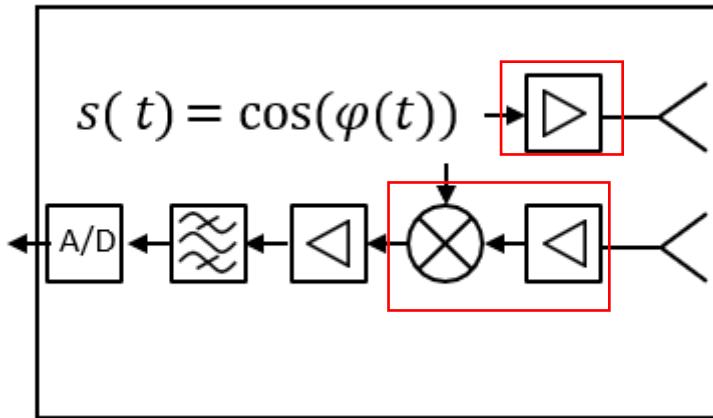


Fig. 6.11: Simplified radar transceiver architecture.

The radar system consists of three main subsystems: the transmitter, receiver, and signal processing. The transmitter is supposed to transmit the signal with high power without distorting the signal. In contrast, the receiver is supposed to maximize the SNR of the received signal by suppressing the unwanted signals, and the signal processing subsystem is considered to extract the required parameters by applying different algorithms based on the chosen waveform, as explained in [Chapter 2].

A simplified system architecture of a single-channel radar transceiver is shown in Fig. 6.11, where the signal generator feeds the Power Amplifier (PA), and the output signal is fed into the antenna to be transmitted. A portion of the transmitted signal is fed into the receiver as the local oscillator signal. On the receiver side, the received signal is received by the antenna and fed to a Low noise amplifier (LNA), then down-converted by a mixer. The down-converted signal is amplified, filtered, and finally applied to an ADC to be digitally processed.

6.3. System specifications

6.3.1. System budget

Received power range

To get the system specs for each block, first, the input power range is calculated using the well-known radar equation:

$$P_R = P_T \frac{\sigma G_T G_R \lambda^2}{(4\pi)^3 L_{ATM} R^4} \quad 6.18$$

Where R is the one-way distance between the transmitter and receiver, GT and GR are the gains of the transmitting and receiving antennas, respectively, PT and PR are the transmitting and receiving powers, respectively, λ is the wavelength, R is the target distance, L_{ATM} is the loss in the atmosphere, which at 77GHZ is around 0.3~0.5dB/km [3], σ is the radar cross-section of the target which is assumed $10m^2$ for passenger cars [4].

For a transmitted power of 10dBm and by assuming antenna gains of 20dBi, the received power range will be given by $P_R = -21 - 40 \log R$, hence for a detection range of 10m: 250m for a LRR, the received power is $-61dBm$: $-117dBm$.

IF frequency range

As explained in Chapter 2, for a chirp sequence waveform, the beat frequency in each chirp (fast time) is dominated by the range and less influenced by the velocity; hence the beat frequency will be given by:

$$f_{beat} = \frac{BW}{T_{chirp}} * \frac{2R}{C} \quad 6.19$$

So, assuming $T_{chip} = 100\mu s$ and chirp BW = 1GHz (long range) the beat frequency will range from 0.6667MHz to 16.6667MHz hence the minimum sampling frequency will equal 33.33MHz.

SNR calculation for chirp sequence FMCW

For one receive channel, the output SNR of the radar is calculated as follows:

$$SNR = P_{in} + 174 - 10 \log RBW - NF, \quad 6.20$$

Where P_{in} is the minimum received power, NF is the maximum acceptable noise figure for the system, and RBW is the equivalent noise bandwidth which, as explained in [Chapter 2], equals the bin width of the FFT.

$$RBW = \frac{1}{T_{chirp}} \quad 6.21$$

And as assumed $T_{chirp} = 100\mu s$ then $RBW = 10kHz$.

But by using the chirp sequence FMCW, there will be coherent integration which, as explained in Chapter 5, the SNR will improve compared to the single chirp case, and this improvement of the SNR equals $10\log(n)$ where n is the number of chirps in each frame for chirp sequence FMCW, this is equivalent to having smaller RBW.

$$RBW_{chirp\ sequence} = \frac{1}{nT_c} \quad 6.22$$

For our design, n is assumed to be 100, so the RBW equals 100Hz. Hence the SNR is calculated as follows:

$$SNR = P_{in} + 154 - NF \quad 6.23$$

6.3.2. Single object model

This model is based on assuming that there is only one object in the radar's field of view, and we want to detect this object with good SQNDR (>16dB). As explained in Eqn. 6.17, the SQNDR is the parallel combination of the SNR, SQNR, and SDR.

For minimum power (-117dBm), the SNR will dominate the total SQNDR, as noise will limit system performance. Hence at minimum power, the SNR should be greater than or equal to 16dB:

$$SNR = P_{in} + 154 - NF > 16dB \quad 6.24$$

Then the total NF is required to be less than or equal to 21dB. At maximum input power (-61dBm), the SDR will dominate the total SQNDR, as the system performance will be limited by nonlinearity, so for maximum power, the SDR should be greater than or equal to 16dB, with the assumption that the nonlinearity of the system is limited by the third order intermodulation product (IM_3) then the requirement on the total IIP3 for the system will be given by:

$$IIP3 = pin + \frac{SDR}{2} \quad 6.25$$

Then the required total IIP3 for the system should be greater than or equal -53dB.

The single object model is simple and is usually used in literature to get the system specs for the radar system. Still, it is not quite accurate as it does not consider the case that there is a far object and a near object. We want to detect both of them as the near object power will dominate the input power of the radar, which will result in poor SNDR for the far object power.

6.3.3. Two object model

This is a model that we assumed, which is more complex than the single-object model. For this model, the lineup analysis is done such that no matter how close the first object is, we should be able to spot the far object as the near object power will dominate the radar's input power, resulting in poor SNDR for the far object power. The model is applied for the worst case where the minimum power (-117dBm) is received simultaneously with the maximum power (-61dBm), and we want to receive both of them with good SNDR.

As the maximum power will dominate the input power, we will calculate the noise and distortion for this maximum power and calculate how the SNDR of the minimum power will be affected as follow:

$$SNR_{dB}|_{min\ power} = SNR_{dB}|_{max\ power} + P_{min,dBm} - P_{max,dBm} \quad 6.26$$

$$SDR_{dB}|_{min\ power} = SDR_{dB}|_{max\ power} + P_{min,dBm} - P_{max,dBm} \quad 6.27$$

By ensuring that the minimum input power is received with good SNDR for this case, any other case will have better SNR as this is the worst case, and this condition is checked by applying lineup analysis on MATLAB.

6.3.4. MATLAB simulation

In our receiver, we have two blocks (the LNA and the Mixer); the lineup is done to get the limits of the specs of each block such that enhancing this spec more than a certain value will not affect the total SNDR of the minimum signal when received with the maximum signal. To perform the lineup analysis, we needed to assume initial specs for each block, and to do that, we have done simple implementations on cadence for each block. Also, we searched the literature to investigate the possible specs that can be achieved at this frequency range with the available technology. Then after assuming these specs, we swept each spec to get the limit, after which this spec will not affect the total SNR.

6.3.4.1. LNA specs

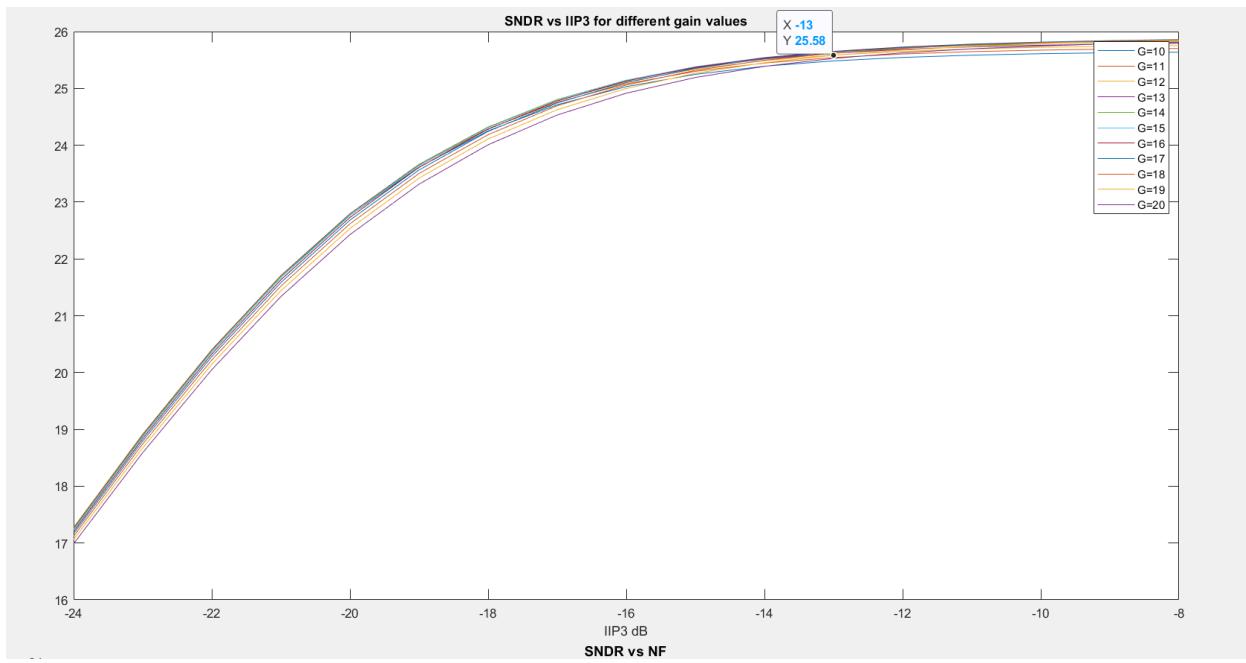


Figure 6.12: SNDR vs IIP3 for different gain values for the LNA

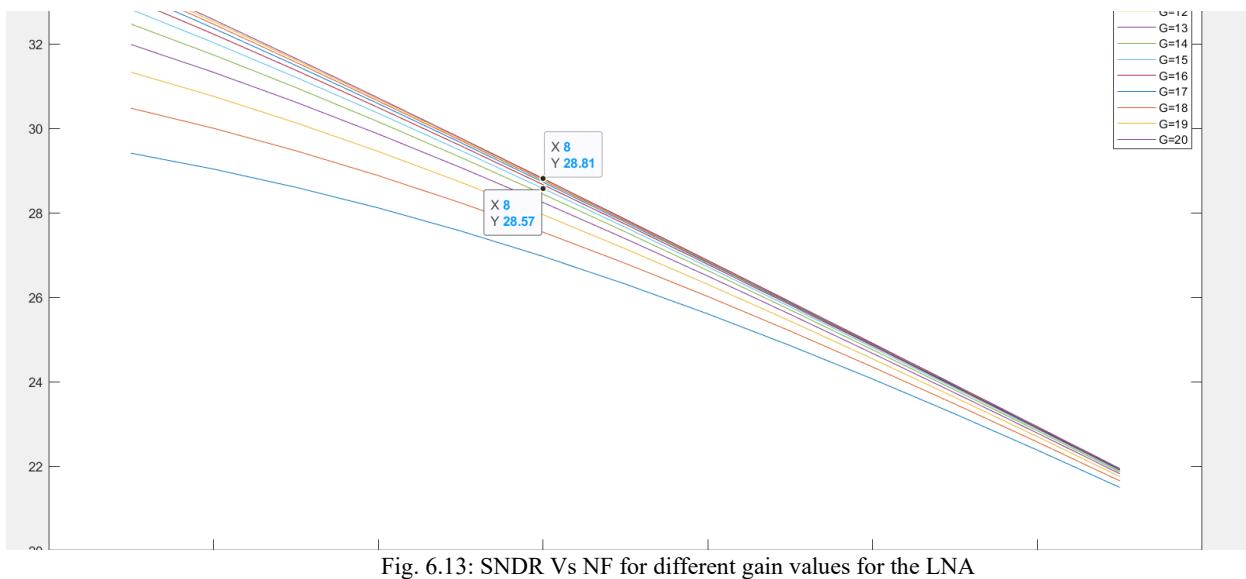


Figure 6.13: SNDR vs NF for different gain values for the LNA

As shown in Fig. 6.12, the SNDR for the minimum power is not affected by increasing IIP_3 of the LNA above a certain value (-13dBm), so this value will be taken as the linearity spec for the LNA.

Fig. 6.13 also shows that the SNDR value is affected by the LNA's NF and gain. But, for the gain, it seems for gain values greater than 15dB, the SNDR does not increase significantly, so the gain spec for the LNA will be 15dB; for the NF, it affects the SNDR, so we will take a reasonable value for the NF that is close to the values achieved in literature and gives high SNR. The NF spec for the LNA is chosen to be 8dB

6.3.4.2. Mixer specs

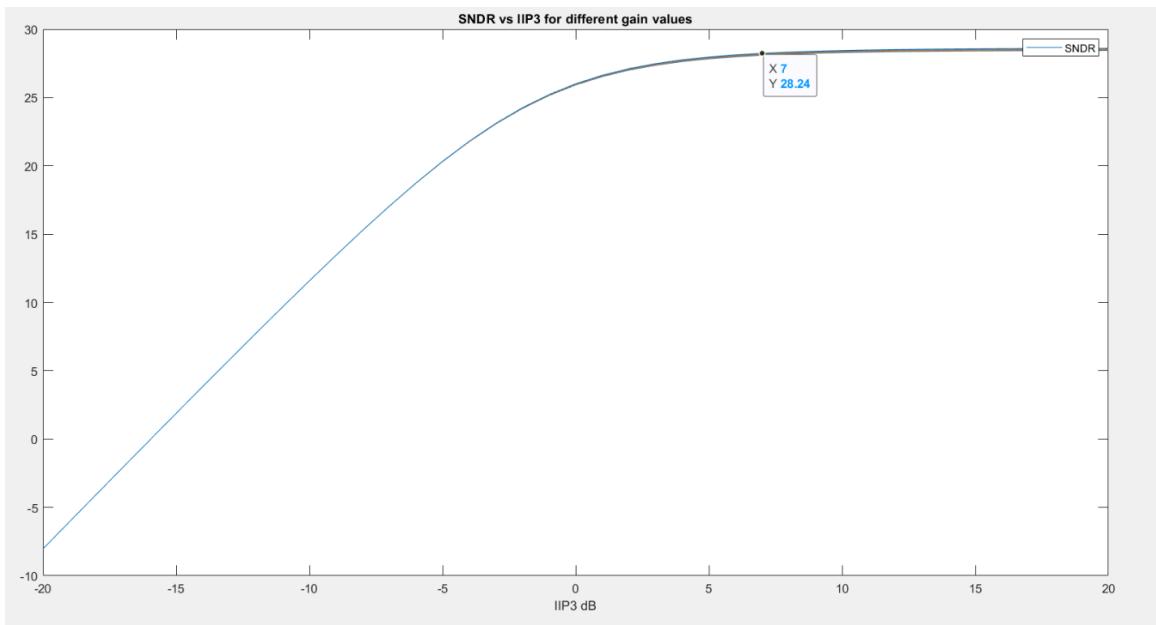
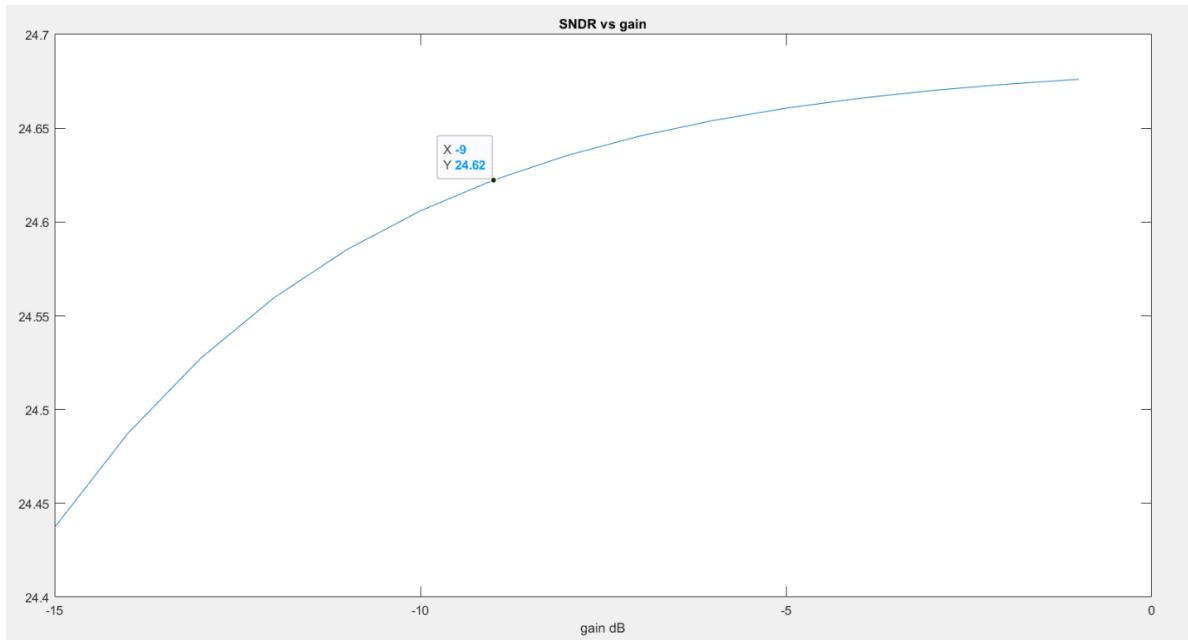


Fig. 6.14: SNDR Vs IIP_3 for different gain values for the MIXER



As shown in Fig. 6.14, the SNDR is not affected by increasing IIP_3 of the MIXER above a certain value

Fig. 6.15: SNDR Vs Gain of the MIXER

(7dBm), so this value will be taken as the linearity spec for the MIXER. Fig. 6.15 also shows that the SNDR value is affected by the MIXER's gain so we will take a reasonable value for the NF that is close to the values achieved in the literature and gives high SNR. The gain spec for the MIXER is chosen -9dB. The targeted output power of the PA is 10dBm. Table 6.1 shows the obtained specs for the LNA and mixer.

Table 6.1: LNA and Mixer specs

Spec	LNA	Mixer
Gain	$> 15dB$	$> -9dB$
NF	$< 8dB$	$< 9dB$
IIP3	$> -14dBm$	Type equation here.

References

- [1]. Razavi, B. (2014) ‘BASIC CONCEPTS IN RF DESIGN’, in *RF Microelectronics*. Inde: Dorling Kindersley.
- [2]. Oppenheim, A.V. and Schafer, R.W. (2014) ‘Digital Processing of Analog Signals’, in *Discrete-time signal processing*. 3rd edn. Harlow: Pearson.
- [3]. J. Hatchet al., “Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band,” *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 3, pp. 845–860, Mar. 2012.
- [4]. “A Review of Automotive Radar Systems-Devices and Regulatory Frameworks,” document SP 4/01, Spectrum Planning Team, Australian Communication Authority, 2001

7

LOW NOISE AMPLIFIER

A low noise amplifier (LNA) is considered the most important block in the receiver front end that has a critical effect on the overall performance of the system. The main function of the LNA is to provide enough gain to minimize the contribution of the subsequent blocks, such as a mixer, to the total noise figure of the receiver chain. As the first active block in the receiver chain, LNA contributes directly to the overall receiver noise figure; therefore, it should amplify incoming signals while adding as little noise as possible. An LNA should have high gain, low noise figure, suitable linearity to accommodate large signals without distortion, high reverse isolation to prevent signals from leaking back to the antenna, good input and output return loss, and it should be stable. Some of these metrics are discussed in the following sections.

7.1. Noise

Noise is a random process; the noise value at any time instant is a random variable. However, noise is deterministic in the frequency domain and has its power distributed across different frequencies in a certain shape known as the noise power spectral density (PSD). The main reason for defining the PSD is that it allows many of the frequency-domain operations used with deterministic signals to be applied to random signals as well. For example, When a noise signal with a PSD of $S_x(f)$ is applied to a linear, time-invariant system with a transfer function $H(s)$, then the output spectrum is

$$S_y(f) = S_x(f)|H(f)|^2. \quad 7.1$$

We can empirically measure the PSD of noise signals coming from different noise sources in an electronic device. Once the power spectral densities are known, we can do noise analysis for a circuit to know the total input or output noise. Following are some of the noise sources in electronic circuits [1].

7.1.1. Resistor thermal noise

The resistor's thermal noise is considered to be white noise; has a flat power spectral density across all frequencies, up to very high frequencies. This means that the generated thermal noise has uncorrelated samples. The thermal noise can be modeled as either a series random voltage source with a PSD of $\overline{v_n^2} = 4kTR$ or as a parallel random current source with a PSD of $\overline{I_n^2} = 4kT/R$, where k is Boltzmann's constant, and T is the absolute temperature in kelvin. Different resistors' thermal noise representations are shown in Fig. 7.1.

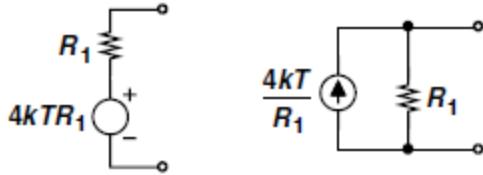


Fig. 7.1: different methods of modeling the resistor's thermal noise.

7.1.2. MOSFET noise sources

7.1.2.1. Transistor thermal noise

For a MOS transistor operating at the saturation region, the thermal noise of the channel can be modeled as either a random current source tied between the source and drain terminals with a PSD $\bar{I_n^2} = 4kT\gamma g_m$ or a voltage source with a PSD of $\bar{v_n^2} = 4kT\gamma/g_m$, as shown in Fig. 7.2.

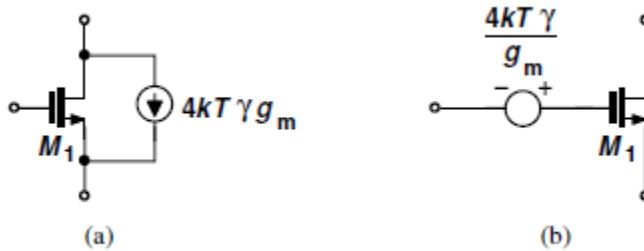


Fig. 7.2: a) Transistor's thermal noise as a current source, b) series voltage source model

Where g_m is the transistor's transconductance, and γ is a coefficient in the range of $2/3$ to 2 and is usually obtained by measurements. Another component of thermal noise in MOSFETs arises from the gate resistance, which becomes more important with the scaling down of the transistor's size.

7.1.2.2. Transistor flicker noise

MOS devices also suffer from a colored noise source called the “flicker” or “ $1/f$ ” noise. Flicker noise is dominant at low frequencies and is modeled by a random voltage source series with the gate with the following PSD:

$$\bar{v_n^2} = \frac{K}{C_{ox}WL} \frac{1}{f}, \quad 7.2$$

Where K is a process-dependent constant.

We can also model flicker noise by a random current source between the drain and the source of the transistor. This is possible because a small-signal voltage source in series with the gate is multiplied by the transistor's transconductance g_m and is transferred into a current between the drain and the source. Thus, the PSD of the flicker noise, when modeled as a current source, is

$$\bar{i_n^2} = g_m^2 \frac{K}{C_{ox}WL} \frac{1}{f}. \quad 7.3$$

For a certain device size and biasing, the flicker noise PSD intercepts with the thermal noise PSD at a certain frequency, known as the flicker noise corner frequency f_c . Shown in Fig. 7.3.

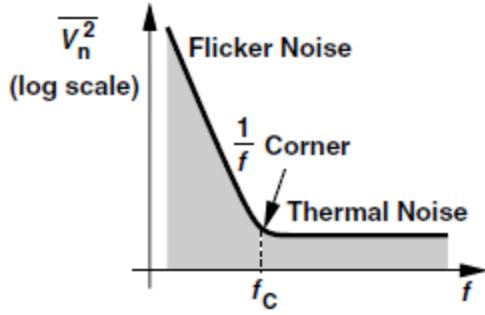


Fig. 7.3: Flicker noise corner frequency.

To find f_c , we equate the thermal noise voltage PSD with the flicker noise voltage PSD as follows:

$$\frac{K}{C_{ox}WL} \frac{1}{f_c} = \frac{4kT\gamma}{g_m}. \quad 7.4$$

It follows that

$$f_c = \frac{K}{C_{ox}WL} \frac{g_m}{4kT\gamma}. \quad 7.5$$

In today's technologies, the corner frequency falls into the range of tens or hundreds of megahertz. When operating beyond the corner frequency, the thermal noise is dominant, and we neglect the flicker noise.

7.1.3. Noise Figure

We have discussed the definition of the noise factor and noise figure in [Chapter 6], and we have also seen how to derive the noise factor of a block. However, we modeled the noise of the block as a single random voltage source referred to the output or the input of the block. At high frequencies, this model of the block noise is not sufficient, and the input-referred noise of the block is modeled as a series voltage source $\overline{v_n^2}$ and a parallel current source $\overline{i_n^2}$ [2], as shown in Fig. 7.4.

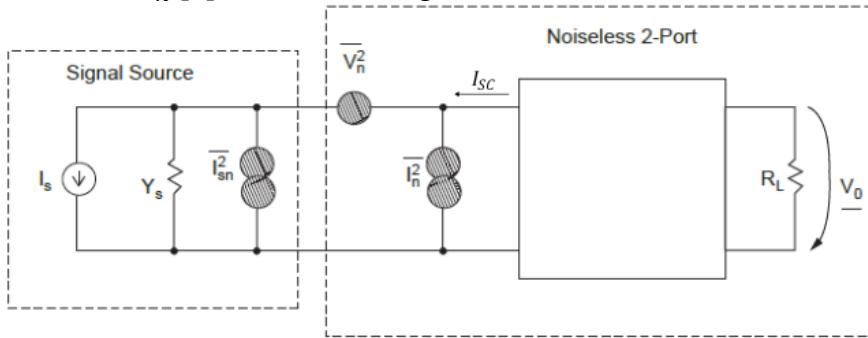


Fig. 7.4: schematic used for the derivation of the noise factor for a noisy two-port network at high frequencies.

$\overline{v_n^2}$ is obtained by shorting the input ports of the noisy network, calculating the output-referred noise PSD and dividing it by the voltage gain. $\overline{i_n^2}$ is obtained by leaving the input ports of the noisy network open, calculating the output-referred noise PSD and dividing it by the trans-resistance gain of the network.

For low-frequency analog circuits with $Z_{in} \rightarrow \infty$, i_n does not find a path to flow into, and the noise is only

modeled by the series voltage source. However, for high-frequency networks, both sources must be considered.

We can calculate the noise factor F of the network using the following equation:

$$F = \frac{\overline{i_{sc,tot}^2}}{\overline{i_{sc,s}^2}}, \quad 7.6$$

Where $\overline{i_{sc,tot}^2}$ is the PSD of the total short circuit noise current while $\overline{i_{sc,s}^2}$ is the PSD of the short circuit noise current due to the signal source only. From Fig. 7.4, we have

$$i_{sc,tot} = i_n + \frac{v_n + v_{sn}}{Z_s}, \quad i_{sc,s} = i_{sn} = \frac{v_{sn}}{Z_s}, \quad 7.7$$

Where v_{sn} is the signal source noise voltage, which has a PSD of $\overline{v_{sn}^2} = 4kTR_s$. For simplicity, let's first assume that the noise sources of the network, i_n and v_n , are uncorrelated, then it follows that

$$\overline{i_{sc,tot}^2} = \overline{i_n^2} + \frac{\overline{v_n^2} + \overline{v_{sn}^2}}{|Z_s|^2}, \quad \overline{i_{sc,s}^2} = \frac{\overline{v_{sn}^2}}{|Z_s|^2}. \quad 7.8$$

If we substitute with Eqn. 7.8 into Eqn. 7.6, we obtain the following expression for the noise factor

$$F = 1 + \frac{\overline{v_n^2}}{4kTR_s} + \frac{(R_s^2 + X_s^2)\overline{i_n^2}}{4kTR_s}. \quad 7.9$$

We notice that the noise factor of the two-port network is a function of its noise sources and the signal source impedance, $Z_s = R_s + jX_s$. There is an optimum source impedance that will minimize the noise factor, and we can find it as follows:

$$\frac{\partial F}{\partial R_s} = 0, \quad \frac{\partial F}{\partial X_s} = 0 \rightarrow R_{s,opt} = \sqrt{\frac{\overline{v_n^2}}{\overline{i_n^2}}} \text{ and } X_{s,opt} = 0. \quad 7.10$$

If we substitute with the optimum source impedance into Eqn. 7.9, we get the following expression for the minimum noise factor

$$F_{min} = 1 + \frac{\sqrt{\overline{v_n^2}\overline{i_n^2}}}{2kT}. \quad 7.11$$

If the noise sources of the network are correlated, we can use the IEEE two-port noise parameters to derive the noise factor [2]. In this case, we assume that the network noise current has an uncorrelated component i_u and a correlated component $i_c = Y_{cor}v_n$, where Y_{cor} is the correlation admittance and is defined as

$$Y_{cor} = \frac{\overline{i_n v_n^*}}{\overline{v_n^2}} = G_{cor} + jB_{cor}. \quad 7.12$$

The noise factor, in this case, is given by

$$F = 1 + \frac{\bar{l}_u^2}{l_{sn}^2} + \frac{\bar{v}_n^2}{l_{sn}^2} |Y_{cor} + Y_s|^2. \quad 7.13$$

After some simplification, we can write the noise factor as follows:

$$F = F_{min} + \frac{R_n}{G_s} |Y_s - Y_{sopt}|^2, \quad 7.14$$

Where $R_n = \bar{v}_n^2 / 4kT$ is the noise resistance, $Y_s = G_s + jB_s$ is the source admittance, $Y_{sopt} = G_{sopt} + jB_{sopt}$ is the optimum source admittance that minimizes the noise factor, and F_{min} is the minimum noise factor and is given by

$$F_{min} = 1 + 2R_n(G_{cor} + G_{sopt}). \quad 7.15$$

Eqn. 7.14 shows that the value of R_n indicates the sensitivity of the noise figure to a mismatch between the source impedance and the optimum source impedance; if R_n is large, then a little deviation from the optimum impedance will cause the noise factor to increase considerably above its minimum value.

To summarize, We notice that F_{min} and $Z_{s,opt}$ depend only on the internal noise sources of the two-port network. The required source impedance to achieve the minimum noise factor is different from that of the maximum power transfer. In other words, matching for noise is different from the gain match.

For a MOSFET device in a common-source configuration, F_{min} and $Z_{s,opt}$ are given by Eqn. 7.16 and Eqn. 7.17, respectively.

$$F_{min} \approx 1 + \frac{2f}{f_T} \sqrt{k_1} \sqrt{g_m(R_s + R_g) + 1}, \quad 7.16$$

$$Z_{s,opt} \approx \frac{1}{\omega(C_{gs} + C_{gd})} \left[\sqrt{\frac{g_m(R_s + R_g)}{k_1}} + j \right], \quad 7.17$$

Where k_1 is a technology-dependent constant and f_T is the unity current gain frequency. For a certain frequency, there is an optimum current density (J_{opt}) that minimizes the value of F_{min} . Therefore, the first step in designing an LNA is that we bias the transistors at this optimum current density. As shown in Fig. 7.5, this optimum current density itself is a function of frequency and typically increases with frequency. It is also clear that the F_{min} increases with frequency, which makes it more challenging to make low-noise amplifiers at very high frequencies.

Eqn. 7.17 is interesting because it shows that we can control the real part of the optimum noise impedance by changing the size of the transistor. Thus, we can size the device such that $Re\{Z_{s,opt}\} = 50\Omega$, which reduces the matching network required at the input to match the antenna impedance to $Z_{s,opt}$ to a series inductor improving the LAN noise figure. This is known as active noise matching.

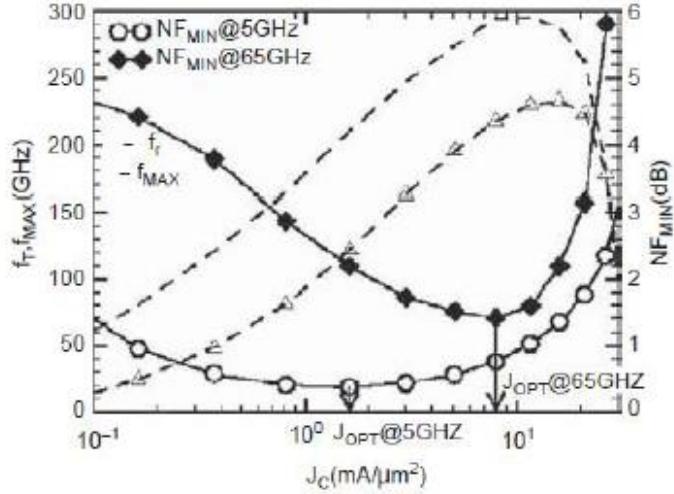


Fig. 7.5: f_{\min} , f_T , and f_{\max} VS current density.

7.1.4. Constant Noise Circles

As discussed in the previous subsection, to achieve the minimum noise figure from an LNA, we need to design a passive matching network to transform the driving antenna impedance, typically 50Ω , to the optimum noise impedance, Z_{opt} . This is not always possible because if we want to achieve maximum power transfer between the antenna and the LNA, we transform the 50Ω to the complex conjugate of the LNA input impedance, Z_{in}^* . When designing an LNA, we try to keep the optimum noise impedance as close as possible to the impedance that achieves the maximum power transfer. This will help with the simultaneous match for gain and noise. When simultaneous matching is not possible, we match to an intermediate point that achieves an acceptable compromise between the noise and input return loss.

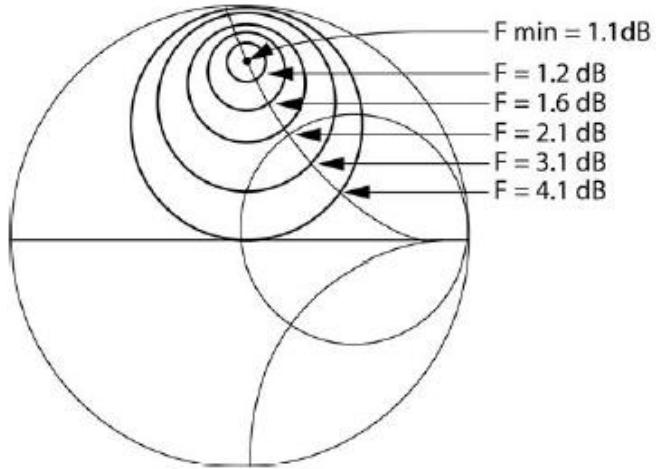


Fig. 7.6: Constant noise circles

If we are willing to deviate by ΔF from F_{min} , one can find the center (c_i) and radius (r_i) of a constant noise circle on the Smith chart; matching the antenna impedance to any point on that circle leads to a noise factor higher than the minimum noise factor by ΔF . Constant noise circles are shown in Fig. 7.6. Similar gain circles can be drawn on the Smith chart to help find the sweet spot that achieves an acceptable noise figure and gain.

7.2. Gain

According to the Friss equation (6.3), the noise contribution of each stage in a cascade of noisy blocks is divided by the gain of the preceding stages. This makes the gain of the LNA of special importance in the receiver chain, as it should be high enough to suppress the noise contribution of the subsequent stages. However, high gain stages at the beginning of the receiver degrade the linearity performance of the subsequent stages due to the increasing signal power through the chain. High signal power levels may cause the blocks to produce high levels of intermodulation distortion. This is a well-known tradeoff between the noise figure and the linearity performance of the receiver. A good compromise can be reached by doing a line-up analysis and making sure that the received signal SNDR is acceptable for high and low input powers.

To achieve the highest gain possible from a two-port network, as the one shown in Fig. 7.8, simultaneous complex conjugate matching at the input and output is essential for maximum power transfer ($Z_S = Z_{in}^*$ and $Z_L = Z_{out}^*$). A little mismatch can be tolerated without reducing the gain by a great amount. We can use source and load mismatch gain circles to find good matching points that will not reduce the gain that much. Combined with the noise circles, one can find a good input-matching point that achieves good gain and noise performance.

7.3. Linearity

Just like any other block, the linearity of the LNA is characterized by the 1-dB compression point, third-order, and second-order intercept points, referred to the LNA input or output. For a degenerated common source stage as the one shown in Fig. 7.7, assuming a large-channel square law device, we can find the output current I_D as a function of the input voltage V_{in} . By implicit differentiation of the obtained relation, we can derive the non-linearity coefficients α_1 and α_3 [1].

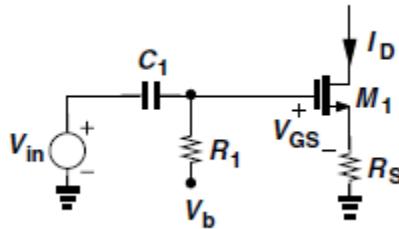


Fig. 7.7: A degenerated common source stage.

From α_1 and α_3 , one can derive an expression for the third-order intercept point as follows:

$$A_{IIP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} = \sqrt{\frac{2g_m}{3R_s} \frac{(1 + g_m R_s)^2}{K}}, \quad 7.18$$

Where $K = (1/2)\mu_n C_{ox}(W/L)$, and g_m is the transconductance at the bias point. This result shows that the linearity of common-source configurations is limited by the overdrive voltage of the main device. In the case of a short-channel undegenerated CS stage, the A_{IIP3} increases with the overdrive voltage until it reaches a maximum at a certain point before degrading again.

7.4. Stability

When designing an LNA, stability must be guaranteed across all frequencies, even out of band, and source impedances. In this section, we discuss some of the stability considerations of a two-port network, such as stability circles and factors.

Consider a two-port network characterized by S parameters measured in a Z_o Ohm system is driven by a source with an internal impedance of Z_s and loaded by a load impedance Z_L , as shown in Fig. 7.8.

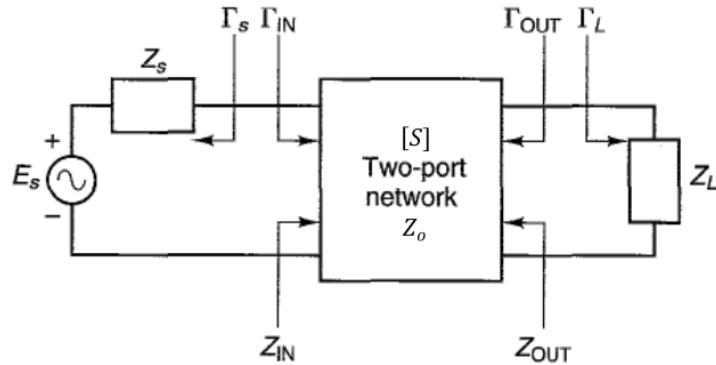


Fig. 7.8: Stability of a two-port network.

The input reflection coefficient (Γ_{in}) is defined as follows [3]

$$\Gamma_{in} = \frac{Z_{in} - Z_o}{Z_{in} + Z_o} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - \Gamma_L S_{22}}, \quad 7.19$$

Where Γ_L is the load reflection coefficient and is defined as $(Z_L - Z_o)/(Z_L + Z_o)$. We notice that the input reflection coefficient is a function of the S parameters of the network and the load impedance. The output reflection coefficient Γ_{OUT} is defined in a similar way

$$\Gamma_{OUT} = \frac{Z_{OUT} - Z_o}{Z_{OUT} + Z_o} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - \Gamma_S S_{11}}, \quad 7.20$$

Where Γ_S is the source reflection coefficient and is defined as $(Z_s - Z_o)/(Z_s + Z_o)$. Once again, the output reflection coefficient is a function of the network parameters and the source impedance.

Oscillations happen when the two-port network presents a negative resistance at either the input or output ports. In terms of reflection coefficients, this occurs when $|\Gamma_{in}| > 1$ or $|\Gamma_{OUT}| > 1$. At a certain frequency, the network is said to be unconditionally stable if the real parts of Z_{in} and Z_{OUT} are not negative for all possible passive load and source terminations. If the network is not unconditionally stable, this means that some passive load and source terminations can produce input and output impedances having negative real parts. Therefore, the conditions for unconditional stability at a certain frequency are

$$|\Gamma_{in}| < 1 \rightarrow \left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - \Gamma_L S_{22}} \right| < 1, \forall \text{passive } Z_L \quad 7.21$$

$$|\Gamma_{OUT}| < 1 \rightarrow \left| S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - \Gamma_S S_{11}} \right| < 1, \forall \text{ passive } Z_S \quad 7.22$$

When the network is potentially unstable, there may be values of Γ_L and Γ_S (i.e., source and load impedances) for which the real parts of Z_{in} and Z_{OUT} are positive. These values of Γ_L and Γ_S (i.e., regions on the Smith chart) can be determined using the stability circles at this frequency.

The load (output) stability circle is a circle on the Smith chart of Γ_L values for which $|\Gamma_{in}| = 1$ that represents the boundary between the stable and unstable regions of Γ_L . Similarly, the source (input) stability circle is a circle on the Smith chart of Γ_S values for which $|\Gamma_{OUT}| = 1$ that represents the boundary between the stable and unstable regions of Γ_S .

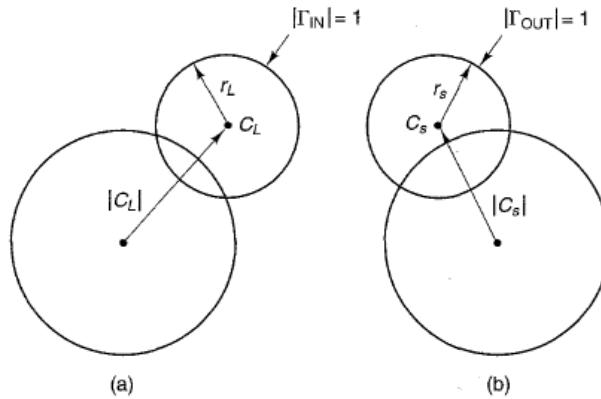


Fig. 7.9: a) Load stability circle b) Source stability circle.

The radius and center of either the output or input stability circles are functions of the S parameters of the two-port network [3]. A necessary and sufficient condition for the unconditional stability of a single-transistor amplifier is

$$K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}S_{21}|} > 1, \quad 7.23$$

and

$$B_{1f} = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 > 0. \quad 7.24$$

Where K is known as Rollet's stability factor and $\Delta = S_{11}S_{22} - S_{12}S_{21}$.

The stability test in Eqns. 7.7. cannot be extended to a chain of several cascaded devices to two-port networks. This condition becomes necessary but not sufficient in the case of cascaded stages, as we have to check that there is no transistor presenting a negative resistance anywhere in the circuit. Larger values of K or B_{1f} do not indicate greater stability. There is another stability factor defined as follows:

$$\mu = \frac{1 - |S_{11}|^2}{|S_{22} - \Delta S_{11}^*| + |S_{12}S_{21}|} > 1. \quad 7.20$$

Larger values of μ imply greater stability which makes it very useful.

7.5. Cascode LNA and Balun Design

7.5.1. Target Specifications

In [Chapter 6], we determined the required specifications for the receiver RF front-end. Assuming a two-object model, we obtained the specifications for the LNA and mixer such that the signal received from the far object has the best signal-to-noise and distortion SNDR ratio. The values of the obtained SNDR were too optimistic because we did not consider the analog baseband, which will limit the overall system linearity causing less SNDR. However, the obtained specifications give the best performance out of the RF front-end.

A literature survey was also conducted to revise the obtained specs of the RF front-end in published papers implementing 76-81GHz automotive radar systems in 65nm CMOS technology. The literature survey comparing other works with the achieved specs of the proposed LNA is shown at the end of this section.

Table 7.1 summarizes the targeted LNA specifications:

Table 7.1: LNA target specifications.

Spec	Target
Gain	$> 14dB$
Noise Figure	$< 10dB$
IIP3	$> -14dBm$
IP1dB	$> -24dBm$
S11	$< -10dB$
S22	$< -10dB$

7.5.2. Topology Selection and Design Choices

We have implemented two topologies for the LNA. In this section, we are investigating the first topology, which is a single-ended, three-stage common source cascode low noise amplifier, shown in Fig. 7.10. In [Section 7.6], the second topology (a four-stage common source) is presented with a comparison between the achieved specs by both topologies. Under the same supply (1.2V), the cascode configuration achieves a higher maximum power gain, better isolation between the input and the output, and is more stable compared to the common source stage. A drawback of this topology is the higher noise figure compared to a common source stage due to the increased number of noise contributors. A cascode stage could be seen as a common source followed by a common gate stage. Intra-stage matching between the CS and CG can improve the performance and resonate out the parasitic capacitance at the internal node [4]. Subtle layout techniques such as shared junction layout [5] can help minimize the capacitance at the internal node improving the performance at high frequencies.

The three stages are identical and biased at a current density of $220\mu A/\mu m$, which is close to the current density for optimum noise performance and achieves a good gain performance simultaneously. The optimal width per finger to maximize the f_{max} is $1\mu m$. The total width of the transistors is $24\mu m$ divided into 2 multipliers with 12 fingers each. Micro-strip transmission lines and metal-insulator-metal (MIM) capacitors are used for inter-stage, input, and output matching networks. The input and output are matched to 50Ω to ease the integration with the mixer. The micro-strip line is made of a signal line in the M9 layer and a ground plane in M2. We used M2 for the ground instead of M1 because it is thicker and provides

better isolation from the substrate. We can also use a stack of M1 and M2 shunted through vias to form a thicker ground plane with no penetrating slots [5]. The characteristic impedance of the used transmission lines is 50Ω which corresponds to a signal line width of $6\mu m$. The measured loss of the transmission lines is about $1.2dB/mm$. A transformer-based balun is used at the output of the LNA to perform the single-ended to differential conversion for the passive mixer.

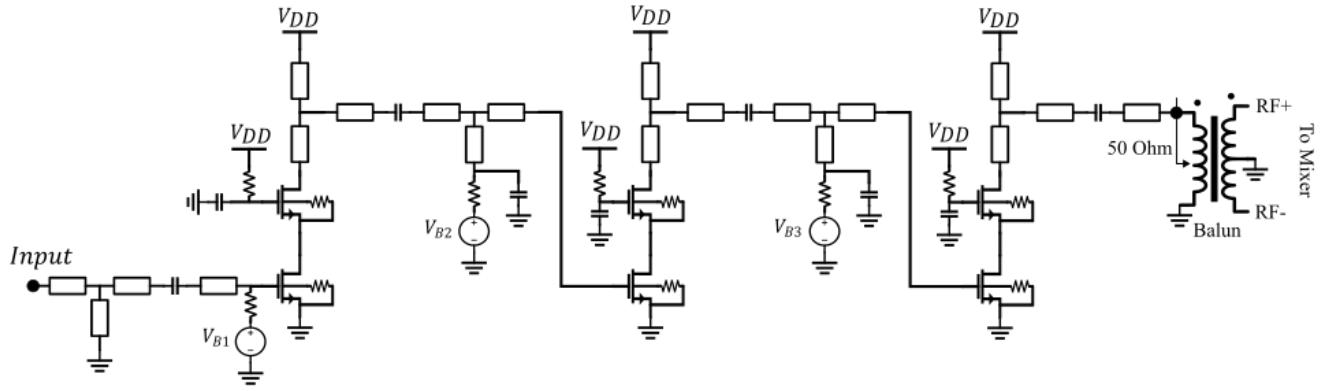


Fig. 7.10: Schematic of the three-stage cascode LNA.

7.5.3. Post-Layout simulation results

The following simulation results were obtained after EMing the full layout of the circuit on Keysight's Advanced Design System (ADS) and performing parasitic extraction (PEX) for the transistors. The LNA was simulated at the nominal corner, then across different corners. First, we present the typical results.

7.5.3.1. Gain

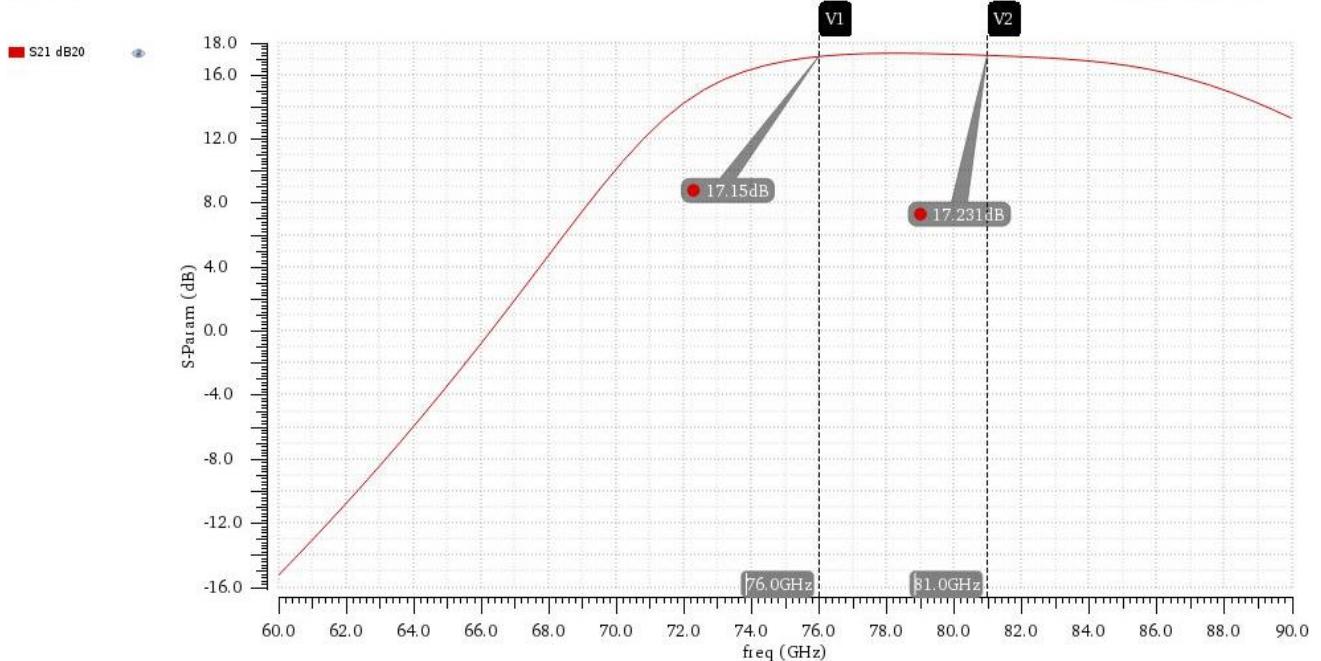


Fig. 7.11: Simulated Gain ($S21$) VS frequency.

As shown in Fig. 7.11, the simulated gain of the LNA is about $17dB$ across the range $76 – 81GHz$ with a gain flatness of about $0.2dB$, which satisfies the required gain spec. The problem with this design is that the gain does not roll off quickly out of the band, which makes the system prone to out-of-band blockers. We can use a band pass filter at the beginning of the receiver to reject out-of-band emissions. This kind of filter would be cheap and easy to integrate because of the low required fractional bandwidth. However, by tuning the matching networks, we obtain another design that trades the flatness with a better out-of-band rejection. The gain of that other LNA design is shown in Fig. 7.12.

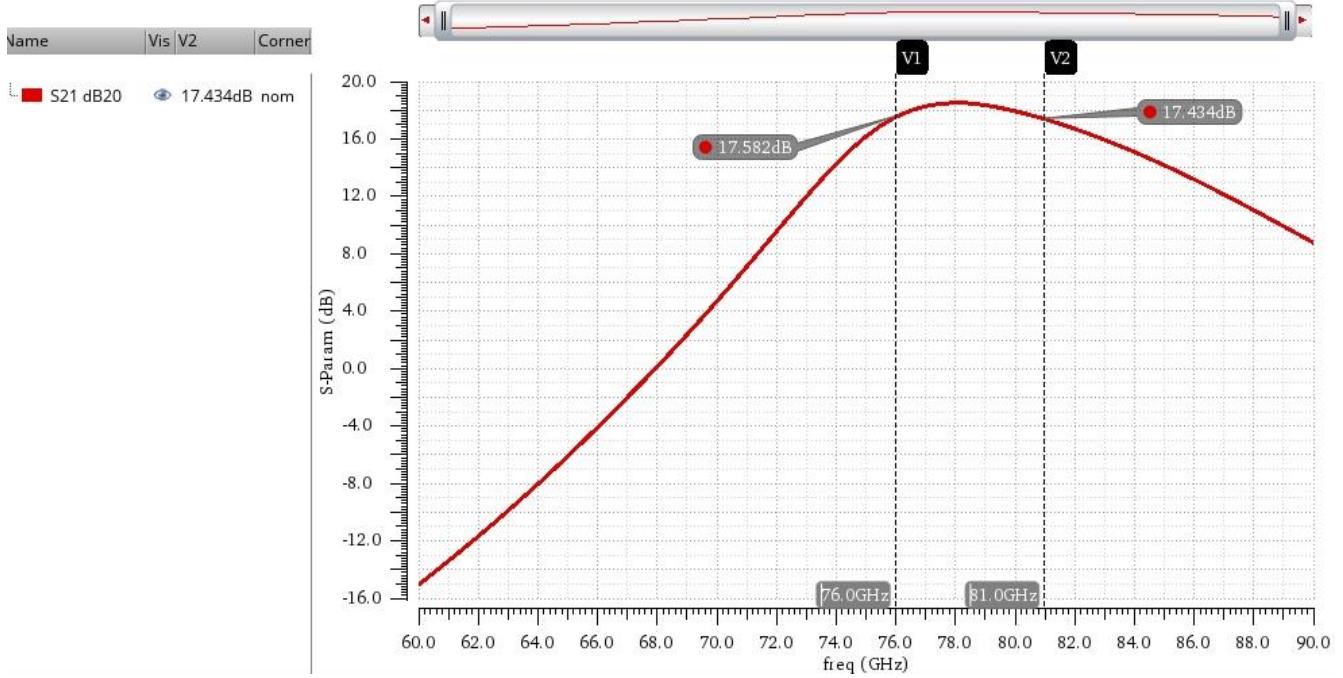


Fig. 7.12: The simulated gain of an alternative LNA design.

Both designs are very similar in other specs and have full layout EMed using ADS. The alternative design has poor gain flatness across the band. The gain peaks at $78GHz$ and reaches about $18.5dB$, which corresponds to a $1dB$ gain variation across the full band. The band $76 – 77GHz$ for the short-range detection does not suffer from much gain variation. However, the long-range band $77 – 81GHz$ suffers from the $1dB$ gain variation. Gain variations might cause the received signal to have a non-constant envelope, which may cause the beat signal to spread more in the frequency domain. All the presented results in this subsection are for the first design.

7.5.3.2. Noise Figure

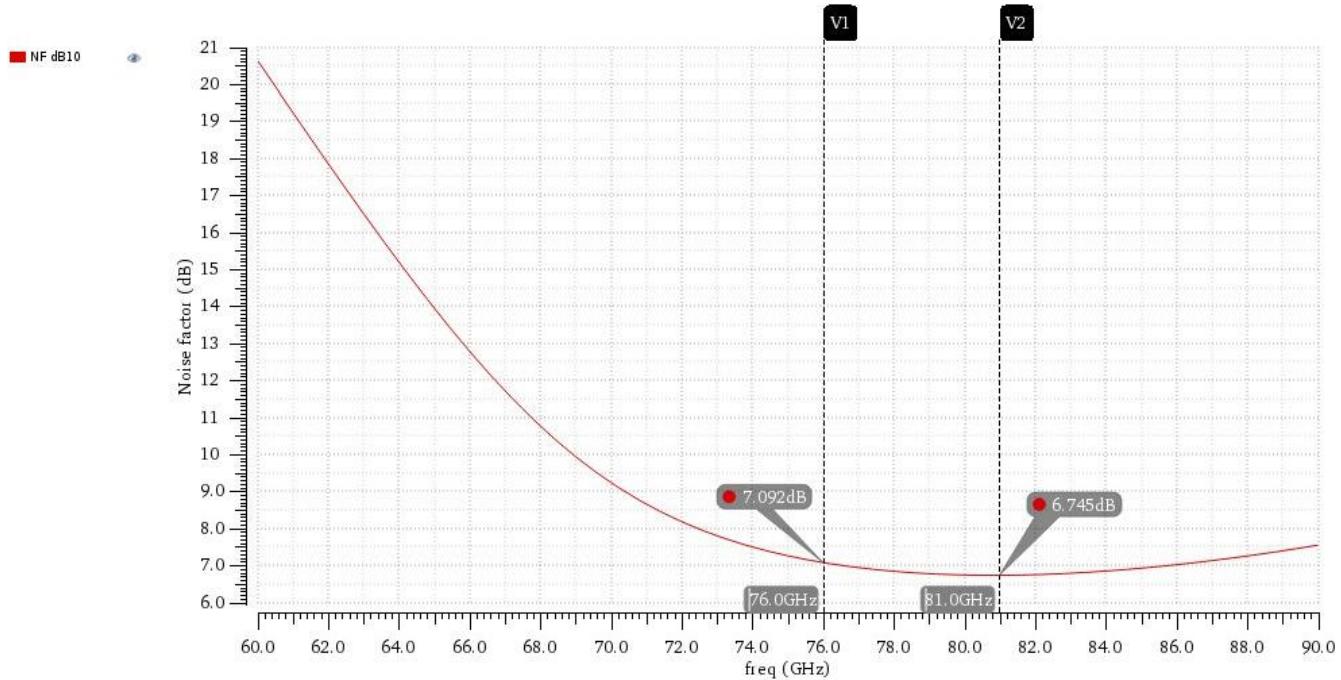
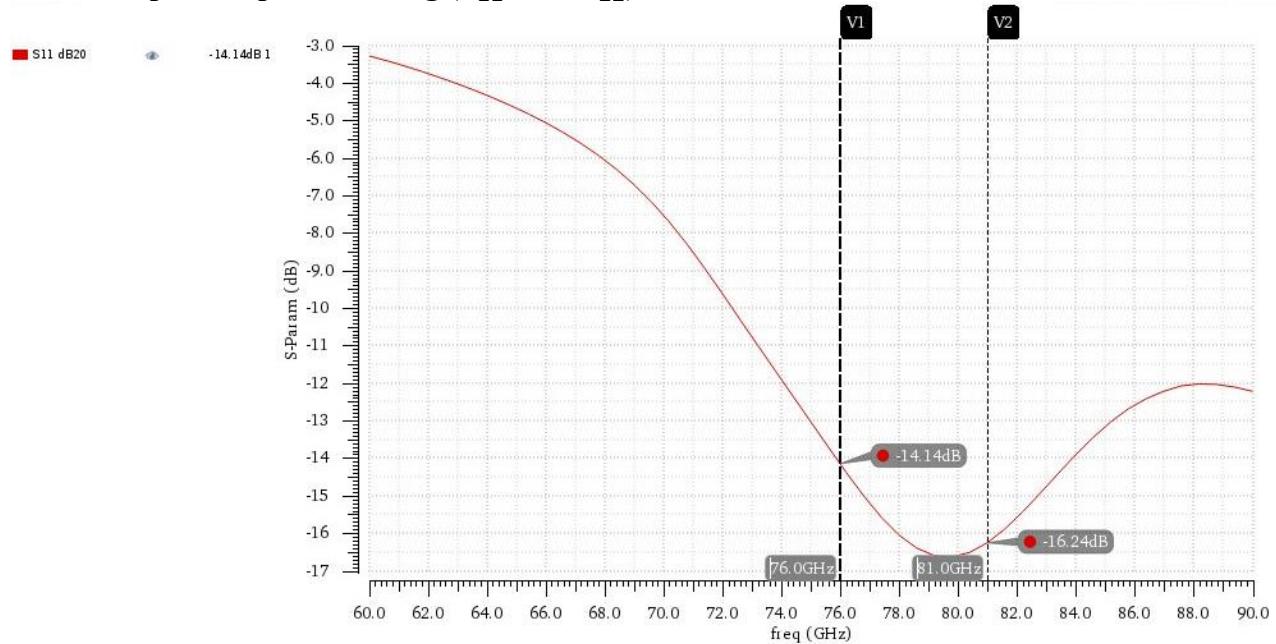


Fig. 7.13 shows the LNA noise figure versus frequency. The lowest noise figure achieved across the band 76 – 81GHz is about 6.7dB which satisfies the target spec ($NF < 10$).

7.5.3.3. Input/Output matching (S_{11} and S_{22})



As shown in Fig. 7.14, the input return loss (S_{11}) is below $-10dB$ for the full range with more skew towards high frequency to achieve gain flatness.

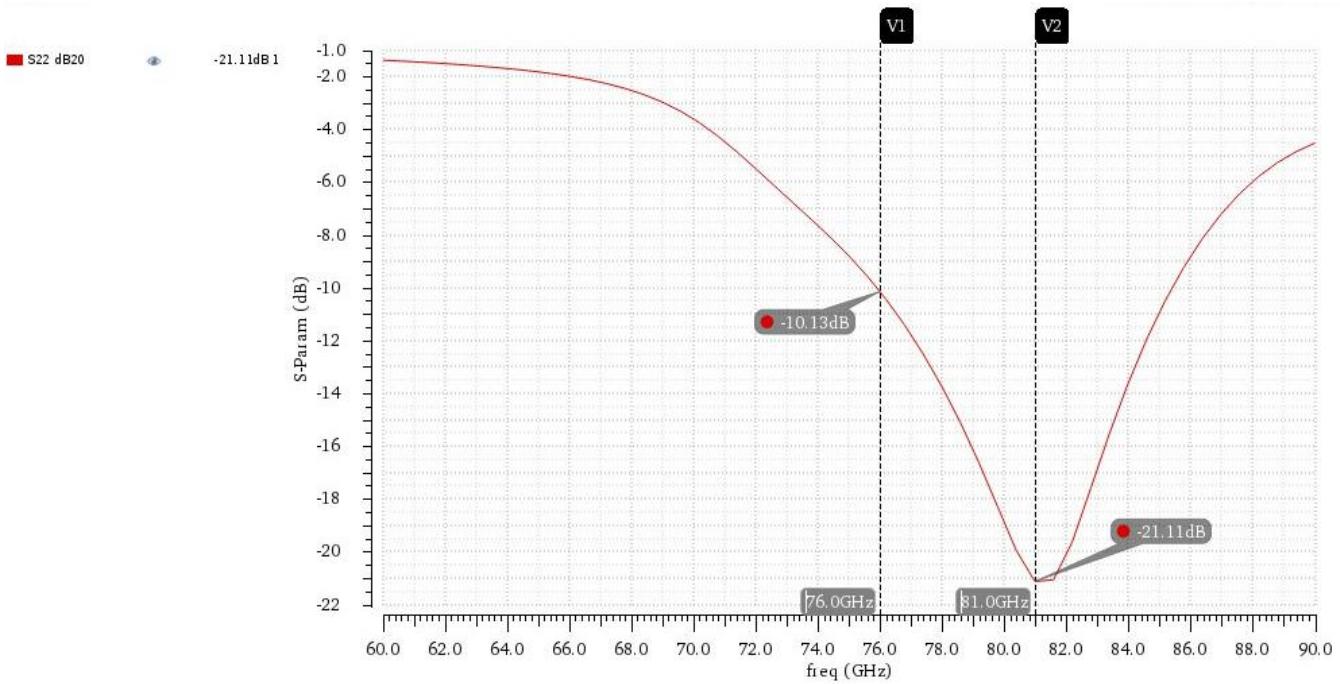


Fig. 7.15: S_{22} vs Frequency.

As shown in Fig. 7.15, the output return loss (S_{22}) is below $-10dB$ for the full range, which satisfies the target spec. The matching is shifted towards higher frequencies to obtain very low gain variation across the band. After EMing the full circuit, the matching shifted even more towards higher frequencies which caused the unwanted gain flatness out-of-band. More iterations to account for this shift after the EMing could improve the situation.

7.5.3.4. Linearity

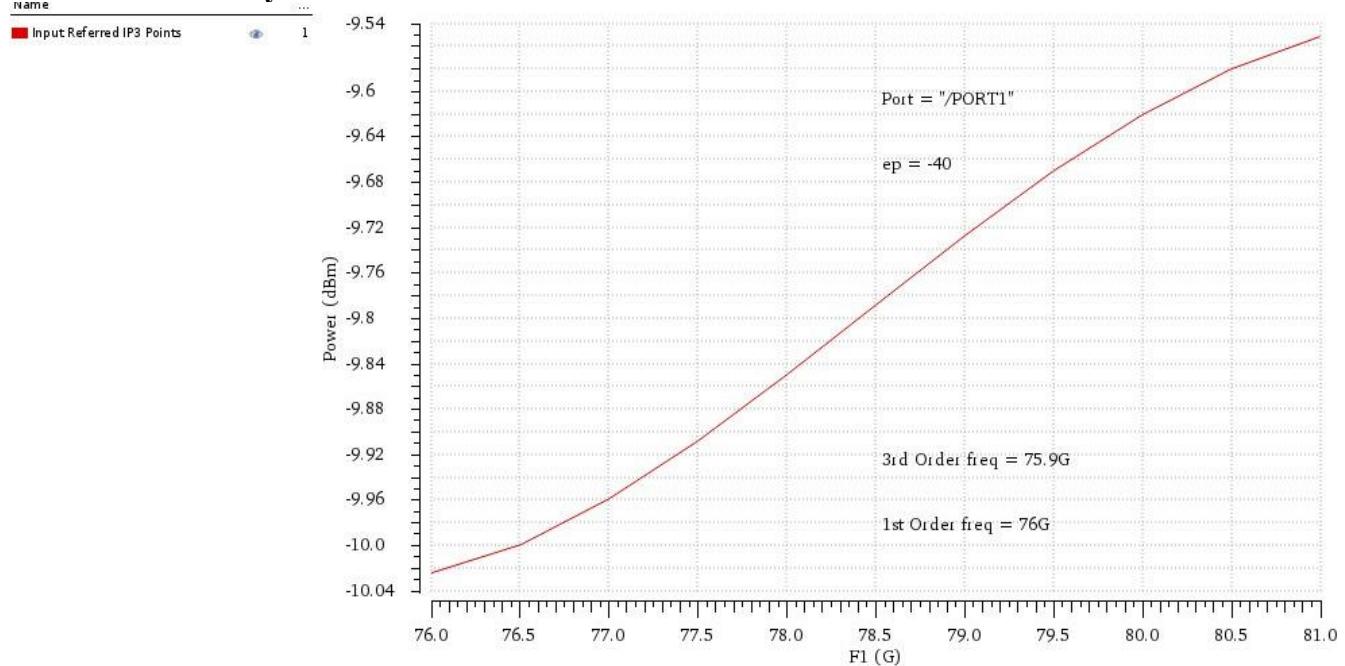


Fig. 7.16: IP3 vs Frequency

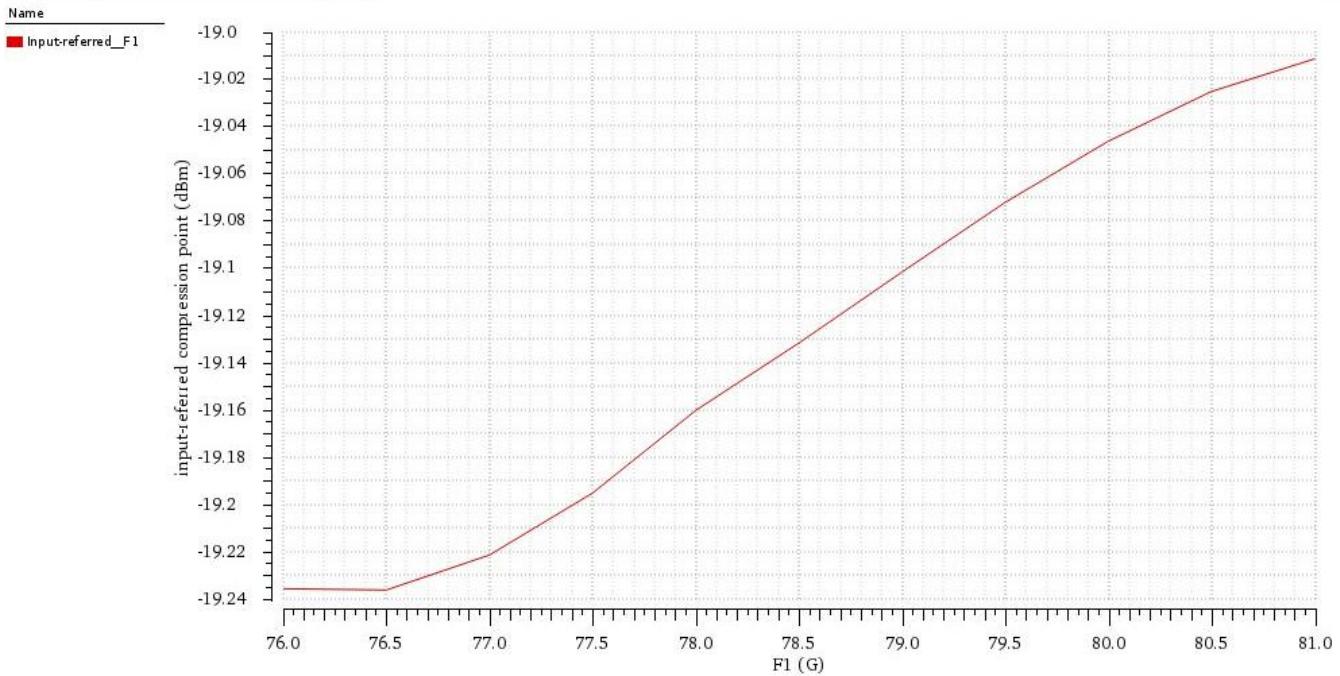


Fig. 7.17: IP1dB vs Frequency.

Fig. 7.16 and Fig. 7.17 show the simulated IIP_3 and IP_{1dB} , respectively. Both metrics satisfy the required linearity specifications ($IIP_3 > -14\text{dBm}$ and $IP_{1dB} > -24\text{dBm}$). The overall receiver linearity is dominated by the subsequent blocks in the analog baseband.

7.5.3.5. Stability

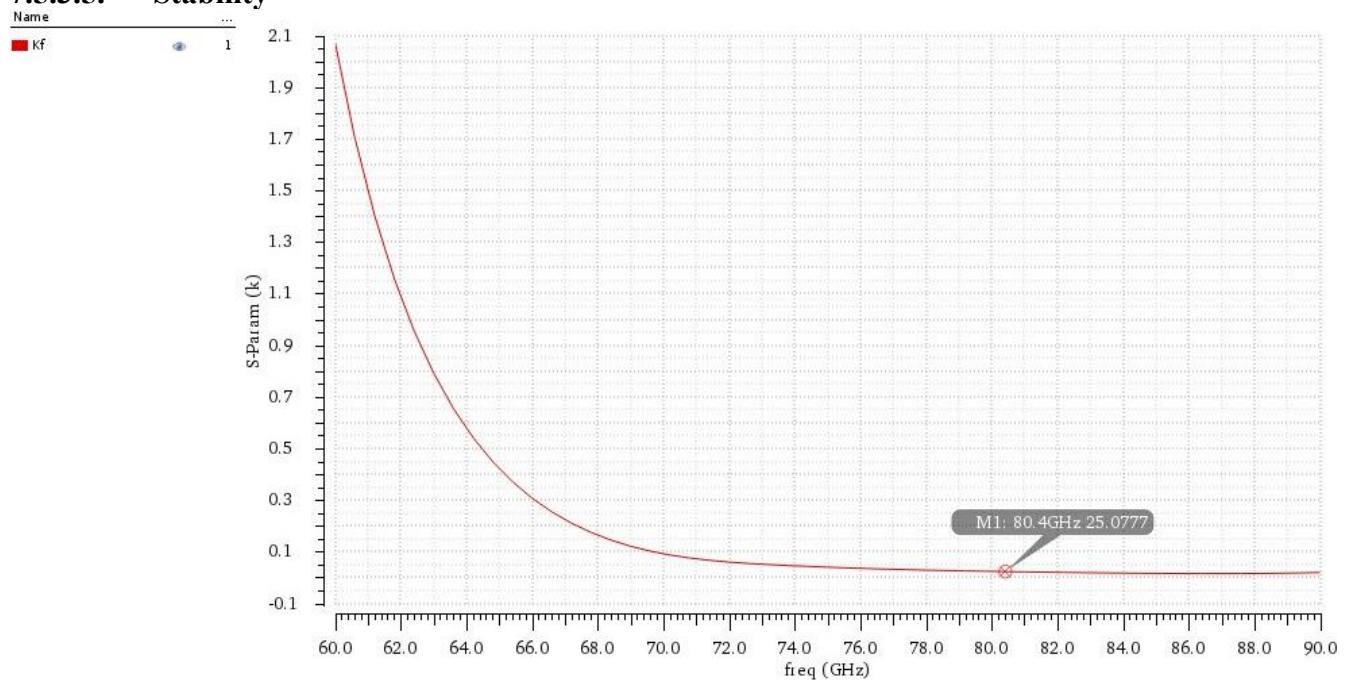


Fig. 7.18: Kf vs Frequency

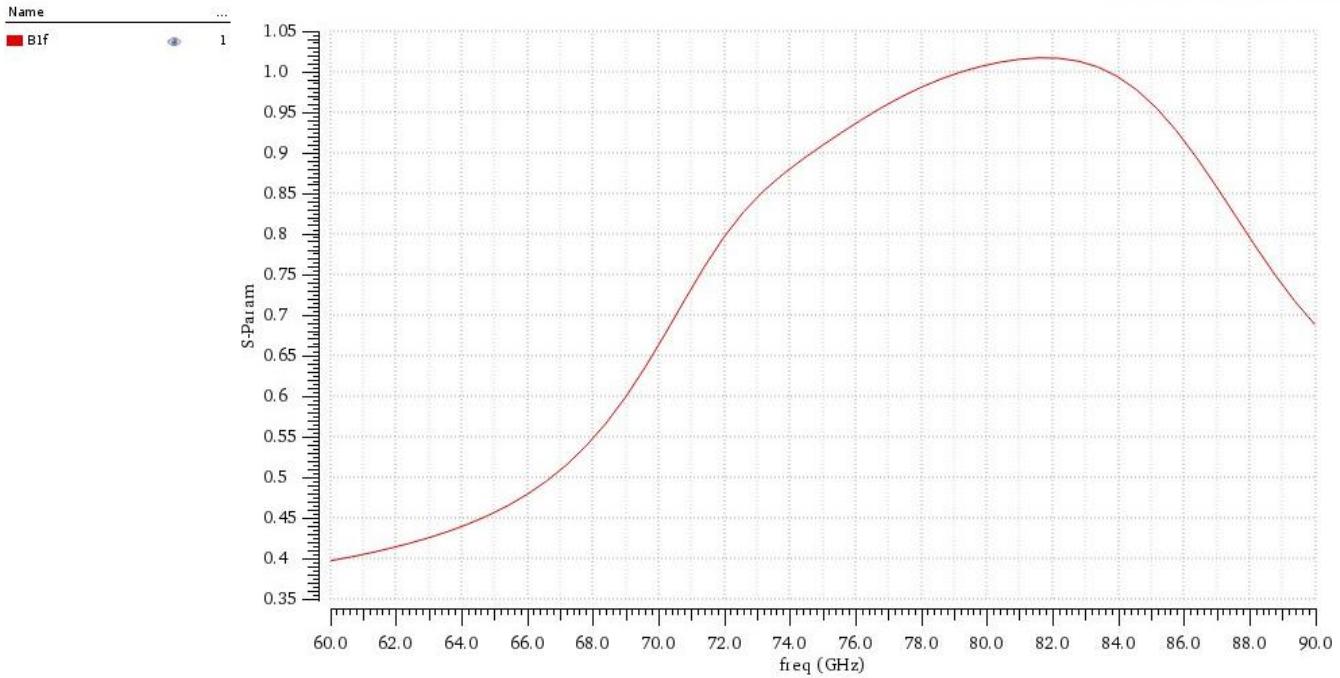


Fig. 7.19: B1f vs Frequency

As shown in Fig. 7.18 and Fig. 7.19, the K-factor is greater than one, and the B1-factor is greater than zero, which are necessary but not sufficient conditions for the unconditional stability of multi-stage amplifiers, as discussed in [Section 7.4]. To guarantee the stability of the multi-stage LNA, the impedance seen from the terminals of each device was verified to have no negative real part across all frequencies.

7.5.3.6. Reverse Isolation (S_{12})

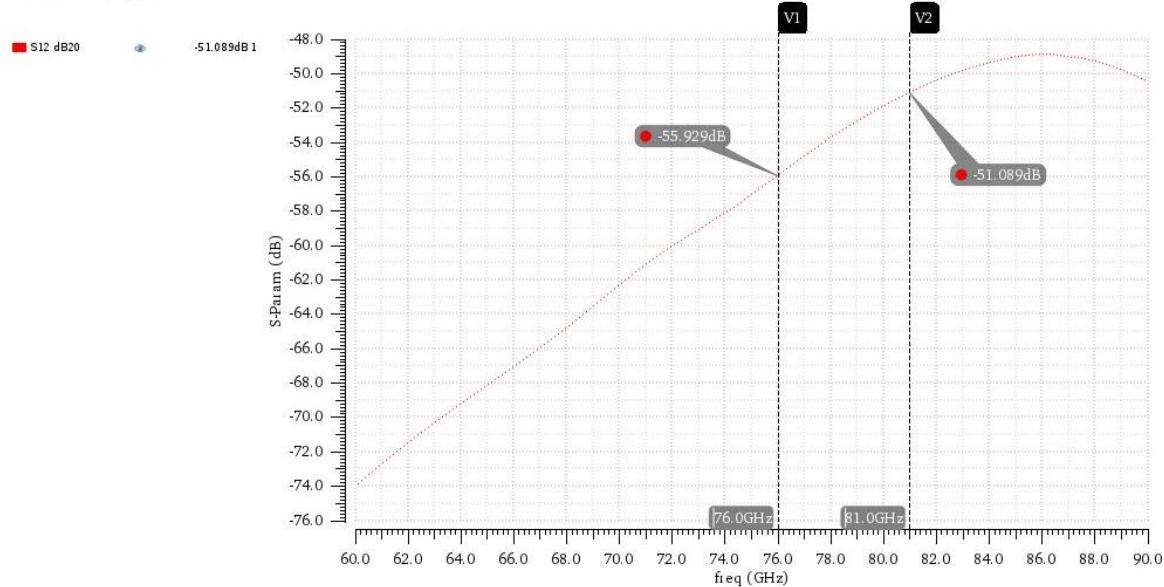


Fig. 7.20: S_{12} vs Frequency

Fig. 7.20 shows that the reverse isolation of the LNA is greater than 51dB across the full band, which is important to prevent signals at the output from coupling back to the LNA input.

7.5.4. Supply variations and Croners Simulation

The LNA response was simulated across three corners. The first is the typical case, where all transistors, capacitors, and resistors use the typical models (“tt” section from the model file) with a temperature of 27°C and the supply voltage at its nominal value (1.2V). The second case is the fast corner, where all the components operate at their fast corner (“ff” section) with a cold temperature of -45°C, and the supply voltage is higher by 10% above its nominal value (1.32V). The last corner is the slow corner, where all the components operate at their slow corner (“ss” section) with a hot temperature of 85°C, and the supply voltage is lower by 10% below its nominal value (1.08V).

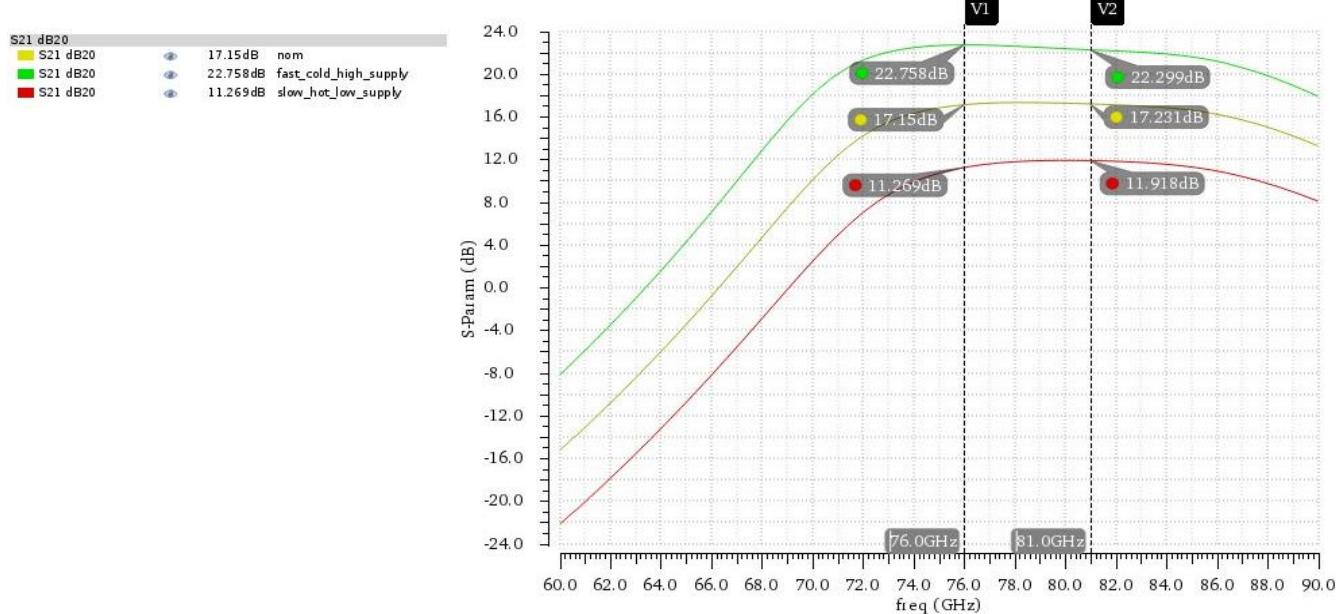


Fig. 7.21: LNA gain (S21) vs frequency across corners

Fig. 7.21 shows the LNA gain across the three corners. The achieved gain degrades at the slow corner and reaches 11dB but is still acceptable.

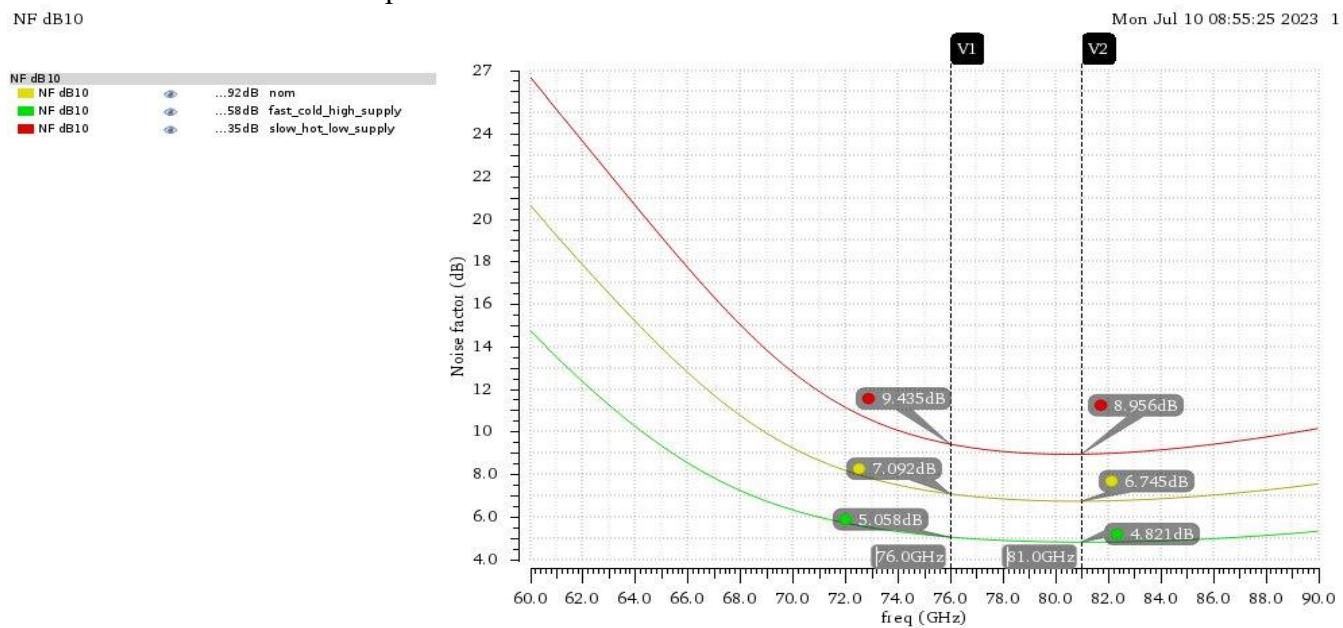


Fig. 7.22: NF vs frequency across corners

Mon Jul 10 08:55:25 2023 1

Fig. 7.22 shows the LNA noise figure across the three corners. The achieved noise figure degrades at the slow corner and reaches $9.4dB$ but is still less than $10dB$.

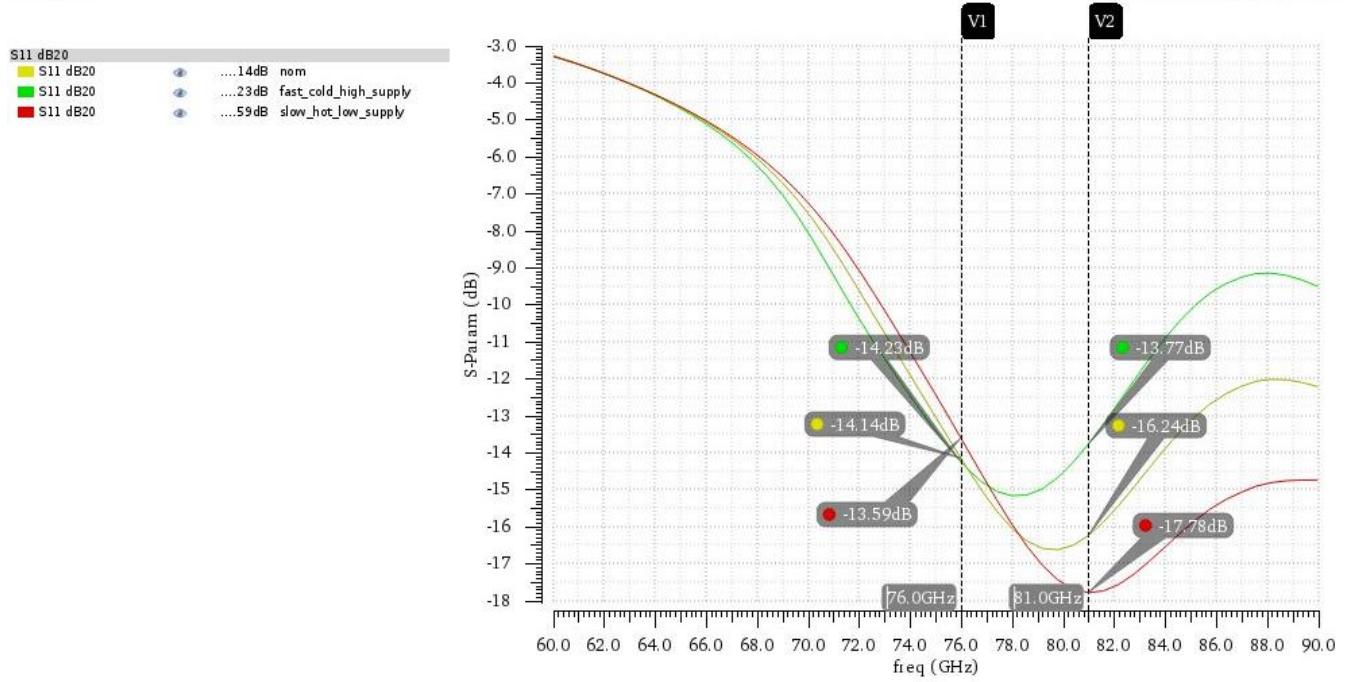


Fig. 7.23: S11 vs frequency across corners.

Fig. 7.23 shows the LNA input return loss (S_{11}) across the three corners. The achieved S_{11} stays greater than $-10dB$ across the band $76 – 81GHz$ across different corners.

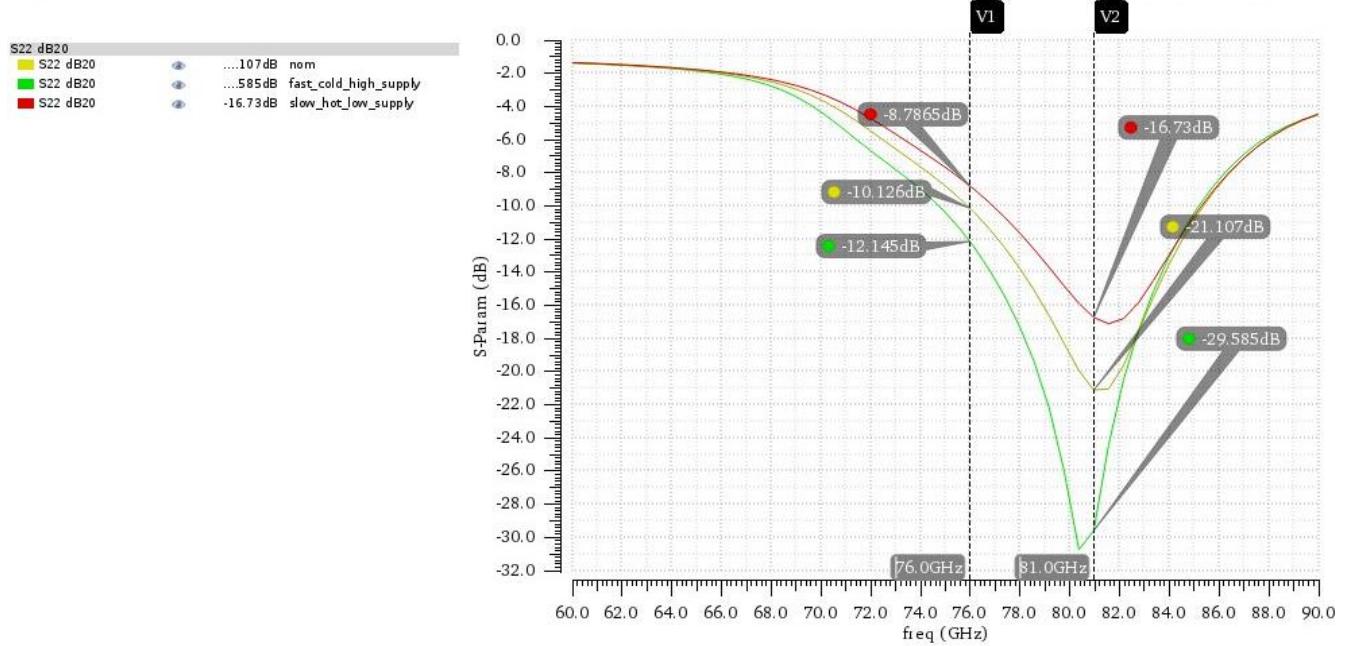


Fig. 7.24: S22 vs frequency across corners.

Fig. 7.24 shows the LNA output return loss (S_{22}) across the three corners. The achieved S_{22} stays greater than $-10dB$ across the band $76 – 81GHz$ across different corners except for the slow corner as it reaches $-8.7dB$ at $76GHz$, but it is still acceptable.

Name	...	Corner
Kf	nom	
Kf	fast_cold_high_supply	
Kf	slow_hot_low_supply	

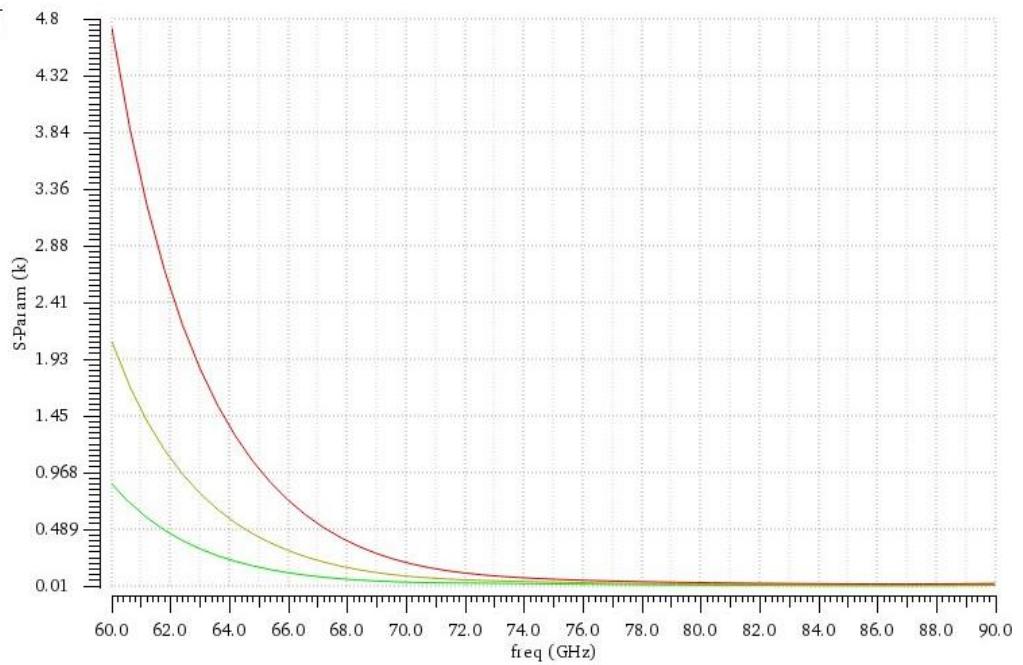


Fig. 7.25: Kf vs frequency across corners.

Fig. 7.25 shows the LNA K -factor across different corners. We can see that it is still above one for all frequencies across different corners, which is necessary but not sufficient, as discussed before.

Name	...	Corner
B1f	nom	
B1f	fast_cold_high_supply	
B1f	slow_hot_low_supply	

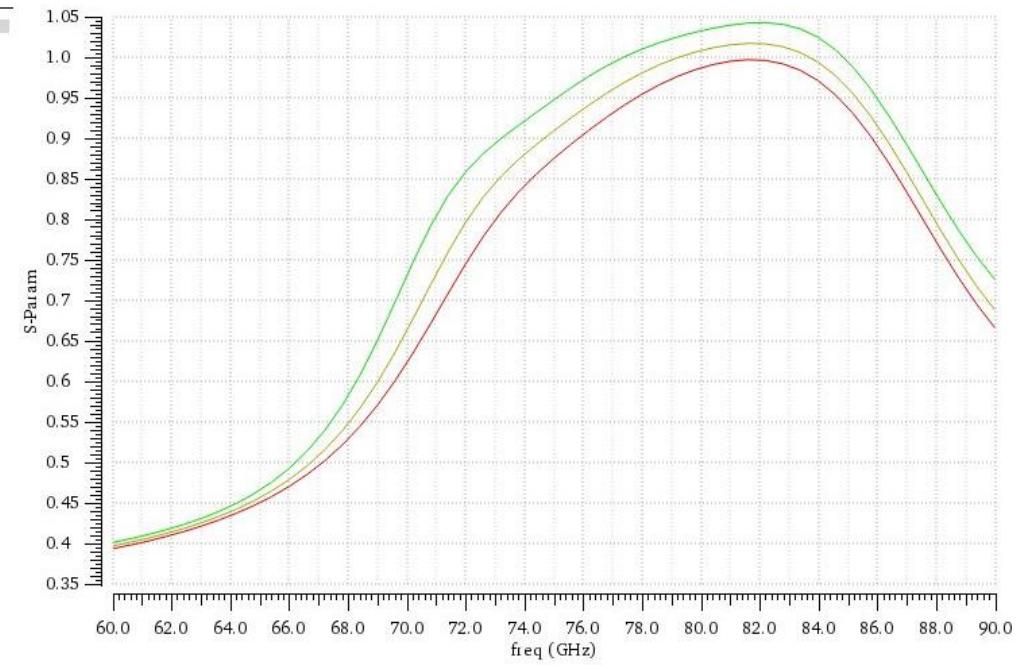


Fig. 7.26: B1f vs frequency across corners.

Fig. 7.26 shows the LNA $B1$ -factor across different corners. We can see that it is still above zero for all frequencies across different corners, which is necessary but not sufficient, as discussed before.

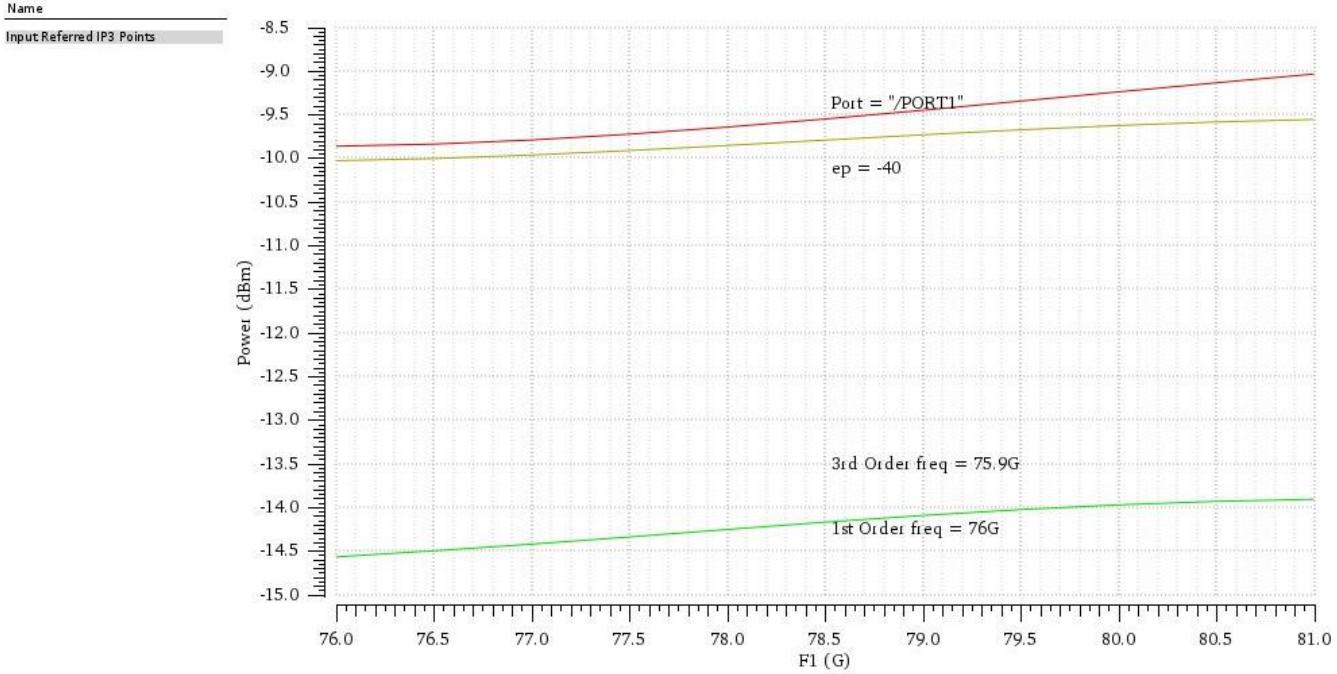


Fig. 7.27: IIP₃ vs frequency across corners.

Fig. 7.27 shows the LNA IIP_3 versus frequency across different corners. We notice that the worst linearity occurs at the fast corner (green curve) as the IIP_3 reaches a minimum of -14.5dBm . this is acceptable compared to the target spec of $IIP_3 > -14\text{dBm}$.

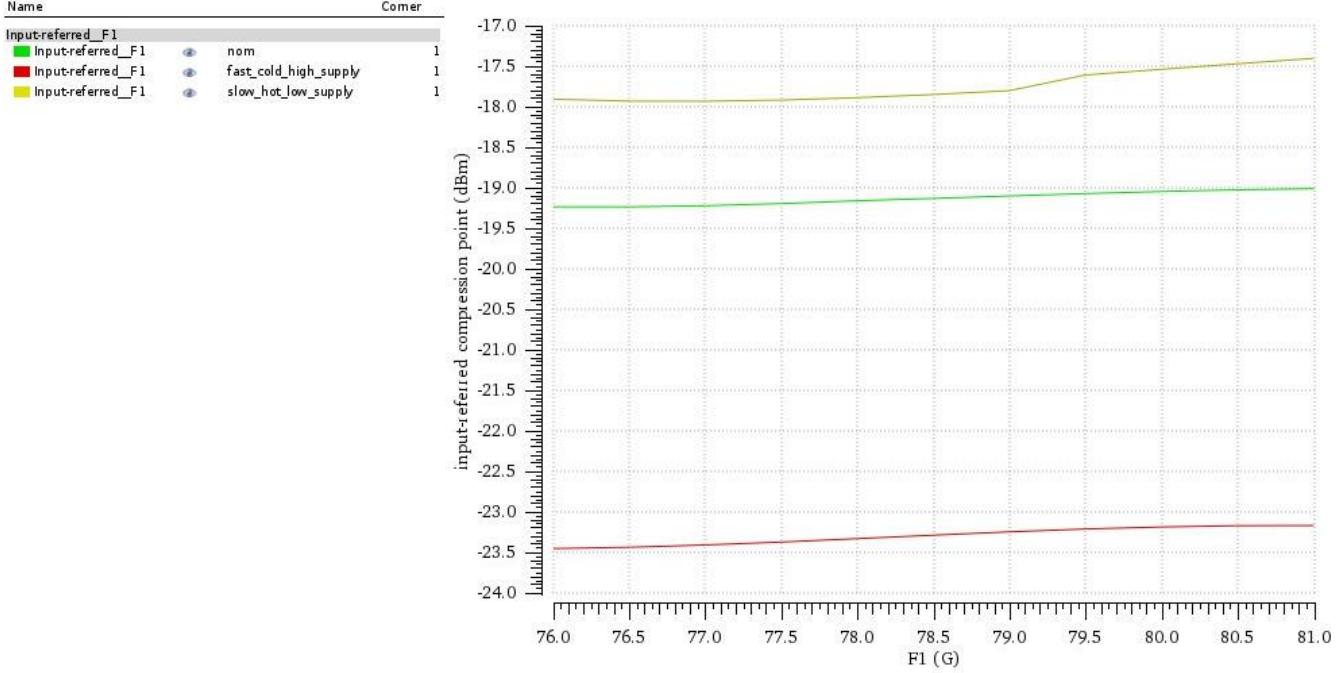


Fig. 7.28: IP_{1dB} vs frequency across corners.

Fig. 7.28 shows the LNA IP_{1dB} versus frequency across different corners. We notice that the worst linearity occurs at the fast corner (red curve) as the IP_{dB} reaches a minimum of -23.5dBm . this is acceptable compared to the target spec of $IP_{1dB} > -24\text{dBm}$.

7.5.5. LNA Layout

Fig. 7.29 shows the full layout of the proposed LNA. Layout iterations were made to try to fit the LNA in the least area possible without causing the transmission lines to come too close to each other to avoid any additional coupling between them that would affect the performance. The total area of the LNA is $922 * 481 \mu\text{m}^2$. Design rules check (DRC) was performed to make sure that the layout did not violate the technology rules. During the layout of the transistors, symmetry was taken into consideration to avoid any relative signal delays at such a high frequency; the signal should reach a transistor's multipliers at the exact same time with no phase shift or delay. The layout of the cascode is shown in Fig. 7.31. A large resistor is connected between the source and bulk of each device to float the bulk at high frequencies reducing the parasitic capacitances seen at the drain and the source of the device. An electromagnetic (EM) simulation for the full layout of the LNA was performed on ADS to see how the performance is affected by the coupling between the transmission lines. All the presented results are after the EMing of the LNA and the parasitic extraction (PEX) of the transistors. Fig. 7.30 shows the EMing of the LNA.

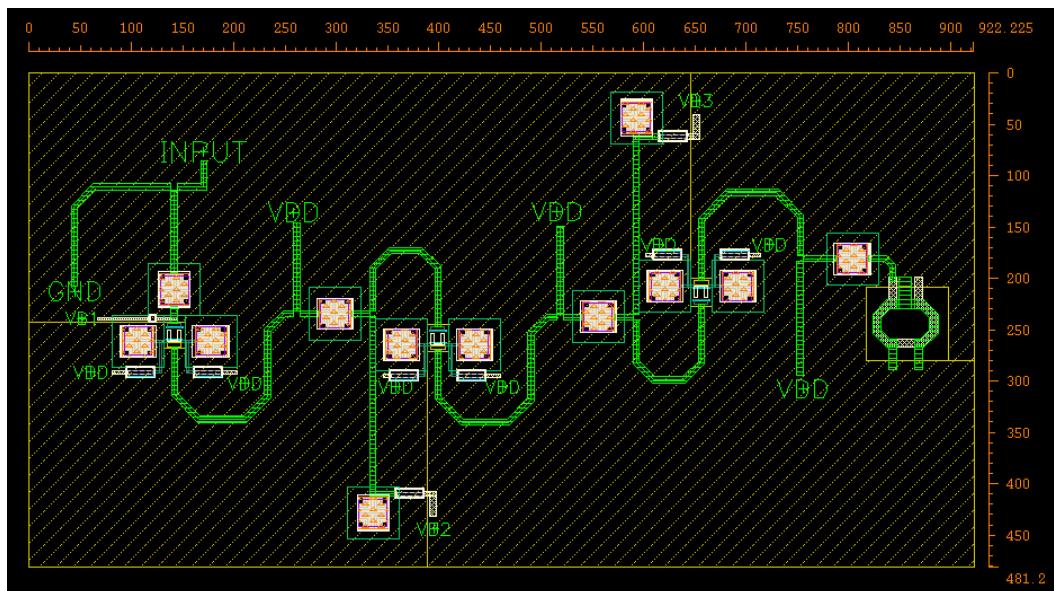


Fig. 7.29: LNA full layout

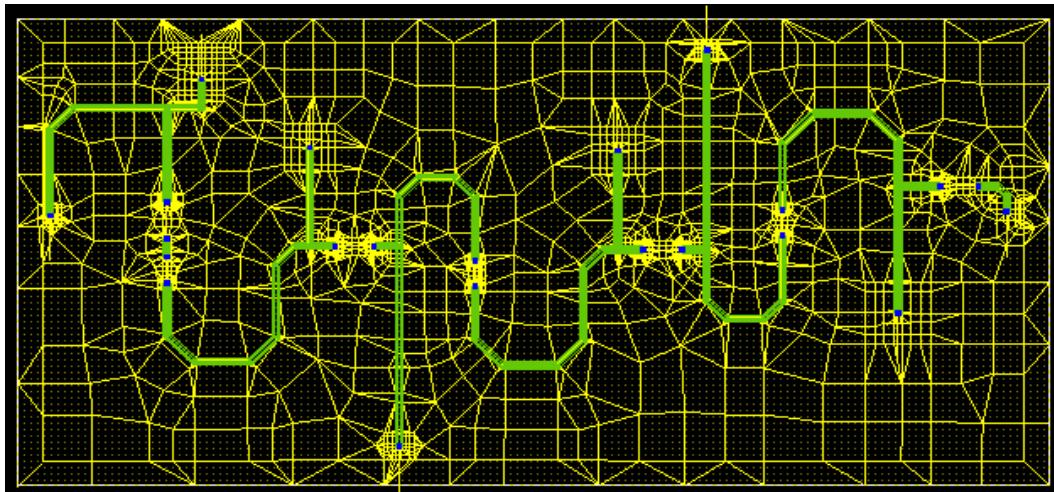


Fig. 7.30: LNA EMing on ADS

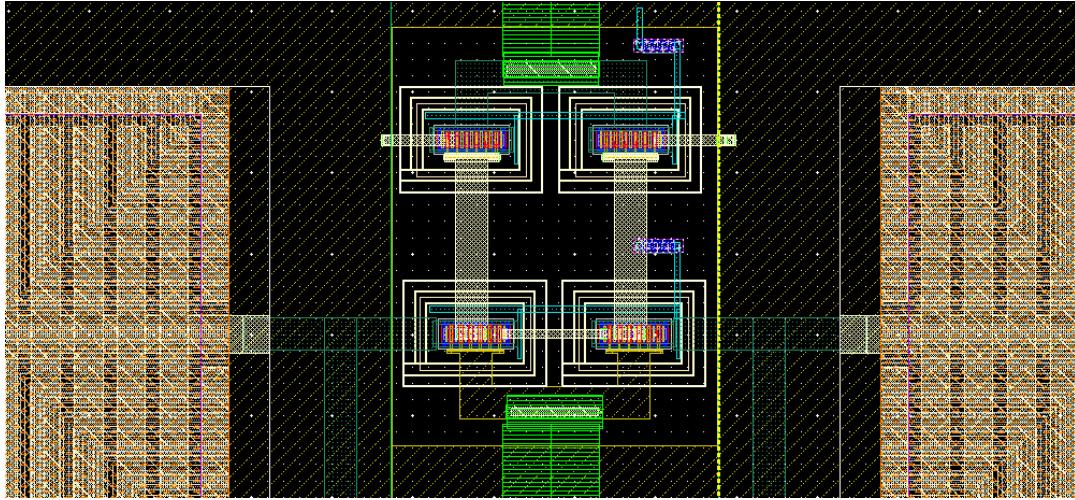


Fig. 7.31: Cascode layout.

Performance Summary and Literature Survey

Table 7.1 summarizes the performance of the proposed LNA:

Table 7.2: LNA performance summary.

Spec	Target	Achieved
Gain (S21)	$> 14dB$	$17.2dB$
Noise Figure	$< 10dB$	$7dB$
IIP3	$> -14dBm$	$-10dBm$
IP1dB	$> -24dBm$	$-19dBm$
S11	$< -10dB$	$-14 \rightarrow -16dB$
S22	$< -10dB$	$-10 \rightarrow -21dB$
S12	-	$> 51dB$
Supply voltage	-	1.2V
Power consumption*	-	$3 * 5.3mA * 1.2V = 19.1mW$
Area	-	$922 * 481\mu m^2$

*This is the power consumption at the nominal corner without the power consumption of the biasing current mirrors (with current mirrors, it becomes 20mW).

Table 7.3 shows the LNA literature survey:

Table 7.3: LNA literature survey.

	[5]	[6]	[7]	[8]	This work
Process	65-nm CMOS	65-nm CMOS	65-nm CMOS	90-nm CMOS	65-nm CMOS
Frequency	75.6-76.3GHz	76-77GHz	76-77GHz	73.5-77.1GHz	76-81GHz
Gain	17.5dB	17.6dB	11dB	14dB	17.3dB
NF	7.4dB	6dB	7.8dB	6.8dB	7dB
IP1dB	-22dBm	NA	-29dBm	-20dBm	-19dBm
IIP3	-12.5 dBm	NA	-19dBm	-10dBm	-10dBm
Power Cons.	30mw	20mw	21.5mw	45mw	19.1mw

7.5.6. Balun Design

A balun between the LNA and the mixer is needed to do single-ended to differential conversion. There are many topologies for the integrated baluns. Baluns can be implemented using active circuits, which is preferable at low frequencies. However, at higher frequencies, baluns based on microwave structures, such as transform-based baluns, Marchand, and rat-race baluns, are preferred. We choose to use a transformer-based balun to make use of it in the matching between the mixer RF input and the 50Ω at the LNA output [?]. The used balun is shown in Fig. 7.32. The three main metrics to characterize the performance of the balun are the amplitude imbalance between the differential outputs (AI), phase imbalance (PI), and insertion loss (IL).

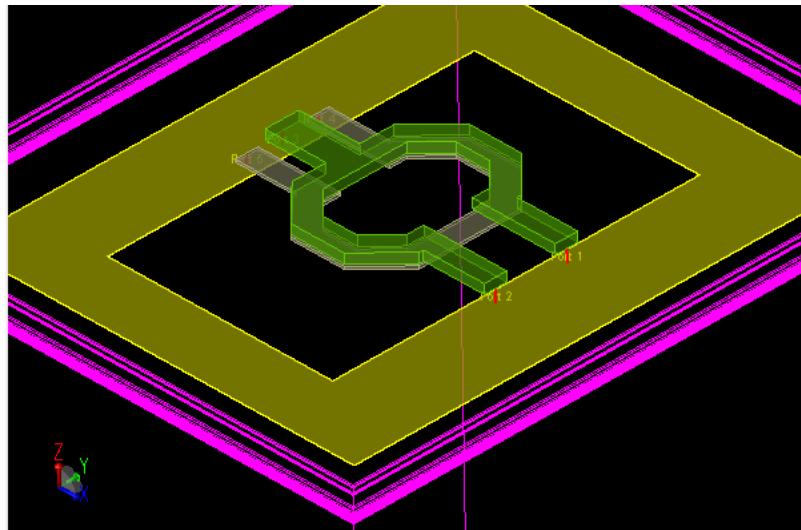


Fig. 7.32: the used transformer-based balun as shown in ADS.

The AI and PI of the balun are primarily affected by the capacitive coupling between the two coils. More capacitive coupling causes worse AI and PI. Thus to improve the AI and PI, we need to decrease the width of the lines to decrease the capacitive coupling. Decreasing the lines' width increases the insertion loss of the balun. A sweet spot can be obtained while keeping certain self-inductances for the primary and secondary coils to achieve the required matching. Fig. 7.33 shows the AI of the balun.

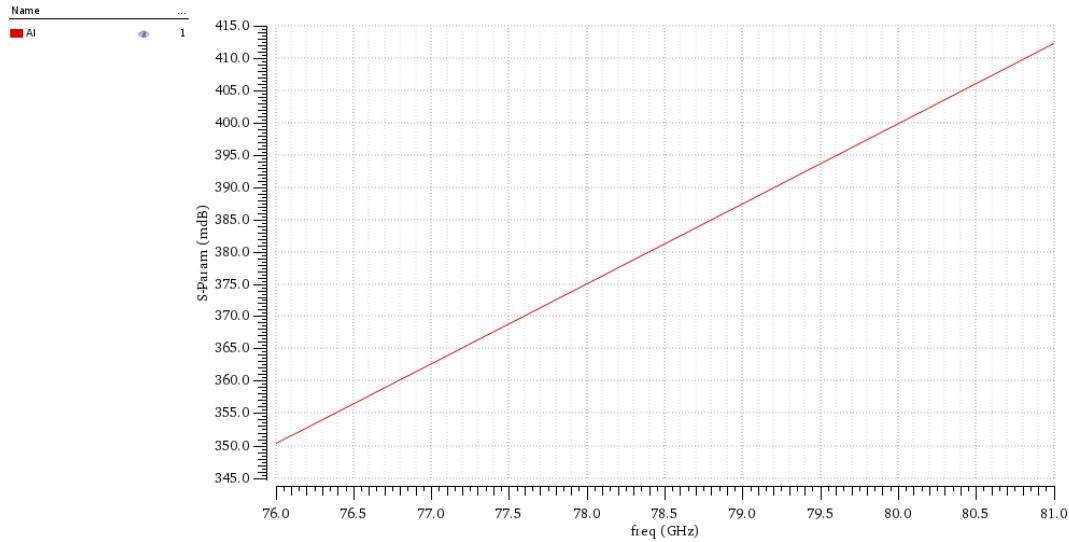


Fig. 7.33: Amplitude imbalance for the Balun.

Fig. 7.34 and Fig. 7.35 show the phase imbalance and the insertion loss of the balun, respectively.

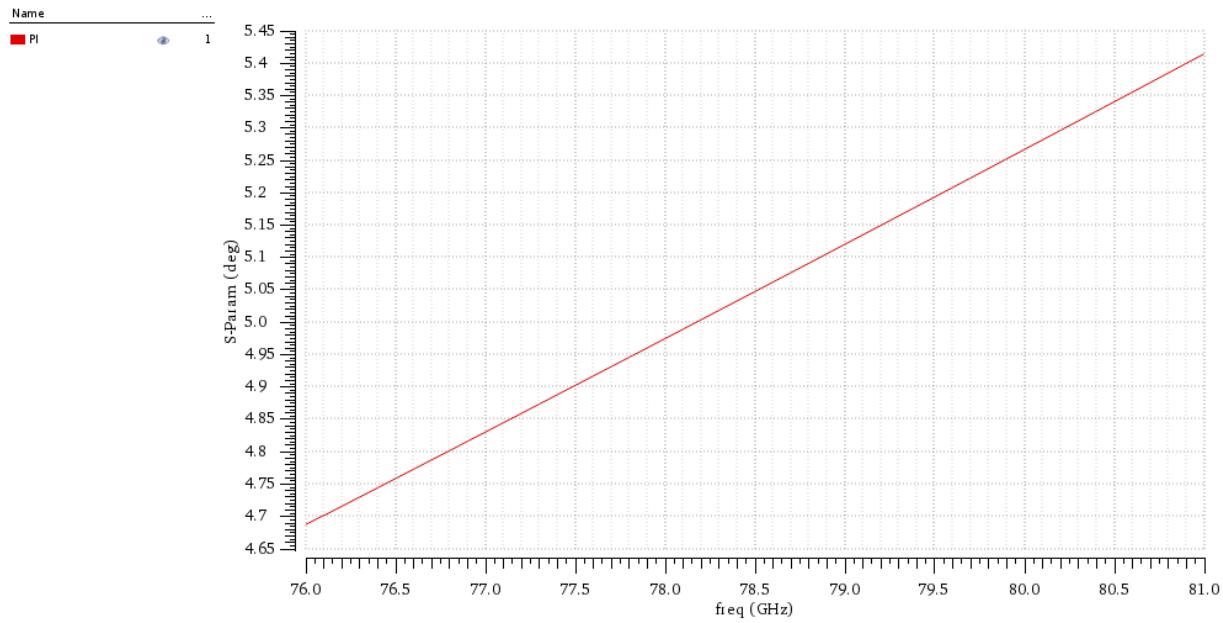


Fig. 7.34: Phase imbalance for the Balun.

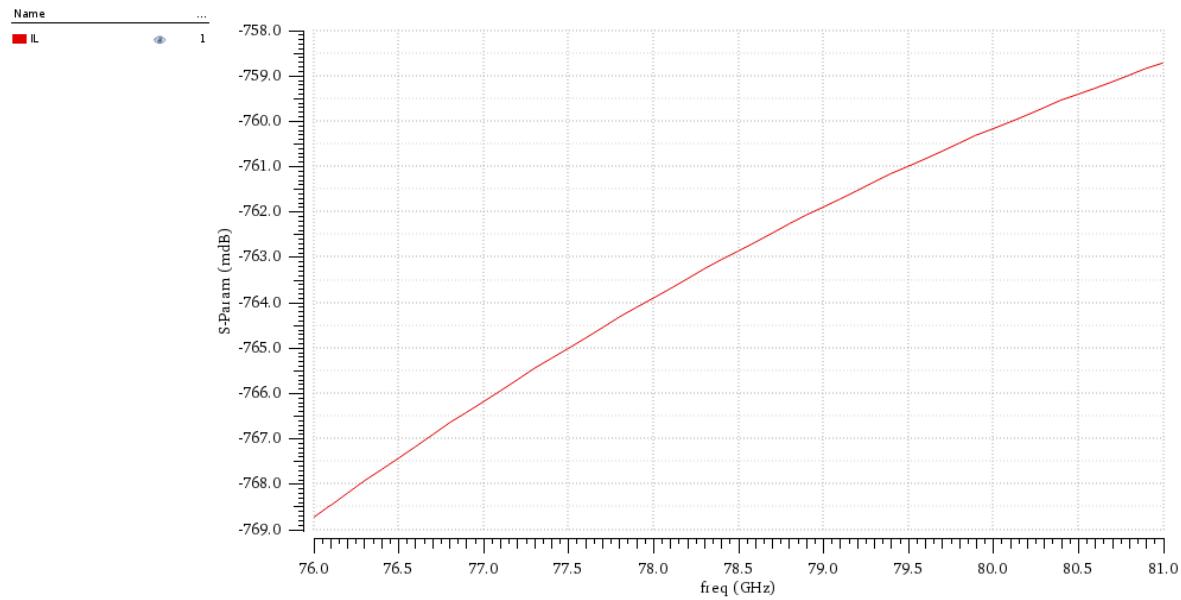


Fig. 7.35: Insertion loss of the Balun.

7.6. Common Source

In this section, the common-source topology is illustrated. As shown in Fig. 7.36, the design consists of a single-ended, four-stage common source low noise amplifier. The common source topology achieves a lower gain than the cascode topology under the same supply (1.2V). Also, the parasitic capacitors of the single transistor introduce some feedback that degrades the isolation compared to the cascode topology. The main advantage of this topology is that it has lower noise as it has a lower number of transistors.

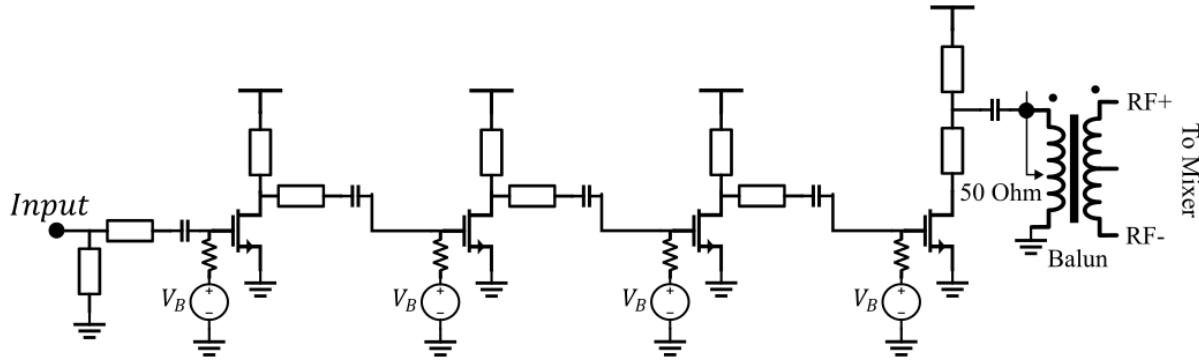


Fig. 7.36: Common Source LNA schematic

The four stages are identical and biased at a current density of $350 \mu A/\mu m$, which is higher than the current density for optimum noise. This is a higher current density to achieve the required gain as the noise spec was not degraded that much. The optimal width per finger to maximize the f_{max} is $1\mu m$. The total width of the transistors is $16\mu m$ divided into 2 multipliers with 8 fingers each. Micro-strip transmission lines and metal-insulator-metal (MIM) capacitors are used for inter-stage, input, and output matching networks. The input and output are matched to 50Ω to ease the integration with the mixer. The micro-strip line is made of a signal line in the M9 layer and a ground plane in M2. We used M2 for the ground instead of M1 because it is thicker and provides better isolation from the substrate. A transformer-based balun is used at the output of the LNA to perform the single-ended to differential conversion for the passive mixer.

7.6.1. Post-Layout Simulation Results

The upcoming shown results are reported after Electro-Magnetic Simulations (EMing) the complete layout of the circuit and Parasitic Extraction (PEX) for the transistors. The LNA was simulated at the nominal corner, then across different corners. First, we present the typical results.

7.6.1.1. Gain

As shown in Fig. 7.37, the reported gain of the LNA is about $15dB$ across the range $76 - 81GHz$ with a gain flatness of about $0.5dBm$, this satisfies the gain spec. The problem with this design is that the gain does not roll off quickly out of the band, which makes the system sensitive to out-of-band blockers. A band-pass filter may be used at the beginning of the receiver to reject out-of-band emissions.

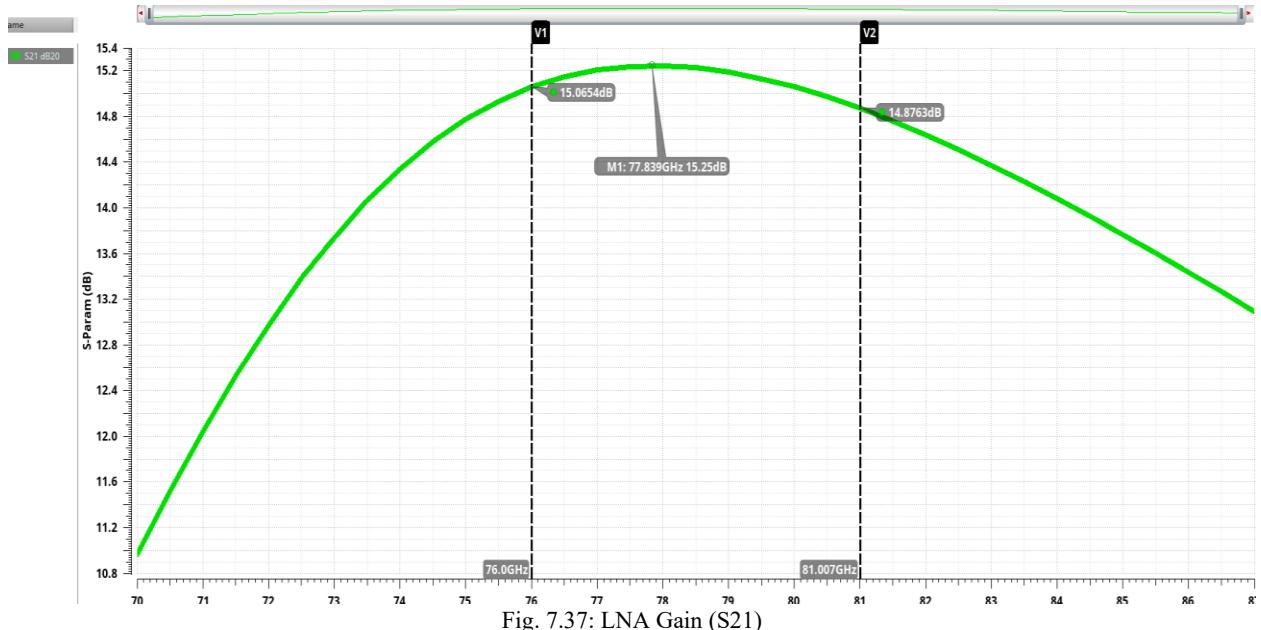


Fig. 7.37: LNA Gain (S₂₁)

7.6.1.2. Noise Figure

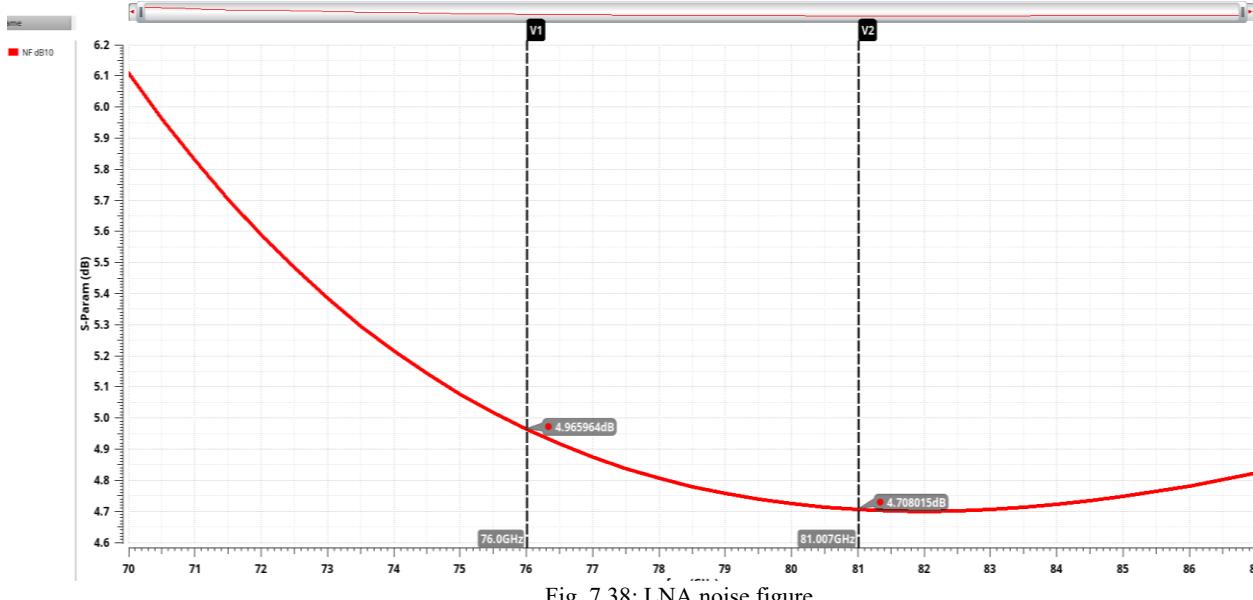
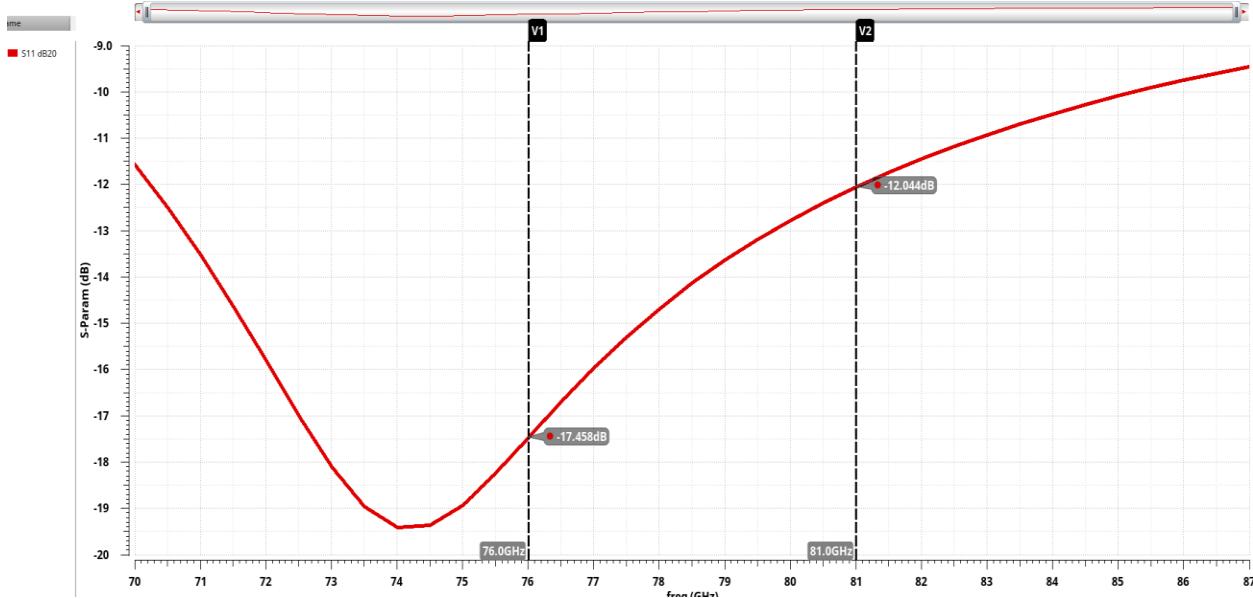


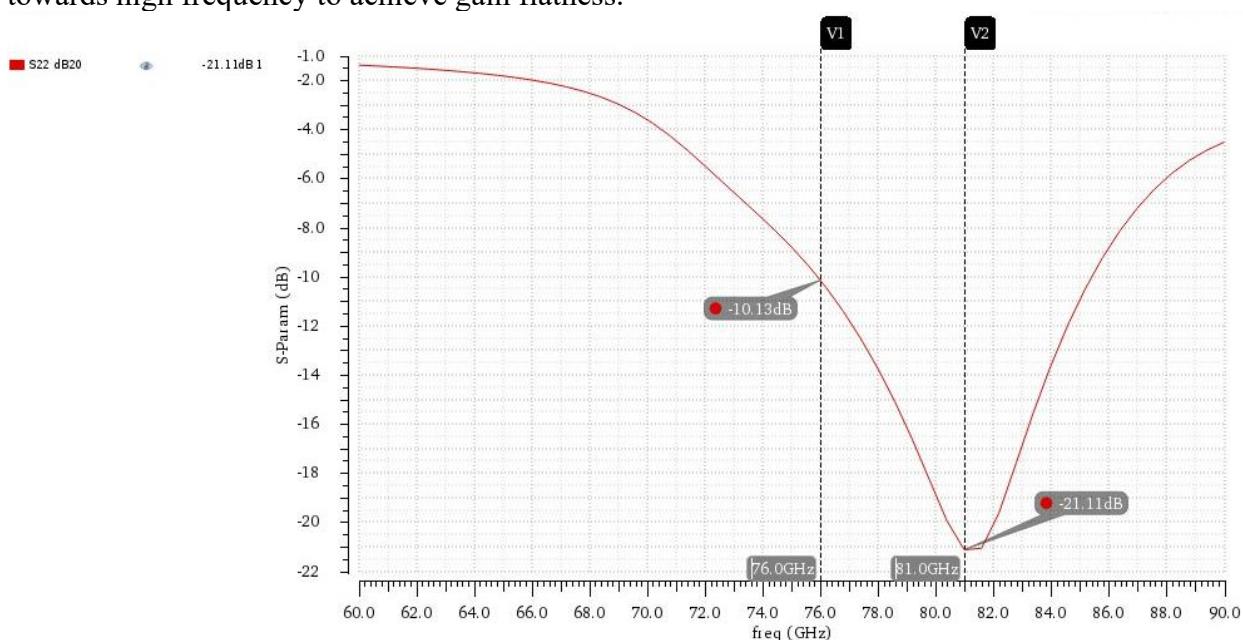
Fig. 7.38: LNA noise figure

Fig. 7.38, shows the noise figure versus frequency. The lowest noise figure achieved within band 76 – 81GHz is about 4.8 dB, which satisfies the target spec ($NF < 8$).

7.6.1.3. Input and Output Matching (S11&S22)



As shown in Fig. 7.39, the input return loss (S_{11}) is below $-10dB$ for the full range with more skew towards high frequency to achieve gain flatness.



As shown in Fig. 7.40, the output return loss (S_{22}) is below $-10dB$ for the full range, which satisfies the target spec. The matching is shifted towards higher frequencies to obtain very low gain variation across the band.

After EMing the full circuit, the matching shifted even more towards higher frequencies which caused the unwanted gain flatness out-of-band.

7.6.1.4. Linearity

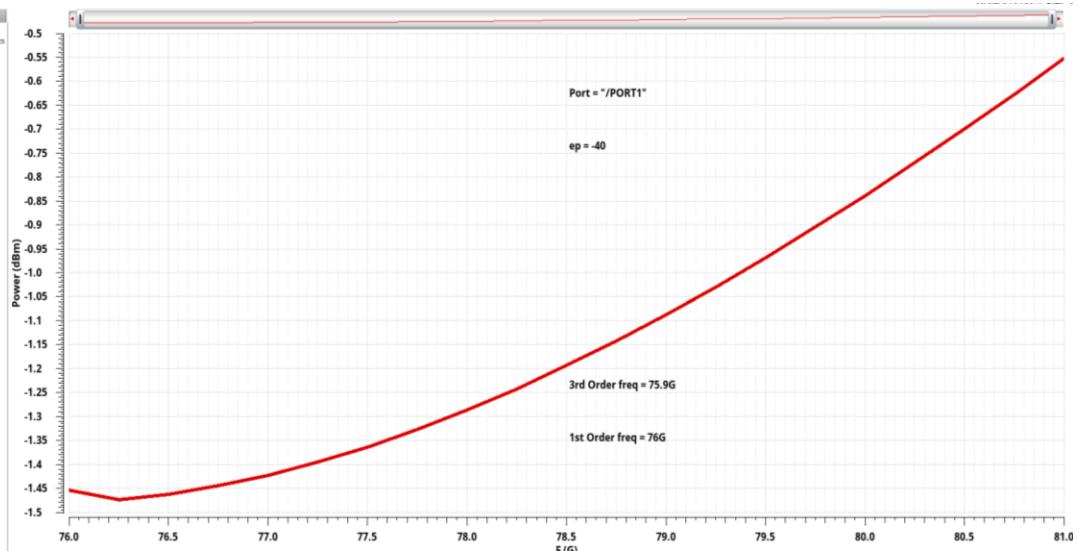


Fig. 7.41: IIP₃ vs Frequency.

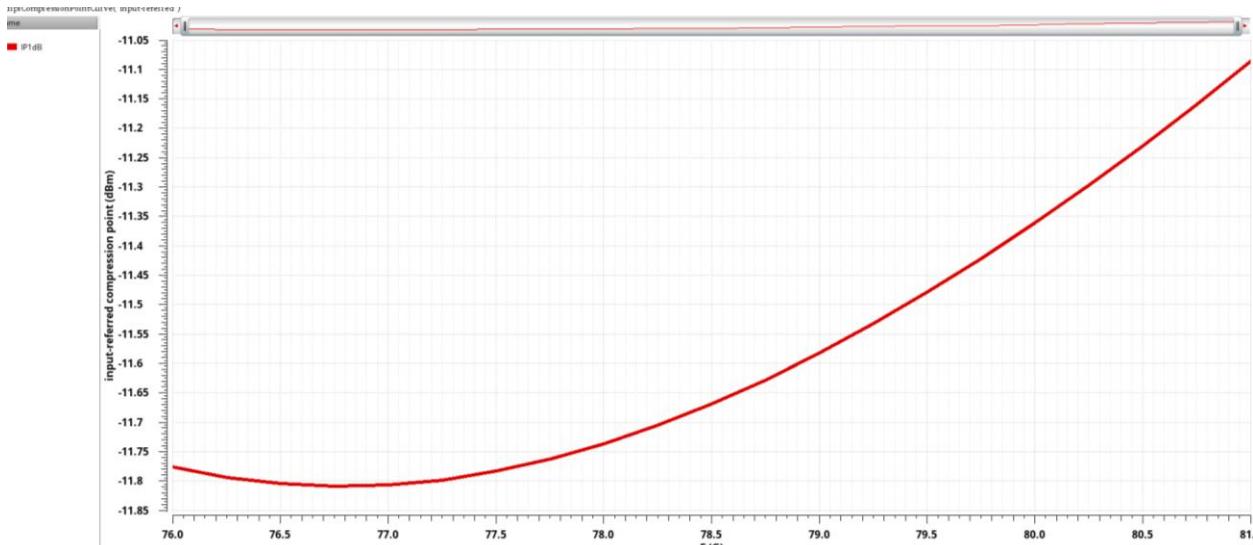


Fig. 7.42: IP_{1dB} vs Frequency.

Figs. 7.41 and 7.42 show the resulting IIP_3 and IP_{1dB} , respectively. Both specs are achieved ($IIP_3 > -14dBm$ and $IP_{1dB} > -24dBm$). The overall receiver linearity is dominated by the subsequent blocks in the analog baseband.

7.6.1.5. Stability

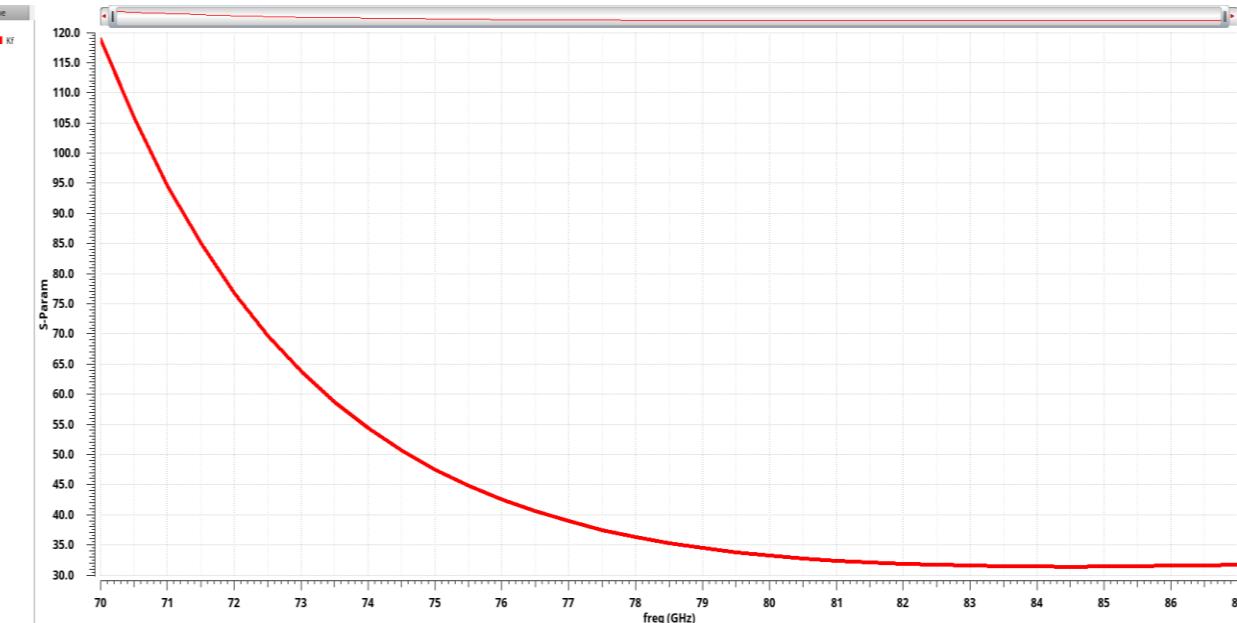


Fig. 7.43: Kf vs Frequency.

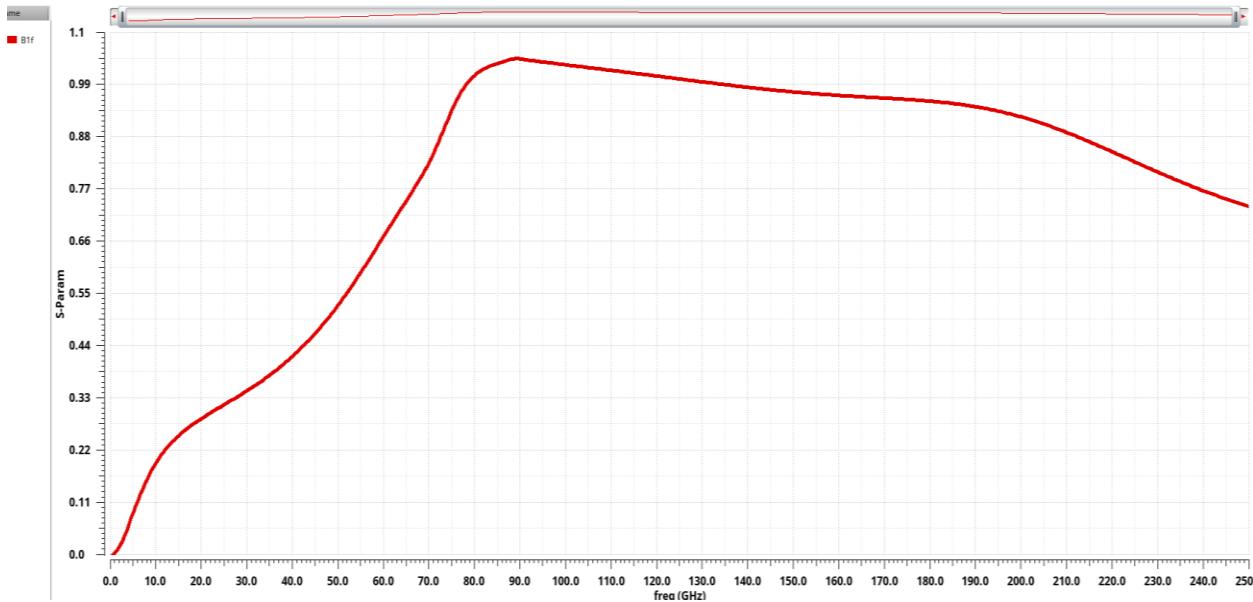


Fig. 7.44: B1f vs Frequency.

As shown in Figs. 7.43 and 7.44, the K-factor is greater than one, and the B1-factor is greater than zero. These are necessary but not sufficient conditions for the unconditional stability of multi-stage amplifiers, as discussed in [Section 7.5]. To guarantee the stability of the multi-stage LNA, the impedance seen from the terminals of each device was verified to have no negative real part components across all frequencies.

7.6.1.6. Reverse Isolation (S12)

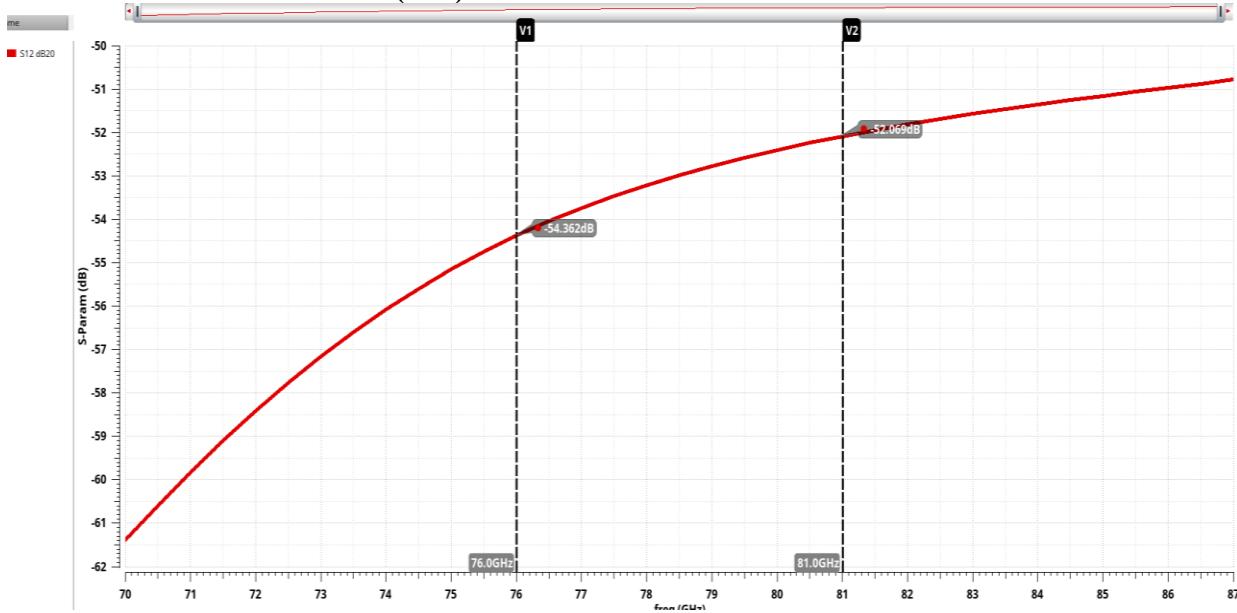


Fig. 7.45: Reverse isolation (S12)

Fig. 7.45 shows that the reverse isolation of the is greater than 52dB across the band of interest. This is important to prevent signals at the output from coupling back to the LNA input.

7.6.2. Supply Variations and Corners Simulation

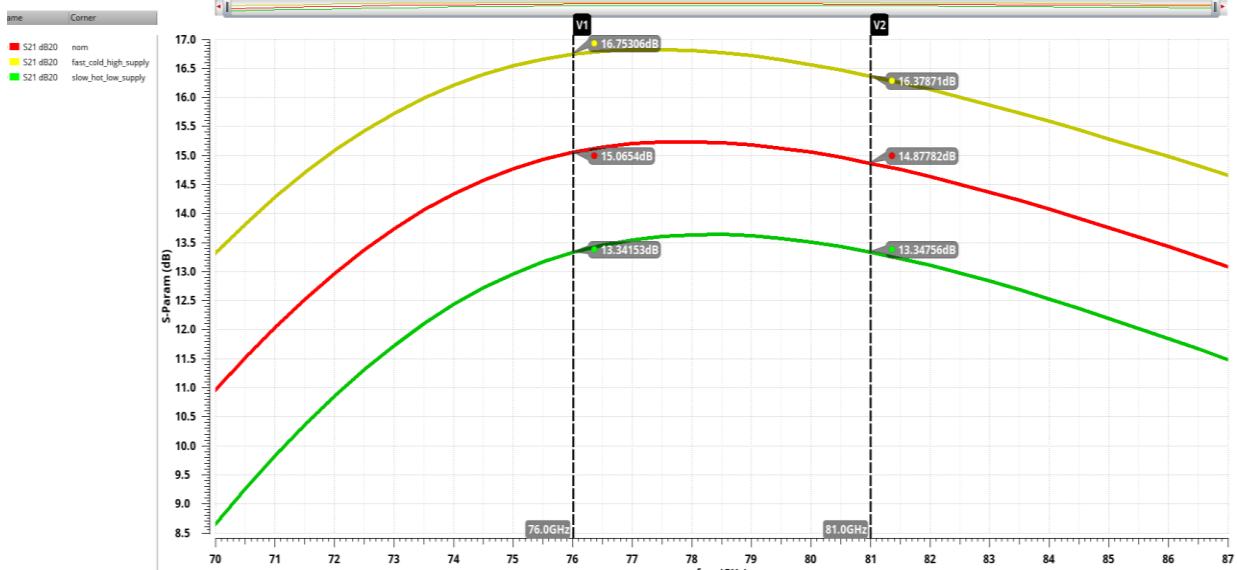


Fig. 7.46: LNA Gain across corners.

Fig. 7.46 shows the LNA gain across the three corners. The achieved gain degrades at the slow corner and reaches 13.5dB but is still acceptable.

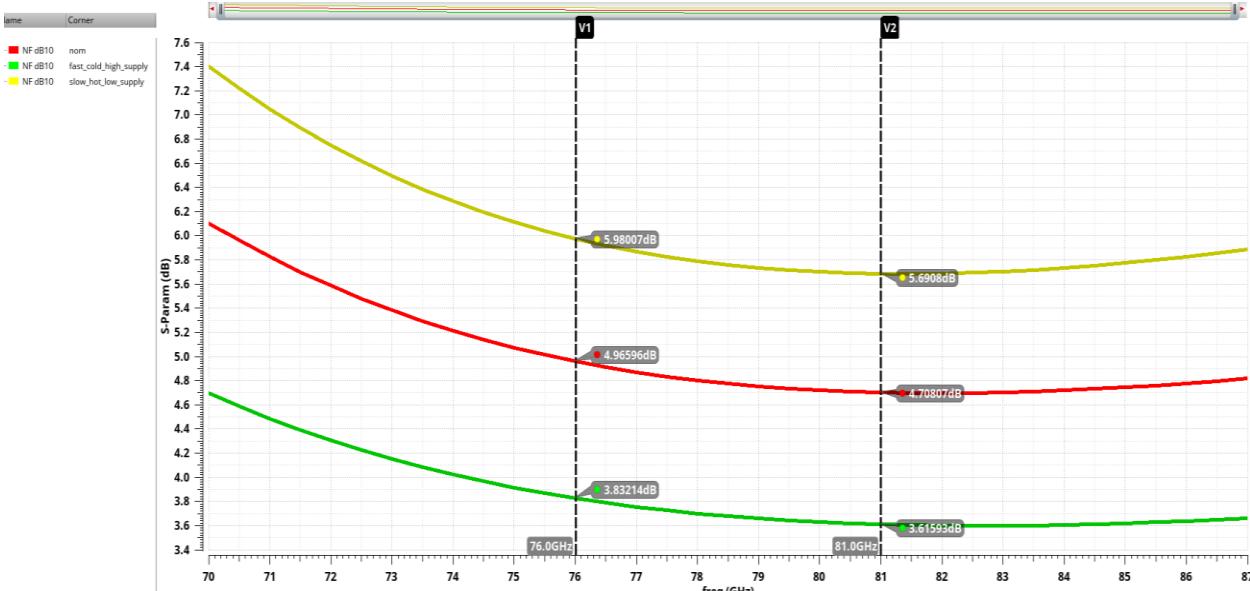


Fig. 7.47: LNA NF across corners.

Fig. 7.47 shows the LNA noise figure across the three corners. The achieved noise figure degrades at the slow corner and reaches 5.9dB but is still less than 8dB .

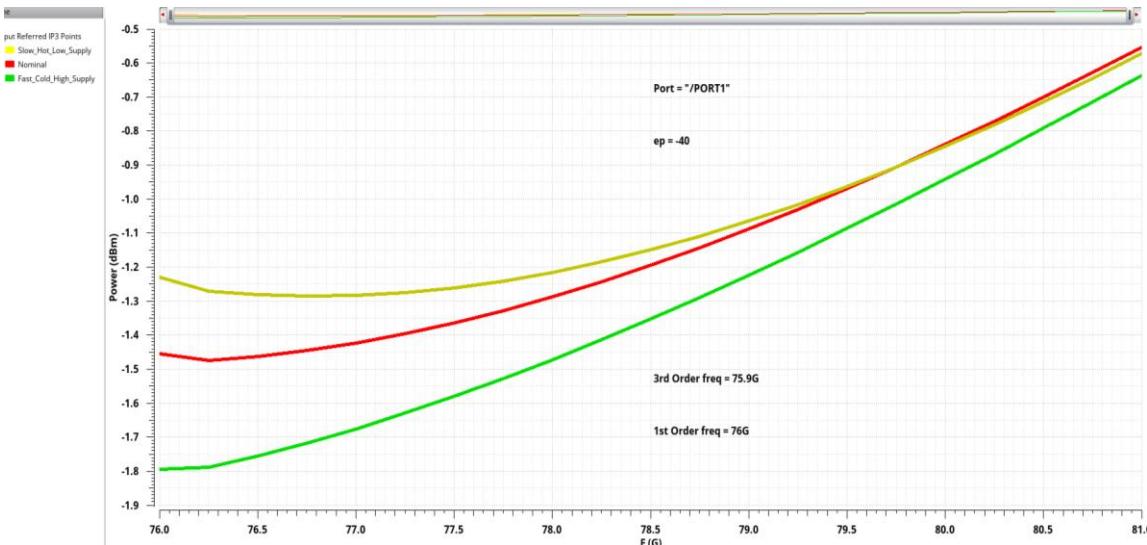


Fig. 7.48: IIP3 vs Frequency across corners.

Fig. 7.48 shows the LNA IIP_3 versus frequency across different corners. We notice that the worst linearity occurs at the fast corner (green curve) as the IIP_3 reaches a minimum of -1.8 dBm . this is acceptable compared to the target spec of $IIP_3 > -14\text{dBm}$.

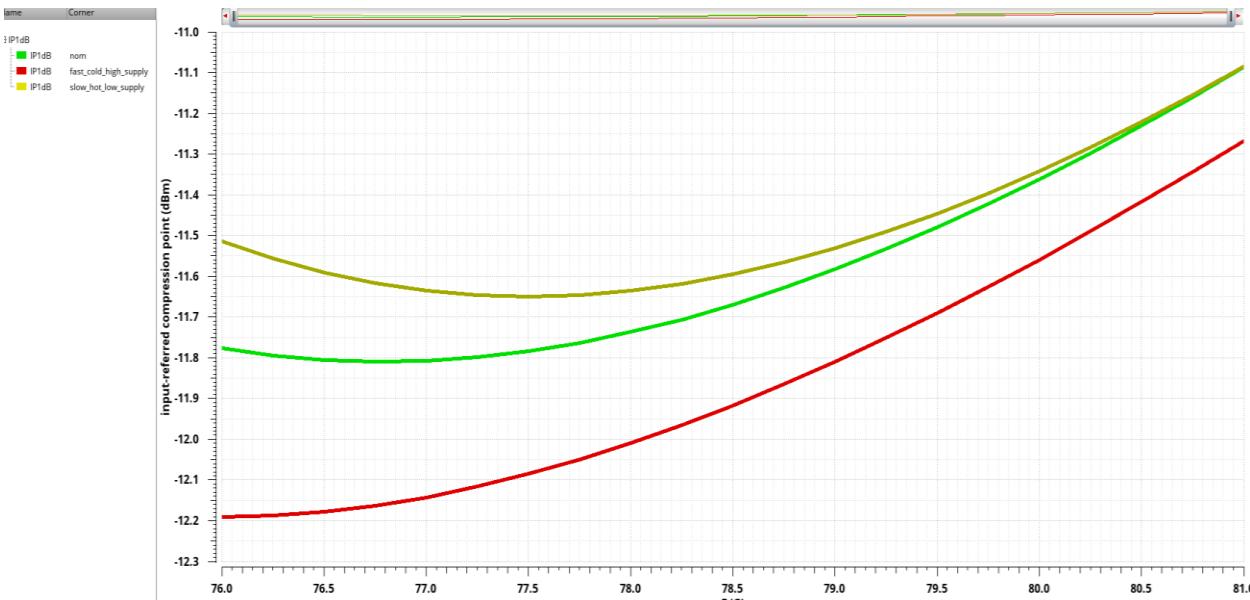


Fig. 7.49: IP_{1dB} vs Frequency across corners.

Fig. 7.49 shows the LNA IP_{1dB} versus frequency across different corners. We notice that the worst linearity occurs at the fast corner (red curve) as the IP_{dB} reaches a minimum of -12.2 dBm . this is acceptable compared to the target spec of $IP_{1dB} > -24 \text{ dBm}$.

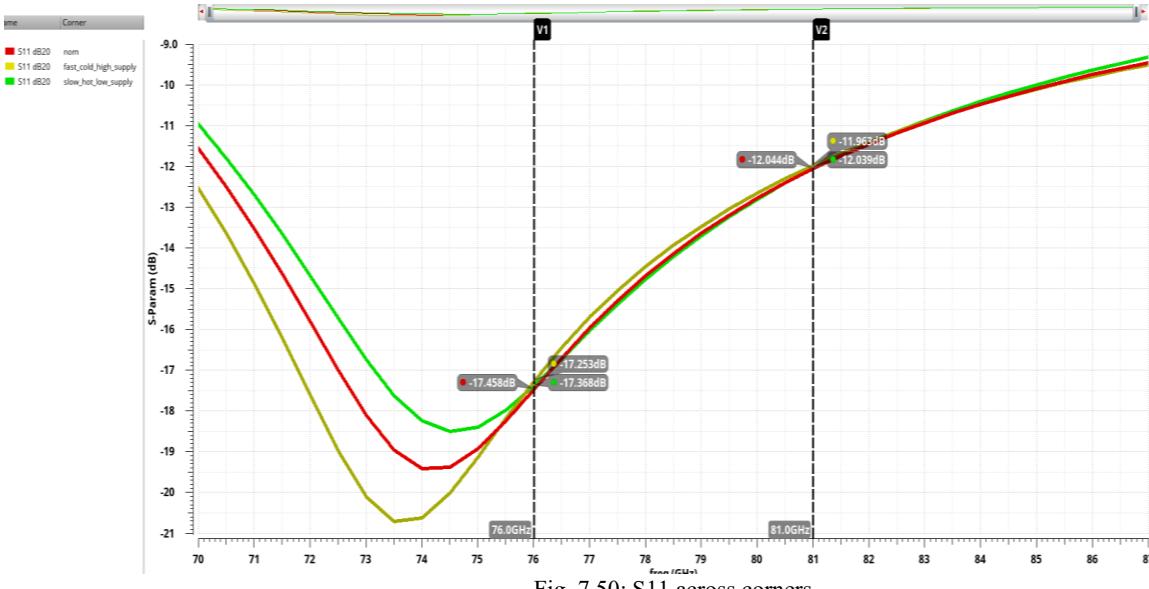


Fig. 7.50: S₁₁ across corners.

Fig. 7.50 shows the LNA input return loss (S_{11}) across the three corners. The achieved S_{11} stays greater than -10 dB across the band $76 - 81 \text{ GHz}$ across different corners.

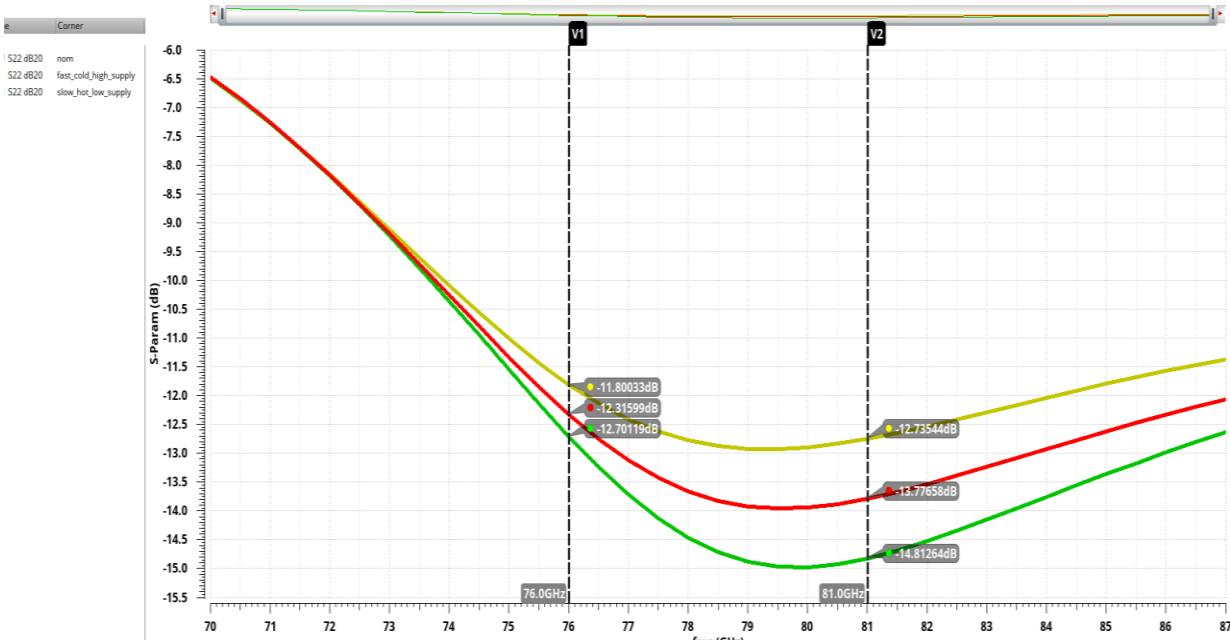


Fig. 7.51: S22 across corners.

Fig. 7.51 shows the LNA output return loss (S_{22}) across the three corners. The achieved S_{22} stays greater than -10dB across the band $76 - 81\text{GHz}$ across different corners.

7.6.3. LNA Layout.

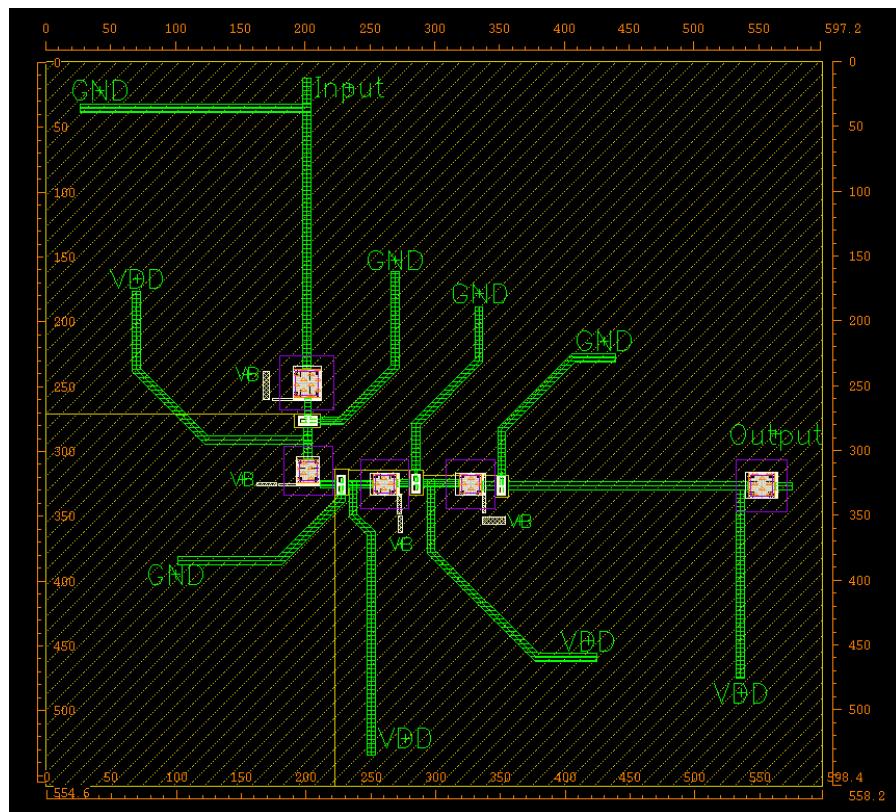


Fig. 7.52: LNA layout.

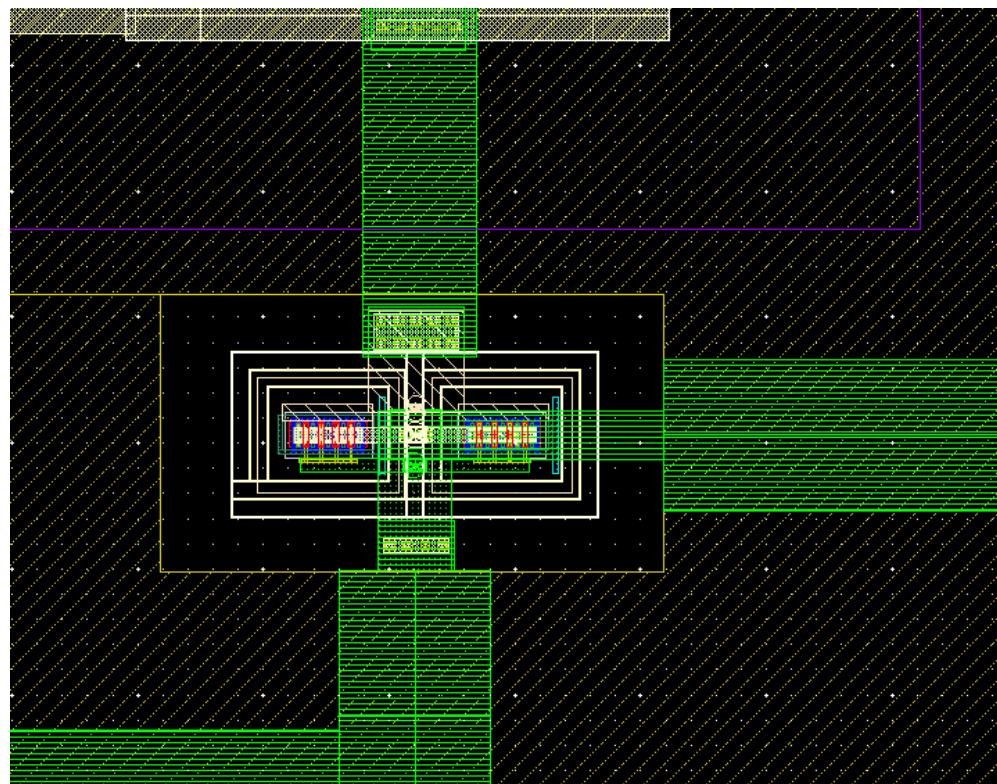


Fig. 7.53: Device Layout.

References

- [1]. Razavi, B. (2014) 'BASIC CONCEPTS IN RF DESIGN', in *RF Microelectronics*. Inde: Dorling Kindersley.
- [2]. Voinigescu, S. (2013) 'High-frequency linear noisy network analysis', in *High-frequency integrated circuits*. Cambridge: Cambridge University Press.
- [3]. Gonzalez, G. (2008) 'Microwave transistor amplifier design', in *Microwave transistor amplifiers: Analysis and Design*. 2nd edn. Taipei: Pearson Education.
- [4]. Heydari, B. et al. (2007) 'A 60-GHz 90-nm CMOS cascode amplifier with interstage matching', 2007 European Microwave Integrated Circuit Conference [Preprint]. doi:10.1109/emicc.2007.4412654.
- [5]. Lee, J. et al. (2010) 'A fully-integrated 77-GHz FMCW radar transceiver in 65-nm CMOS technology', *IEEE Journal of Solid-State Circuits*, 45(12), pp. 2746–2756. doi:10.1109/jssc.2010.2075250.
- [6]. Luo, T.-N., Wu, C.-H.E. and Chen, Y.-J.E. (2013) 'A 77-GHz CMOS automotive radar transceiver with anti-interference function', *IEEE Transactions on Circuits and Systems I: Regular Papers*, 60(12), pp. 3247–3255. doi:10.1109/tcsi.2013.2265974.
- [7]. Le, H.V. et al. (2012) 'A 77 GHz CMOS low noise amplifier for automotive radar receiver', 2012 *IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)* [Preprint]. doi:10.1109/rfit.2012.6401651.
- [8]. Y. Kawano, T. Suzuki, M. Sato, T. Hirose and K. Joshin, "A 77GHz transceiver in 90nm CMOS," 2009 IEEE International Solid-State Circuits Conference - Digest of Technical Papers, San Francisco, CA, USA, 2009, pp. 310-311,311a, doi: 10.1109/ISSCC.2009.4977432

8

MIXER

The mixer is a block that translates a signal from one frequency to another. In the receiver, it translates the signal from radio frequency (RF) to intermediate frequency (IF) or baseband according to the receiver architecture. This is called down-conversion mixing. And in the transmitter translates the signal from (IF) or baseband to radio frequency (RF). This is called up-conversion mixing. The mixer has nonlinear characteristics like a diode or a transistor. In transistors, the two signals which will be mixed can be applied to two different terminals. And in diodes, some circuitry is needed for the two signals to be mixed. The nonlinear circuit can produce the sum and the difference of the two-input signals, RF and local oscillator (LO) in the case of the receiver, or IF and LO in the case of the transmitter, $\omega_1 \pm \omega_2$, where ω_1 and ω_2 are the two-input signals.

8.1. Mixer basic operation

Assuming a down-conversion mixer, the IF signal can be produced by multiplying the RF signal and LO signal:

$$\cos(\omega_{RF}t) * \cos(\omega_{LO}t) = 0.5 * [\cos((\omega_{RF} - \omega_{LO})t) + \cos((\omega_{RF} + \omega_{LO})t)] \quad 8.1$$

Equation (8.1) can be represented by a circuit equation in which a current waveform is generated by the product of the RF signal and a time-varying conductance, $g(t)$ which varies at LO frequency:

$$i(t) = v_{RF}(t) * g(t), \text{ where } g(t) = \alpha v_{LO}(t) \quad 8.2$$

This can be generated using a voltage-controlled circuit element:

$$g[V] = a + bV + cV^2 + dV^3 + \dots, \text{ where } a, b, c, d \text{ are constants.} \quad 8.3$$

So, the time-varying signal which is generated by the LO signal has a large number of harmonics that mix with the RF signal and result in a lot of mixing products at IF output. These mixing products can be represented by:

$$\omega_n = \pm\omega_{RF} \pm n\omega_{LO}, n = 0, 1, 2, \dots \quad 8.4$$

But in reality, RF signal does not have only one component at RF frequency, but it has harmonics. So, eqn. (8.4) can be represented as:

$$\omega_{m,n} = \pm m\omega_{RF} \pm n\omega_{LO}, m, n = 0, 1, 2, \dots \quad 8.5$$

Using a nonlinear circuit element gives the desired mixing product and undesired spurious mixing products that should be filtered out.

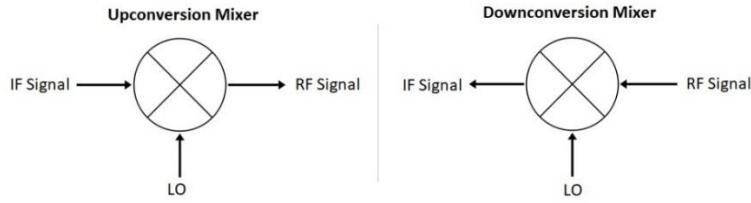


Fig 8.1: Up-conversion and down-conversion mixers.

And the most likely frequency spectrum of IF receiver architecture:

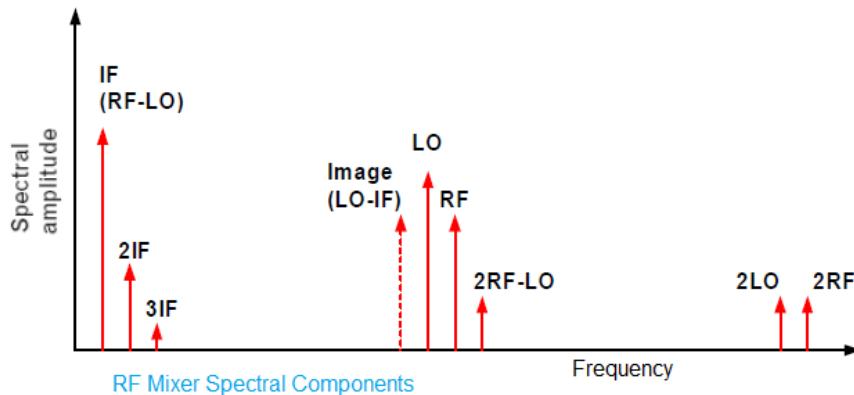


Fig 8.2:frequency spectrum of the mixer.

8.2. Mixer performance metrics

8.2.1. Conversion gain/loss

Conversion gain or loss is the ratio between available RF(IF) output power and the IF(RF) input power. In a single-sideband (SSB) system, only one sideband of the signal is converted, hence, 3 dB of conversion loss is unavoidable. Other sources of losses are input and output mismatch losses, and device and circuit losses. In a double-sideband (DSB) system, both RF signals which are below and above the LO are converted into IF. They are called “images”. Notice that active mixers can provide conversion gain instead of conversion loss in the case of using passive mixers. Down-conversion mixers should provide sufficient gain to adequately suppress the noise contributed by subsequent stages. However, low supply voltages make it difficult to achieve a gain of more than roughly 10 dB while retaining linearity.

8.2.2. Noise

The noise factor is the ratio between total output noise and total input noise. The most important noise components that mix and appear at the output are the noise at the fundamental input frequency and the image frequency. Mixer noise has two types, SSB and DSB noise. DSB noise contains the noise at the fundamental and image frequencies. On the other hand, SSB noise contains the noise at the fundamental frequency but not the image frequency noise, as shown in fig. (8.3) and fig. (8.4). In a receive chain, the

input noise of the mixer following the LNA is divided by the LNA gain when referred to the RX input.

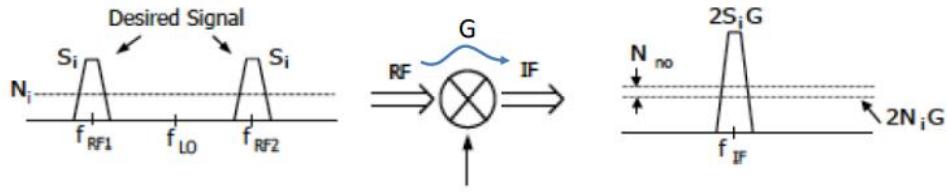


Fig 8.3: DSB Noise Figure.

$$NF_{DSB} = \frac{SNR_i}{SNR_o} = \frac{S_i}{N_i} \frac{2N_i G + N_{no}}{2S_i G} = 1 + \frac{N_{no}}{N_i G} \quad 8.6$$

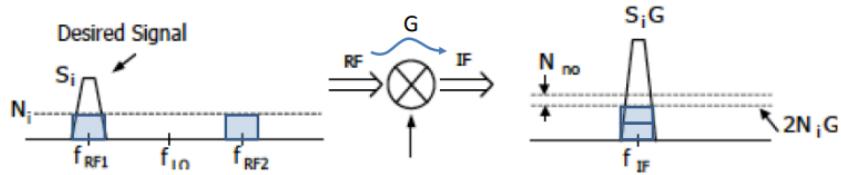


Fig 8.4: SSB Noise Figure.

$$NF_{SSB} = \frac{SNR_i}{SNR_o} = \frac{S_i}{N_i} \frac{2N_i G + N_{no}}{S_i G} = 2 + \frac{N_{no}}{N_i G}. \quad 8.7$$

8.2.3. Intermodulation and linearity

As in microwave amplifiers, an intermodulation (IM) test is done on the mixer. As said in the previous sections, a mixer is a nonlinear circuit device. Hence, the output components can be considered IM products except for the harmonics. This causes intermodulation distortion (IMD) products at the output spectrum. Assuming a down-conversion mixer, unwanted mixing products are in the form of $(\pm m\omega_{RF} \pm n\omega_{LO})$. When doing the IM test, 3 tones are applied, one large LO tone, ω_{LO} and two small RF(IF) tones, ω_{RF1} and ω_{RF2} . In addition to the intended components, $\omega_{LO} - \omega_{RF1}$ and $\omega_{LO} - \omega_{RF2}$, there are undesired components that also appear at the output:

$$\omega_{LO} - [N\omega_{RF1} - (N-1)\omega_{RF2}] \text{ (upper IM sidebands)} \quad 8.8$$

$$\omega_{LO} - [N\omega_{RF2} - (N-1)\omega_{RF1}] \text{ (lower IM sidebands)} \quad 8.9$$

Where $N= 2, 3, 4, \dots$. The most significant ones are the third-order ones, $\omega_{LO} - (2\omega_{RF1} - \omega_{RF2})$ and $\omega_{LO} - (2\omega_{RF2} - \omega_{RF1})$ because they are the closest ones to the desired IF output component, as shown

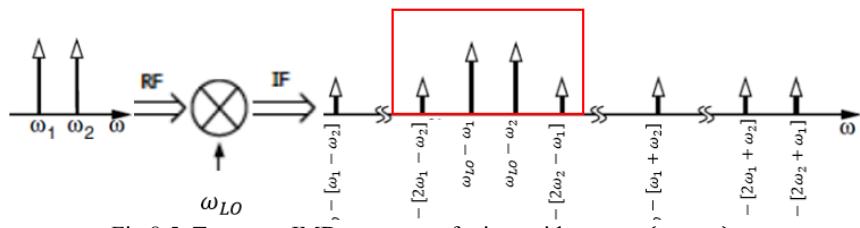


Fig 8.5: Two-tone IMD spectrum of mixer with $\omega_{LO} > (\omega_1, \omega_2)$

in Fig. (8.5).

8.2.4. Harmonic suppression

Harmonic suppression is very important for mixer linearity. LO harmonics can be suppressed by using a suitable mixer topology that inherently suppresses some of them, by using harmonic traps, or at least resorting to using a filter. There are mixers, which are called sub-harmonically pumped mixers that use the harmonics for the mixing operation. They are used at mm-wave range where LO frequency is very large which costs power in the LO chain and complexity in the phase locking.

8.2.5. Isolation

Isolation means low leakage between different mixer ports, there are multiple types of isolation: LO-to-RF, LO-to-IF, and RF-to-IF isolation. Port isolation is so important in determining the amount of filtering required before and after the mixer. Since the LO signal is quite large as compared to the RF signal, any LO-RF feedthrough or leakage, if not filtered out, may cause problems in the following stages of the chain, in addition, large RF and LO feedthrough signals at the IF output may saturate the IF port and decrease the PldB of the mixer. For example, if LO leaks to RF in up-conversion mixers, it may saturate the following block because the LO signal is very large. In single-ended mixers, isolation is achieved using filters, but in balanced mixers, isolation is inherently achieved due to circuit symmetry. In the direct-conversion receiver, LO-RF feedthrough is very critical. It is determined by the symmetry of the mixer circuit and LO waveforms, The LO-IF feedthrough is heavily suppressed by the baseband low-pass filter. The RF-IF feedthrough corrupts the baseband signal by the beat component resulting from even-order distortion in the RF path. And in Heterodyne Receivers, the LO-RF feedthrough is relatively unimportant, The LO-IF feedthrough, becomes serious if ω_{IF} and ω_{LO} are too close to allow filtering of the latter.

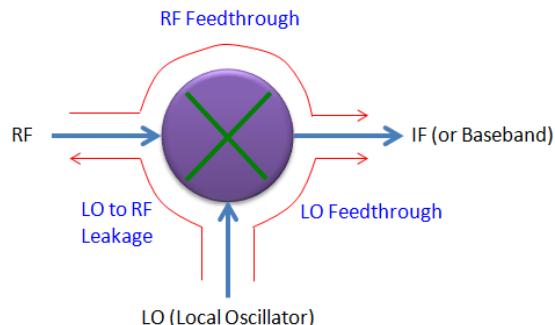


Fig 8.6: Types of feedthrough in mixers.

8.2.6. Operating frequency range

The operation band of the transceiver determines the RF band and the IF band of the mixer. LO frequency could be high-side injection where LO frequency is greater than RF frequency or low-side injection where LO frequency is smaller than RF frequency. Notice that, RF and IF ports are matched to achieve the maximum power transfer flow. Whereas, the LO port is matched to deliver the required LO swing at the LO port of the mixer.

8.3. Mixer Topologies

There are many topologies of mixers depending on the frequency range of operations:

8.3.1. Diode mixer

Diodes are nonlinear circuit elements due to their exponential nature characteristics. Minimum loss can be achieved by dissipating minimum input power in the nonlinear elements and terminating all undesired IMs reactively. The former can be achieved by pumping the LO port strongly; hence the states are fully on or off.

8.3.2. Single-ended diode mixers

It consists of the mixing element which is the diode and passive networks for impedance matching, signal isolation, and LO harmonic and undesired mixer IM products termination. In narrow-band designs, LO harmonics can be terminated using stubs. The disadvantages of single-ended diode mixers are special filtering to isolate LO and RF ports, and noise figure and conversion gain degradation due to the directional coupler or loop filter, which is used to bring the LO signal.

8.3.3. Single-balanced diode mixers

Single-balanced diode mixer (SBM) can be implemented using two single-ended diode mixers with the need for a balun. The balun can be replaced by 90° or 180° hybrids. In the former, if the balun was transformer-based, the RF signal can be applied at the center tap of the balun. The advantage of the balance-diode mixers is the suppression of some undesired spurious IM mixing products of $\pm m\omega RF \pm n\omega LO$ depending on the hybrid type used or using a balun. The use of 180° hybrid or balun for having differential LO signal suppresses all mixing products corresponding to n even. The use of 90° hybrid suppresses all the mixing products corresponding to m and n are even. In addition to suppression of some undesired spurious IM mixing products, the 180° hybrid gives better LO/RF isolation. On the other hand, a 90° hybrid gives better LO and RF matching. Single-ended and single-balanced diode mixers have the same problems, which is insufficient LO-RF isolation.

8.3.4. Double-balanced diode mixers

Double-balanced diode mixer solves the problems of single-ended and single-balanced diode mixers. It can be implemented by combining two SBMs in parallel and 180° out of phase. Due to increasing the number of diodes, it requires more LO power than the previous topologies.

8.3.5. Resistive passive mixer

The double-balanced NMOS passive ring mixer is shown in Fig. (8.7). It is called a passive mixer because no DC in transistors. The main disadvantage is that it does not have gain. The conversion loss without mismatch and device losses is:

Transistors are sized such that the design can provide good switching, maximize linearity, and minimize noise and loss.

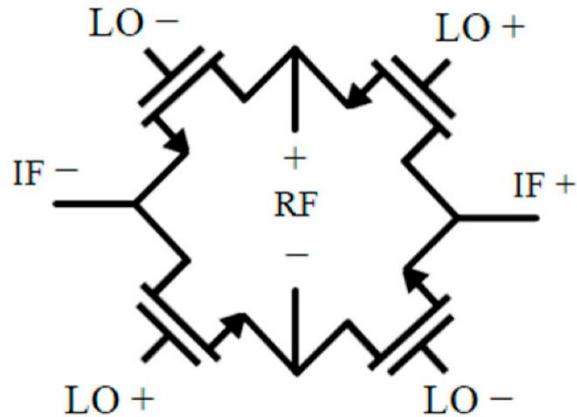


Fig 8.7: Double-balanced NMOS passive ring mixer.

And the process of it can be described in fig. (8.8) as shown:

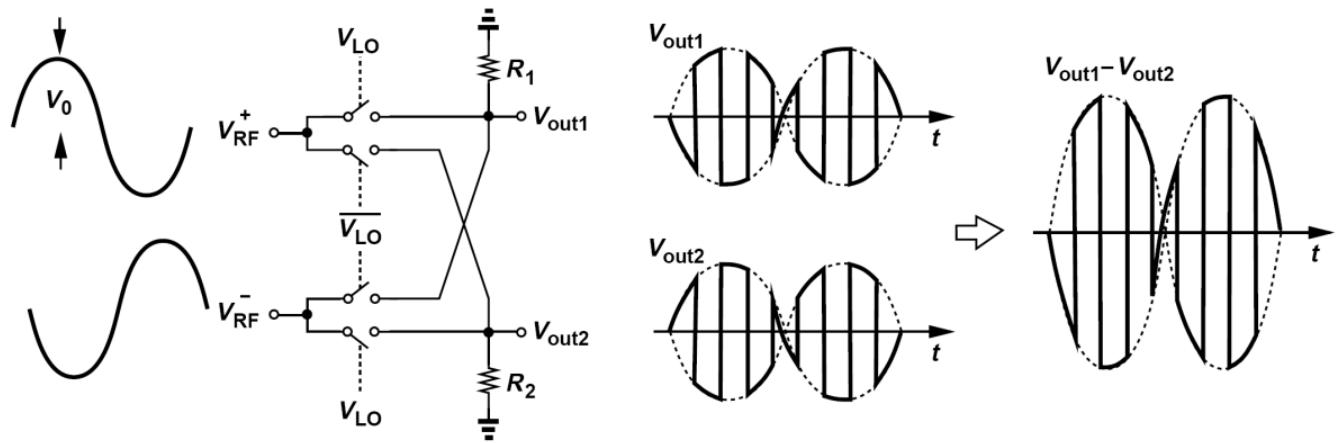


Fig 8.8: the whole process of the passive ring mixer.

8.3.6. Active FET mixer

The input of the mixer is a gain stage. The gain stage converts the voltage into amplified current and this current passes to the switching stage. The switching stage is excited by a square wave that steers the current between the two sides periodically as shown in Fig. (8.9):

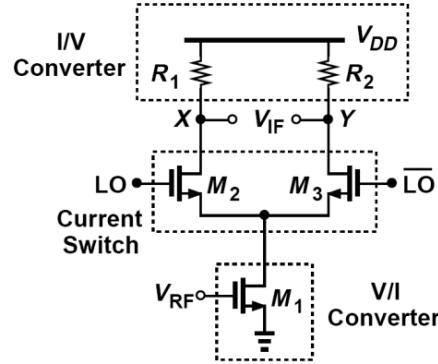


Fig 8.9: Single-balanced Active FET Mixer.

8.3.7. Single-ended FET mixers

Single-gate single-ended FET mixer is shown in fig. (8.9). LO signal modulates the FET transconductance. Then, the RF signal mixes with gm . As mentioned in the single-ended diode mixer, a filter is required at LO/RF input for proper isolation. IF short circuit termination is required at the gate to prevent amplifying IF noise which may lead to instability issues. It has the same disadvantages as the single-ended diode mixer.

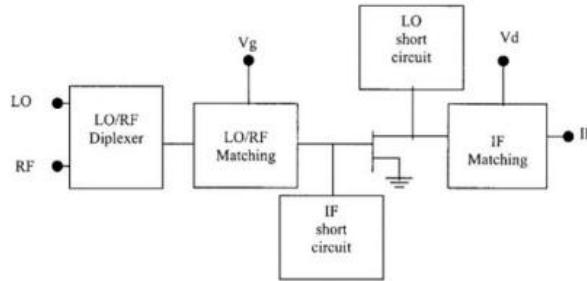


Fig 8.10: Single-ended FET mixer.

8.3.8. Single-balanced FET mixers

It can be established by combining two single-ended FET mixers in parallel as shown in fig. (8.10). The LO differential voltage causes the source-drain node to be a virtual ground which leads to LO/RF isolation. Single-ended and single-balanced FET mixers have the same disadvantages as single-ended and single-balanced diode mixers which are insufficient isolation.

8.3.9. Double-balanced FET mixers

It is based on the Gilbert cell as shown in Fig. (8.11). Assuming a down-conversion mixer, the gain is:

$$\frac{v_{IF}}{v_{RF}} = u(V_{LO}) * gmRD \quad 8.10$$

Where $u(V_{LO})$ is the function in LO swing, which is a square wave:

$$u(v_{LO}) = \frac{4}{\pi} [\sin(\omega_{LO}t) + \frac{1}{3} \sin(3\omega_{LO}t) + \frac{1}{5} \sin(5\omega_{LO}t) + \dots] \quad 8.11$$

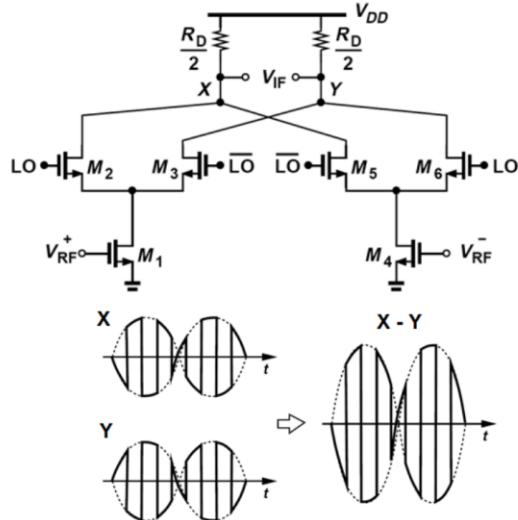


Fig 8.11: Active double-balanced FET mixer based on Gilbert Cell.

8.4. Down-converter mixer

The downconverter mixer core designed in this thesis is a double-balanced passive MOSFET ring mixer. This architecture is selected due to its potential to realize an excellent compromise between the various requirements discussed above especially since the system design level does not require gain from the mixer. This core is then used in an IR architecture to receive one of the sidebands only. This is done instead of using a band-pass filter as the two sides are close to each other and therefore it is not easy to design a very sharp filter.

8.4.1. Mixer core

The mixer is a very non-linear device that is responsible for frequency translation. The layout is significantly critical to the performance of the total chain and therefore needs to be optimized as possible. The LO signal and the RF signal are close to each other in frequency due to using low IF. Hence, it is required to minimize the coupling between them by avoiding any crossing, as shown in Fig. (8.12).

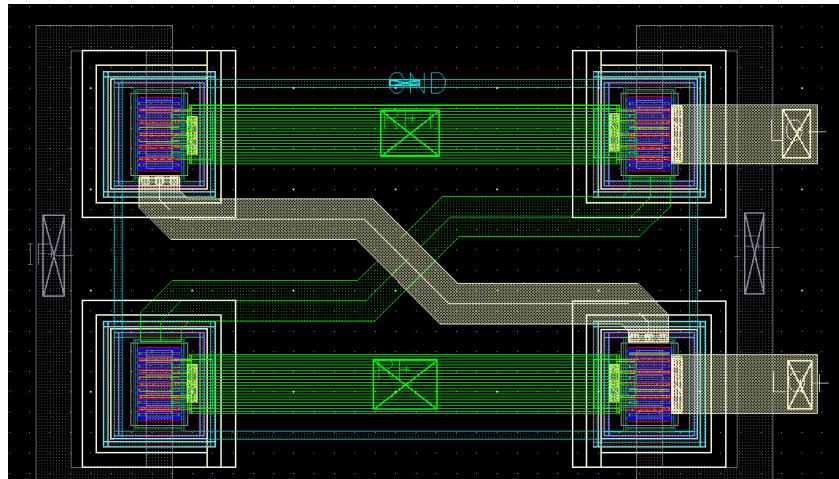


Fig 8.12: Downconverter Mixer core layout.

8.4.2. Mixer Core Simulation Results

The following are the results of the mixer core after EMing all the Transmission Lines and extracting the layout at three different corners corner.

8.4.2.1. Output (IF) matching

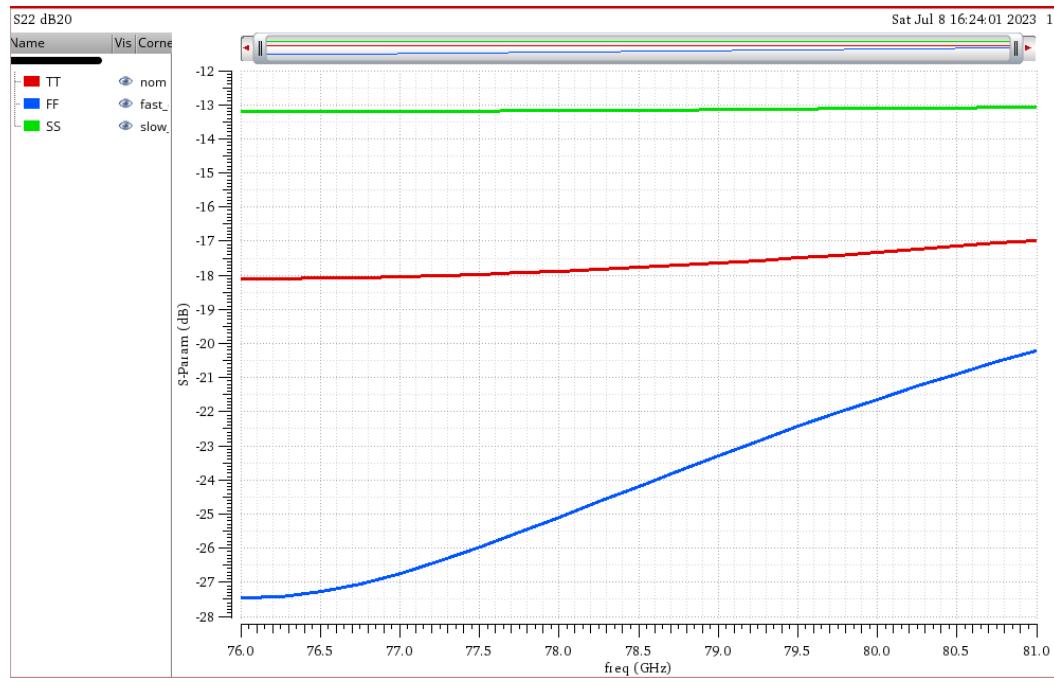


Fig. 8.13: Downconverter Mixer core: S22

Fig. (8.13) shows that $S22 < -13 \text{ dB}$ across the band of operation band within the different corners.

8.4.2.2. Input (RF) matching

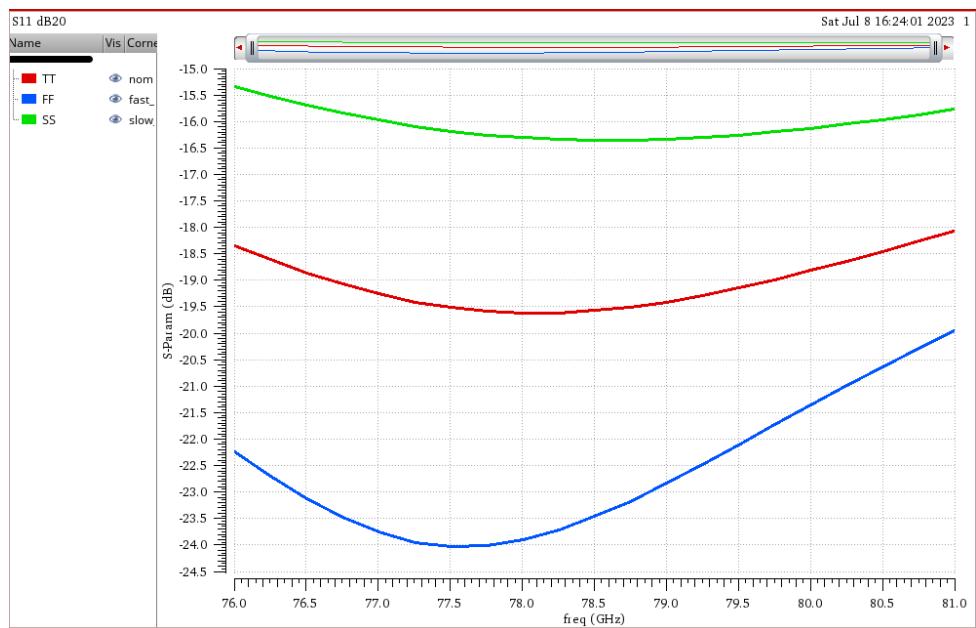


Fig. (8.14) shows that $S_{11} < -15 \text{ dB}$ across the whole band within the different corners.

8.4.2.3. Conversion gain

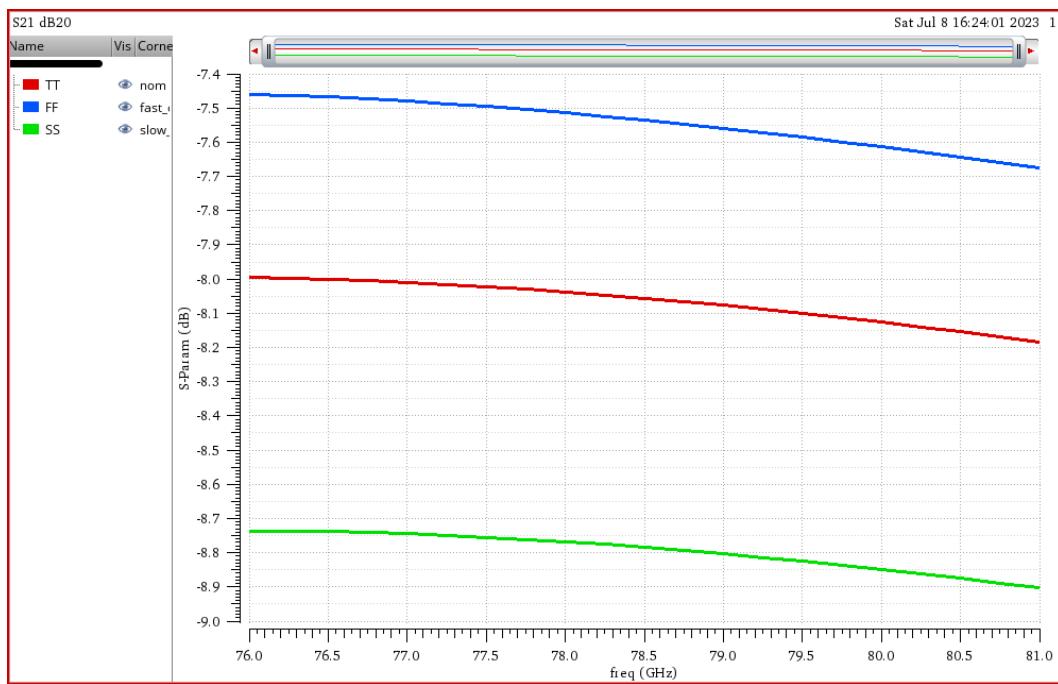


Fig. 8.15: Downconverter Mixer core: S21

Fig. (8.15) shows that S_{21} varies from -7.5 dB to -8.5 dB with maximum variations of 1 dB across the band of operation within the different corners.

8.4.2.4. Noise Figure

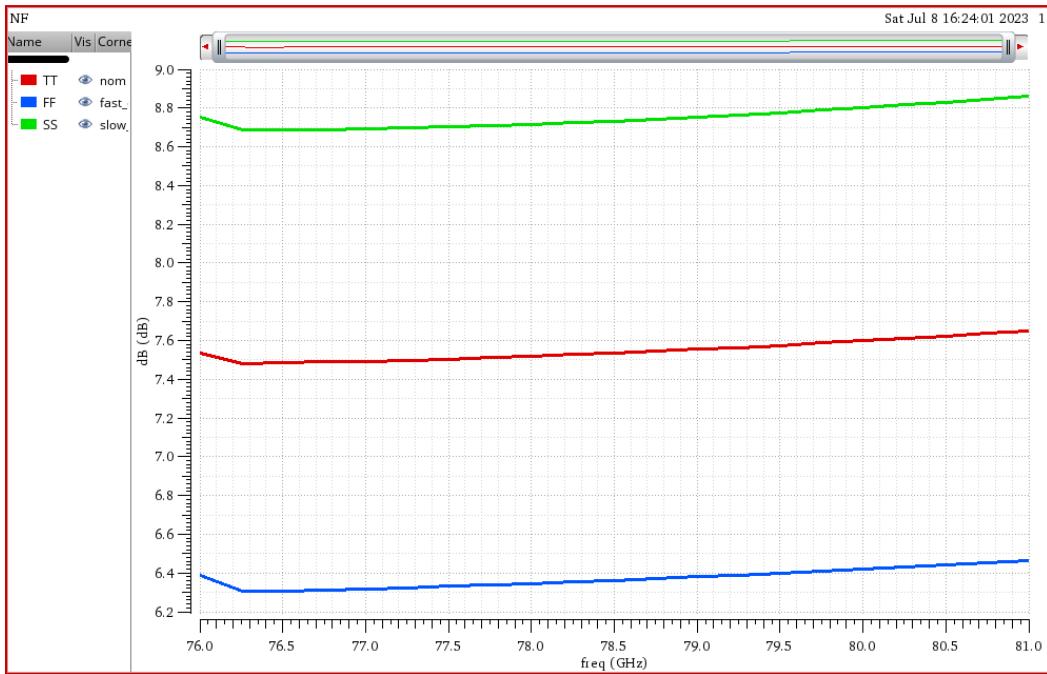


Fig. (8.16) shows that **NF** varies maximum variations of **2 dB** within the different corners.

OIP3.

Fig.8.16: Downconverter Mixer core: S21.

from **8.5 dB** to **6.5 dB** with across the band of operation

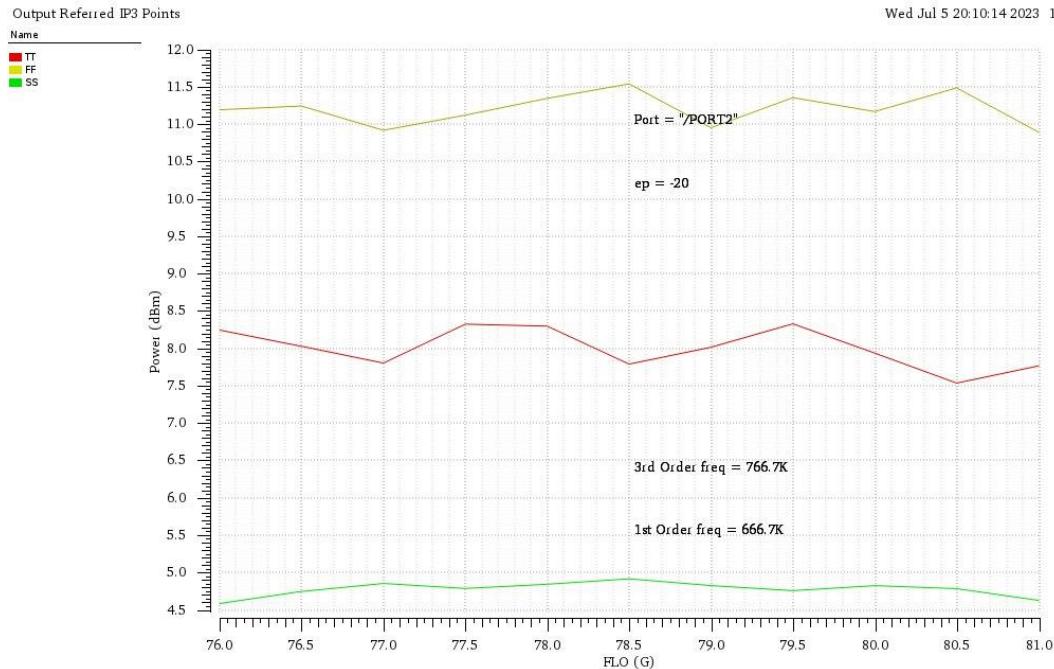


Fig. 8.17: Downconverter Mixer core: OIP3.

Fig. (8.17) shows that **OIP3** varies from **11 dB** to **5 dB** with maximum variations of **6 dB** across the band of operation within the different corners.

8.4.2.5. OIP3

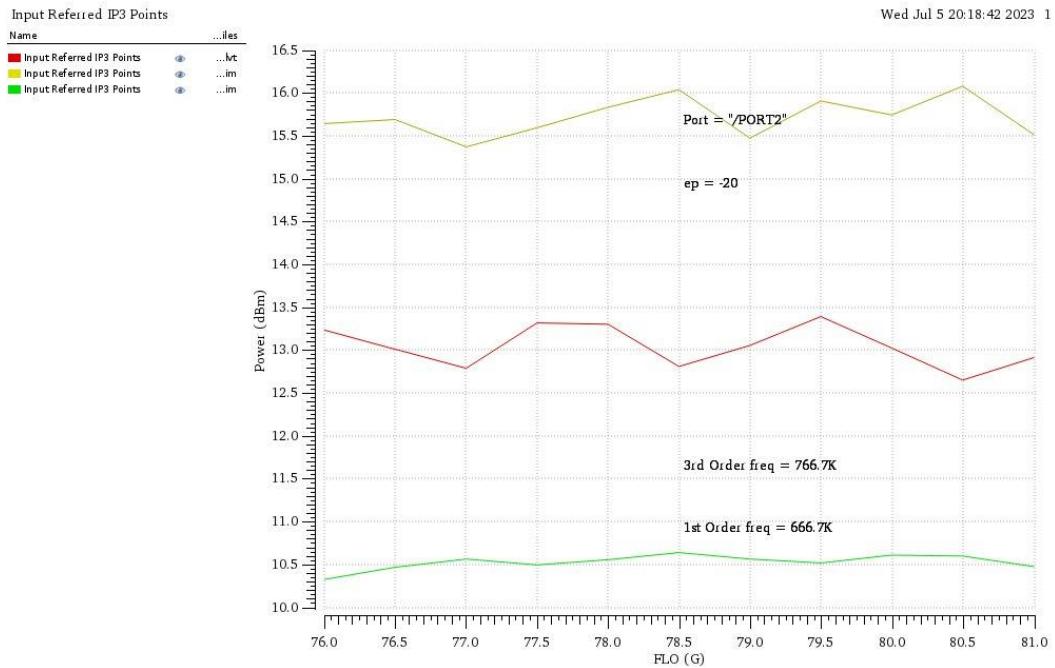


Fig. (8.18) shows that **IIP3** varies maximum variations of **5 dB** within the different corners.

Fig. 8.18: Downconverter Mixer core: IIP3

from **15.5 dB** to **10.5 dB** with across the band of operation

8.4.2.6. LO-RF Port Isolation

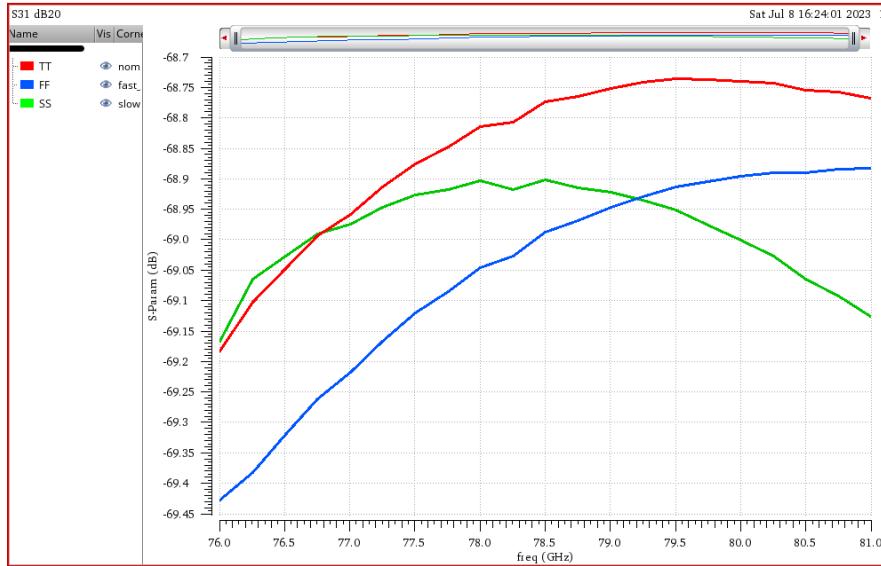


Fig. (8.19) shows that **S31 < -68 dB** over the whole band within the different corners.

8.4.2.7. LO-IF Port Isolation

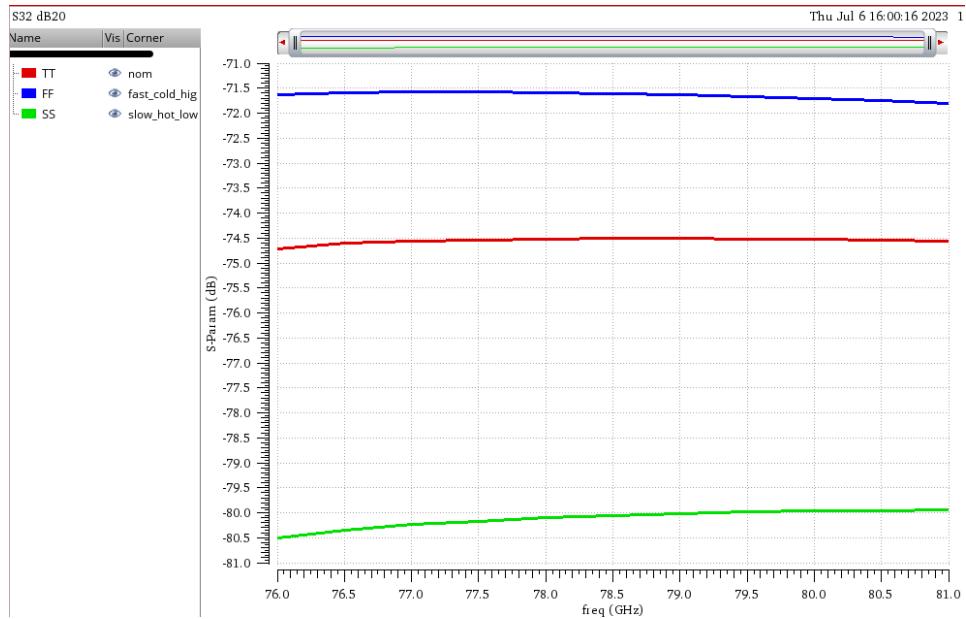


Fig. 8.19: Downconverter Mixer core: S31

Fig. (8.20) shows that $S32 < -71 \text{ dB}$ over the whole band within the different corners.

Table 8.1: Required VS achieved specifications.

Down Conversion Mixer		
	Required	Achieved
RF Frequency	76G : 81G	76G : 81G
LO Frequency	76G : 81G	76G : 81G
IF Frequency	0 : 5G	0 : 5G
Conversion gain	>-10 dB	-8.1 dB
NF	<10 dB	7.5 dB
OIP3	>4 dB	5 dB
Area	--	
Power consumption	--	

Table 8.2: LNA literature survey

Ref	[1]	[2]	[3]	[4]	[5]	[6]	This work
Process	90-nm CMOS	65-nm CMOS	SiGe (200 GHz f_T)	90-nm CMOS	90-nm CMOS	0.18-um CMOS	65-nm CMOS
Topology	Double-Balanced Gilbert-Cell with 2 Marchand Baluns	Double-Balanced Gilbert-Cell with 2 Baluns	Double-Balanced Gilbert-Cell with 2 Baluns	Drain pumped	Gate pumped	Sub-harmonic	Passive fully balanced ring mixer
RF Band (GHz)	75:85	76:77	75:110	35	27	28	76:81
IF Band (GHz)	0.1	6	0.5	2.5	2.5	1.6	0:5
CG (dB)	1.5	-8	14.4	-4.6	-10.3	-11.15	-8.1
IIP3 (dbm)	2.7	2.5	-6	2	12.7	8	13
NF (dB)	23.3	17.8	19.5	7.9	11.4	11.6	7.5
LO-RF isolation	49.2 dB	21 dB	30 dB	11 dB	24 dB	63 dB	70 dB
LO-IF isolation	64.5 dB	32 dB	--	45 dB	22 dB	--	71 dB
IF-RF isolation	39.4 dB	--	--	48 dB	33 dB	--	56 dB

Table 1.2: This work VS other papers

References

- [1] Y. Lin and G. Li, "W-band down-conversion mixer in 90 nm CMOS with excellent matching and port-to-port isolation for automotive radars," 2014 11th International Symposium on Wireless Communications Systems (ISWCS), Barcelona, 2014.
- [2] N. Zhang, H. Xu, H. T. Wu, and Kenneth K. O., "W-Band Active Down-Conversion Mixer in Bulk CMOS," IEEE Microwave and Wireless Components Letters, vol. 19, no. 2, pp. 98-100, Feb. 2009.
- [3] J. Kim, K.T. Kornegay, J. Alvarado Jr., C. H. Lee and J. Laskar, "W-band double-balanced down-conversion mixer with Marchand baluns in silicon-germanium technology," Electronics Letters, vol. 45, no. 16, pp. 841–843, Jul. 2009.
- [4] M. T. K. M. J. H. T., "30-40-GHz drainpumped passive-mixer MMIC fabricated on VLSI SOI CMOS technology," Microwave Theory and Techniques, IEEE Transactions, vol. 52, no. 5, pp. 1382-1391, 2004.
- [5] F. Ellinger, "26.5–30-GHz Resistive Mixer in 90-nm VLSI SOI CMOS Technology with High Linearity for WLAN," Microwave Theory and Techniques, IEEE Transactions, vol. 53, no. 8, pp. 2559-2565, 2005.
- [6] T.-Y. Y. a. H.-K. Chiou, "A 28 GHz sub-harmonic mixer using LO doubler in 0.18-/spl mu/m CMOS technology," IEEE Radio Frequency Integrated Circuits (RFIC) Symposium, 2006.

9

POWER AMPLIFIER

The power amplifier is a crucial component in radar transmitters, it is required to deliver sufficient power at millimeter-wave frequencies to drive the antenna and transmit the radar signal effectively. The power amplifier establishes the power requirements (and so too the battery performance) of the system as it is the most power-hungry block in the chain, it is required to deliver a specified output power while maintaining high linearity and efficiency.

9.1. Power matching and load-pull technique

One of the main differences between power amplifiers and linear RF amplifiers is the power matching at the output which is not the same as the usual conjugate (gain) match the concept of power match will be explained in this section. The maximum power transfer theorem states that to deliver the maximum power from a source to a load, the load impedance needs to be conjugately matched to the source impedance and that is the concept for conjugate matching which maximizes the gain. But this concept takes no account for the possibility that the generator will have physical limits such as the maximum current it can supply or the maximum voltage it can sustain across its terminals.

As shown in Fig. 9. 1, the gain matching gives a higher gain but with a lower current swing (hence lower output power), this is because the voltage or source is not ideal and has a maximum voltage that it can generate. But in power match the load impedance is matched such that the generator has both maximum voltage and current swing, hence maximum output power given the physical limits of the generator. So hypothetically if there is an ideal generator that has no limits on its voltage and current the gain match

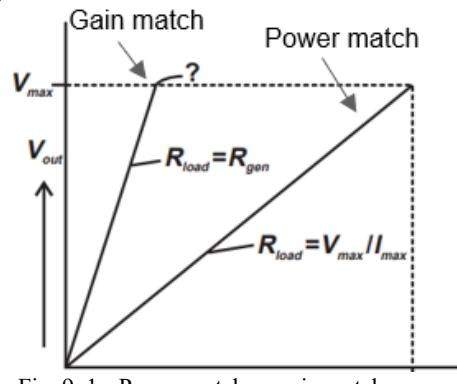


Fig. 9. 1: Power match vs gain match.

will be equivalent to the power match and both maximum gain and maximum power would be achieved but in real cases, this cannot be achieved.

For power match optimum load is calculated such that maximum output voltage and current swings achieved available for the amplifier and theoretically as shown in figure x is given by:

$$R_{LOPT} = \frac{V_{MAX}}{I_{MAX}} \quad 9.1$$

This value R_{LOPT} is usually used as a first guess for the actual optimum load impedance Z_{LOPT} which is calculated by load-pull simulation.

Load-pull simulation is a technique in which the output impedance seen by the amplifier is varied and measures the resulting output power for each impedance the output of the load-pull simulation is represented by load-pull contours as shown in Fig. 9. 2 plotted on the Smith chart representing the output power corresponding to each impedance the point at the center is the optimum load impedance which gives the maximum output power and the first contour represents the impedances at which the output power is less 1dB from the maximum output power and each point inside this contour has powered higher than that at the contour and so on for other contours.

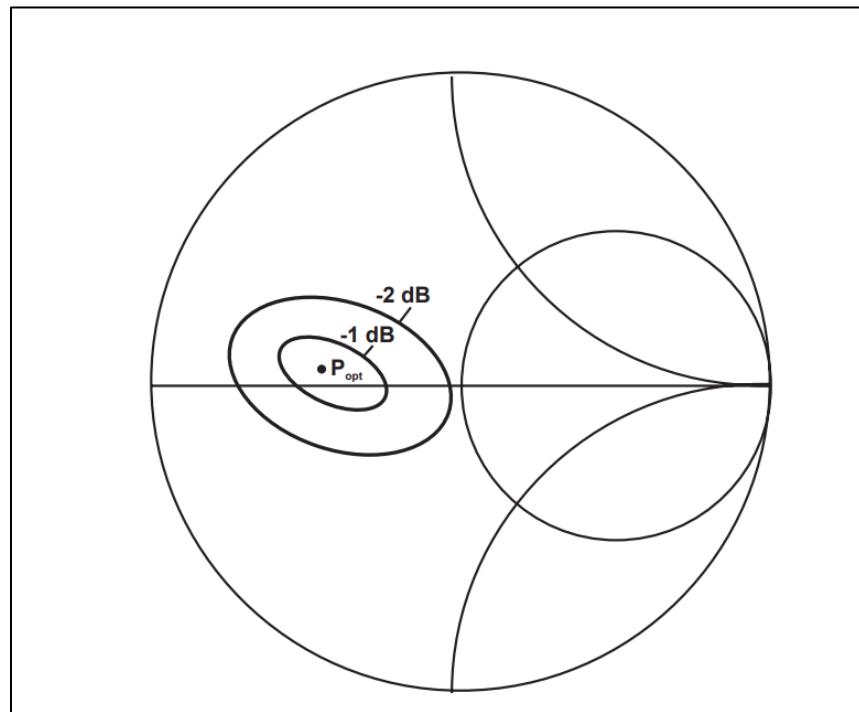


Fig. 9. 2: Typical load-pull data.

9.2. Main power amplifier specifications

9.2.1. Output power

The output power of the PA is usually characterized by the 1 dB compression point OP_{1dB} which (as described in Chapter 3) is the point at which the output power of the power amplifier drops by 1 dB from its linear region. For radar systems as the output signal is frequency modulated and the output power is constant so it can operate at saturation to maximize the power efficiency of the amplifier. So, for radar systems, the output power is usually described by saturation power which is the maximum power that the power amplifier can deliver.

9.2.2. Efficiency

It is a measure of how well the power amplifier utilizes the dc power to deliver a certain output power, higher efficiency indicates less wasted power and heat dissipation and more effective use of resources.

There are two metrics usually used to express efficiency, the first one is drain efficiency (or collector efficiency) which is the ratio between the output power of the PA to the dc power consumed.

$$\eta = \frac{P_L}{P_{DC}} \quad 9.2$$

Where P_L is the average output power delivered to the load of the amplifier and P_{DC} is the average power drawn from the supply.

The other metric is the power added efficiency which takes the driving power of the PA into account and it is the ratio between the power added by the PA to the power dissipated and it is defined as follows:

$$PAE = \frac{P_L - P_{in}}{P_{DC}} \quad 9.3$$

Where P_{DC} is the input power to the power amplifier. For PAs with high gain the PAE approximately approach η .

9.2.3. Linearity

The linearity of the power amplifier is very important, as the nonlinearity causes amplitude distortion and spectral emissions for adjacent channels, the nonlinearity of the system may be described by the following performance metrics:

- 1- OP_{1dB}: The 1 dB compression power is the power at which the gain of the amplifier drops by one dB and this reduction in gain is due to the third-order non-linearity coefficient α_3 which is usually negative which decreases the gain.
- 2- OIP₃: The third-order intercept point which is the power at which the fundamental signal and the third-order intermodulation distortion products intersect.

- 3- Amplitude to amplitude distortion (AM-AM): The AM-AM distortion is a measure of the variation in gain as a function of input amplitude.
- 4- Amplitude to phase modulation (AM-PM): The AM-PM distortion is a measure of the variation in the output phase as a function of the input amplitude.
- 5- Adjacent Channel Power Ratio (ACPR): The ACPR is a measure of the power level of unwanted spectral emissions in adjacent frequency channels which is due to the non-linearity of the PA.

9.3. Classes of power amplifiers

RF power amplifiers consist of a transistor, input network, output network, and RF choke. For the transistor, it can operate as:

- As a dependent-current source.
- As a switch.
- In overdriven mode (Partially as a dependent source and partially as a switch)

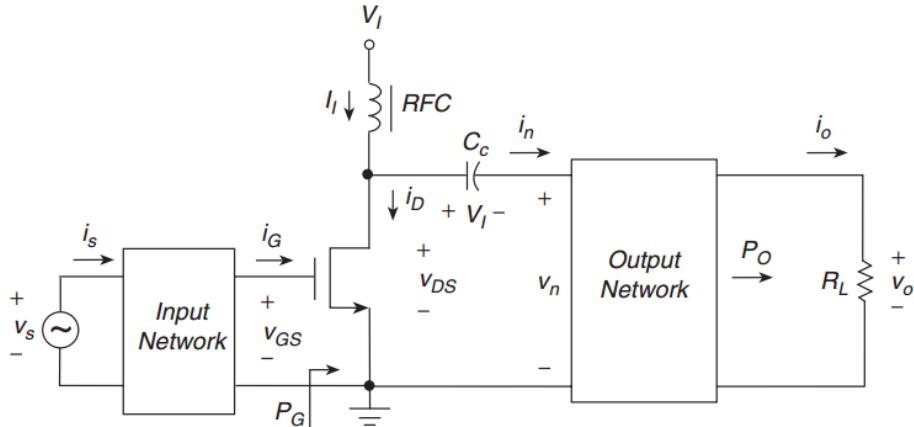


Fig. 9. 3: Block diagram of RF power amplifiers

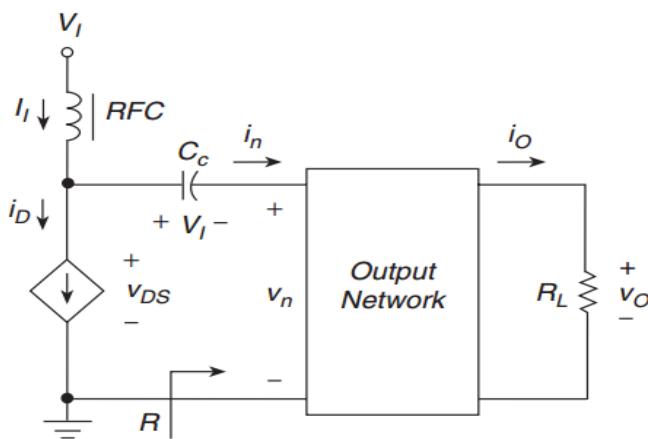


Fig. 9. 4: RF power amplifier with a dependent-source transistor.

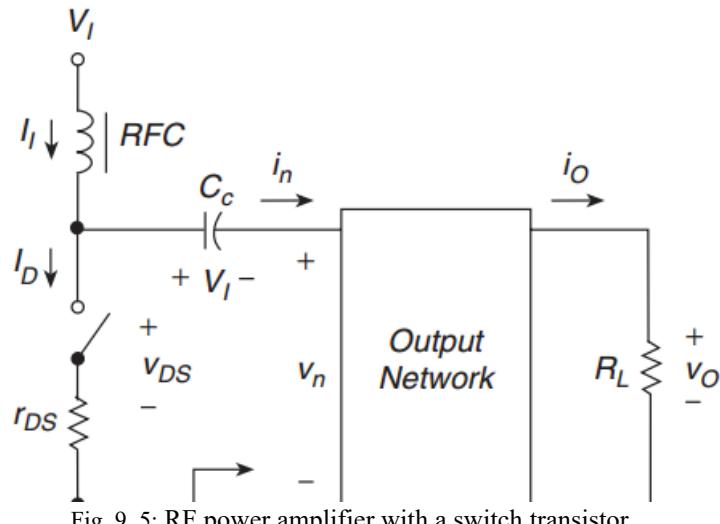


Fig. 9. 5: RF power amplifier with a switch transistor.

The classification of power amplifiers with a transistor operating as a voltage-dependent-current source depends mainly on the conduction angle 2Θ of the drain current. By conduction angle, we mean the portion of the periodic time at which the transistor is on.

9.3.1. Class-A power amplifier

In class-A power amplifiers, the conduction angle of 2Θ is 360^0 which means that the transistor is always on. This is accomplished by making V_{GS} is sufficiently larger than $V_{threshold}$. Specifically, V_{GS} – maximum AC voltage should be greater than $V_{threshold}$. The drain current should also be greater than the maximum value of the AC current.

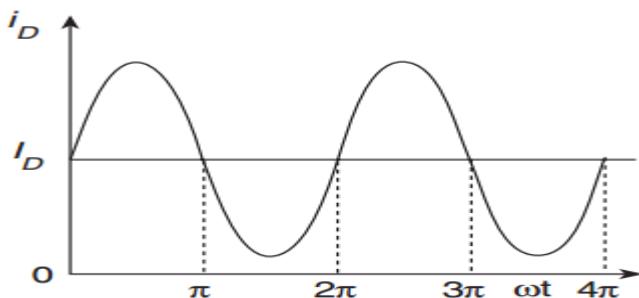


Fig. 9. 6: Waveform of the drain current of class A.

9.3.2. Class-B power amplifier

In class-B power amplifiers, the conduction angle of 2Θ is 180^0 which means that the transistor is on during half of the period. This is accomplished by making V_{GS} equals $V_{threshold}$. So, when the AC component adds to V_{GS} , the transistor conducts only half of the cycle. Consequently, the drain current is zero.

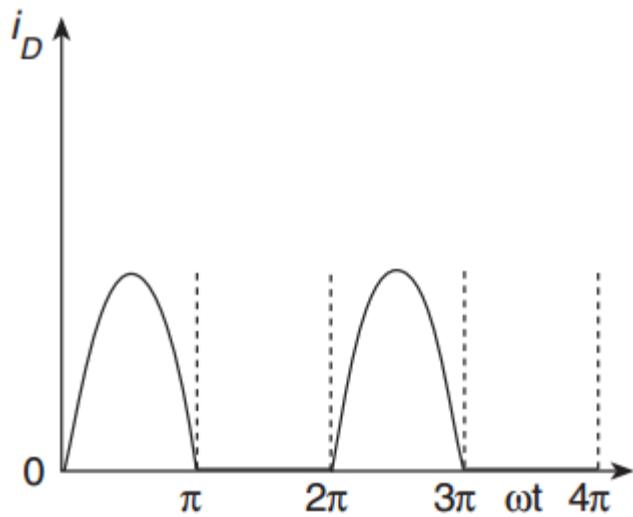


Fig. 9. 7: Waveform of the drain current of class B.

9.3.3. Class-AB power amplifier

This is an intermediate class between class A and class B. The conduction angle is between 180° and 360° which means that the transistor is on more than half of the period and less than the whole period. V_{GS} is slightly higher than $V_{threshold}$ and the transistor is biased at a small drain current.

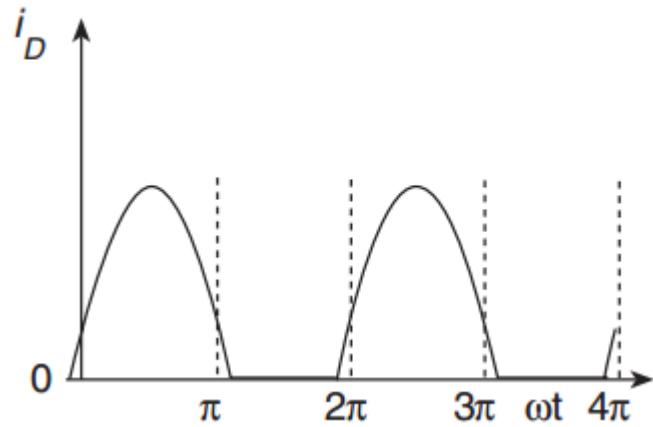


Fig. 9. 8: Waveform of the drain current of class AB

9.3.4. Class-C power amplifier

In class-C power amplifiers, the conduction angle is less than 180° . V_{GS} is set to be less than $V_{threshold}$. Hence, the drain current is set to zero.

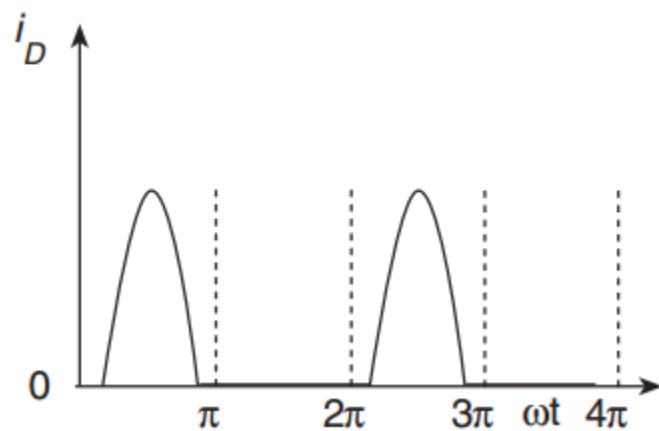


Fig. 9.9: Waveform of the drain current of class C.

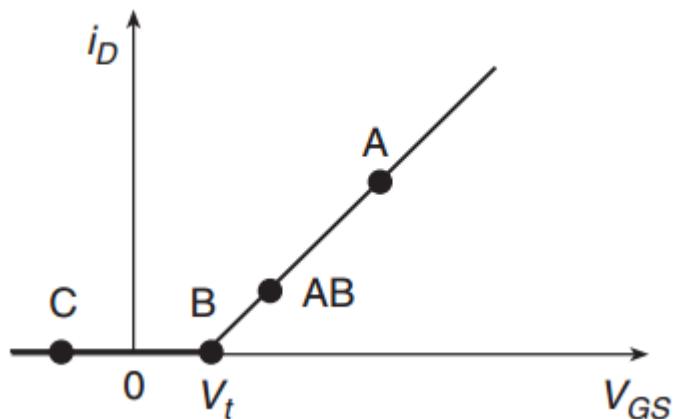


Fig. 9.10: Bias points for classes A, B, AB, C

Switching classes such as class D, E, and F are used to provide very high efficiency (ideally 100%) but their linearity is so poor because transistors here are used as switches.

	Efficiency(ideally)	linearity
Class A	50%	High
Class B	78.5%	low
Class AB	50%: 78.5%	moderate
Switching classes	100%	Very low

Table 9.1: Comparison between ideal efficiencies and linearity of power amplifier classes.

9.4. Common source implementation

For the early implementations of FMCW radar transceivers in 65nm the most common implementation for the power amplifier was using the cascode topology but then in the most recent papers, the design of a PA is done using differential common source stages. In this section, two implementations of the PA are proposed one with CS stages and the other using cascode stages to demonstrate the advantages and disadvantages of each topology. This section includes the common source topology and the next section includes the cascode topology.

9.4.1. required specs

The required spec for the PA from system level requirements was to provide an output power of 10dBm, the rest of the specs were not specified, the rest of the specification was based on the results in the literature using 65nm technology.

Spec	Target specs based on Literature
Gain	>15dB
Saturation power	>13dBm
PAE	>13.5%
P1dB	-

Table 9. 2: target specs for CS implementation

9.4.2. Structure

9.4.2.1. Neutralization capacitors

The two main disadvantages of CS amplifiers at mm-wave operation are its low gain, as at this frequency range the maximum achievable gain of a single transistor is very low which will lead to very low PAE, and that it is hard to maintain its stability. To overcome the disadvantages of CS, the pseudo-differential topology of the PA is chosen to apply a very useful technique which is neutralization capacitors.

The neutralization capacitors are used to compensate for the feedback from the output to the input, this

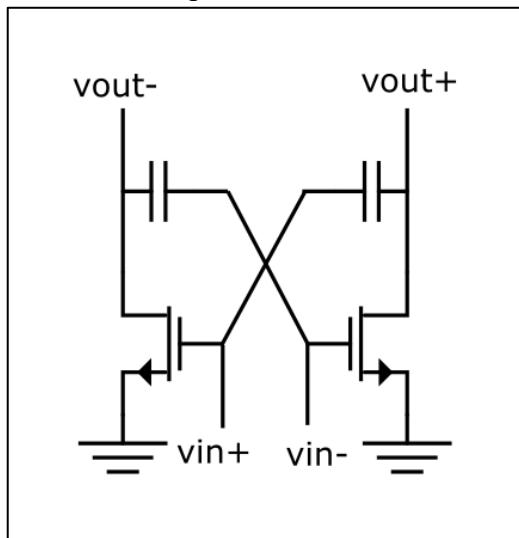


Fig. 9. 11: Neutralization capacitors connection.

feedback results from the parasitic capacitance C_{gd} between the output and input which causes instability, bad reverse isolation, and lower gain (i.e., the Miller effect).

The neutralization is done by providing a compensating current to cancel this feedback and this is done with two additional capacitors cross-connected between the gate and the drain terminals with the value of C_{gd} as shown in Fig. 9.11.

The neutralization gives the CS amplifier a near unilateral behavior which causes unconditional stability, good reverse isolation, and higher gain with no penalty on the output power.

9.4.2.2. Transformer-based matching

At this high-frequency matching with inductors and capacitors was not realizable, as it would require very small inductance values that cannot be manufactured (ex. 20pH), the second choice was TLs matching which had relatively high losses that would degrade the output power and efficiency, so the choice for matching was transformer-based matching which exhibits lower losses compared to transmission lines and provides compact size.

A transformer with self-inductances L_1 and L_2 and a mutual magnetic coupling coefficient k , can be modeled by an equivalent circuit as shown in Fig. 9.12, with an ideal transformer with turns ratio N , a leakage inductor $(1 - k^2)L_1$ and a magnetizing inductor k^2L_1 where $N = k\sqrt{L_1/L_2}$.

For a capacitive input impedance $Z_s = R_s - jX_s$ to a capacitive output impedance $X_L = R_L - jX_L$ the quality factor for both source and load impedances is given by:

$$Q_{s,l} = -Im(Z_{s,l})/Re(Z_{s,l}) \quad 9.4$$

Using this simple equivalent circuit shown in Fig. 9.12 to match Z_L to Z_S it can be proved that the required primary and secondary inductances are given by [6]:

$$L_1 = \frac{R_s}{\omega} \times \frac{\tilde{Z}}{1 - k^2} \quad 9.5$$

$$L_2 = \frac{\tilde{Z}R_L}{\alpha\omega} \times \frac{1 + Q_l^2}{1 + (Q_s - \tilde{Z})^2} \quad 9.6$$

Where ω is the operating frequency, and $\alpha = (1 - k^2)/k^2$ and \tilde{Z} has two values and is given by:

$$\tilde{Z} = \frac{Q_s(2\alpha + 1) + Q_l}{2(\alpha + 1)} \pm \frac{\sqrt{(Q_s(2\alpha + 1) + Q_l)^2 - 4\alpha(\alpha + 1)(Q_s^2 + 1)}}{2(\alpha + 1)} \quad 9.7$$

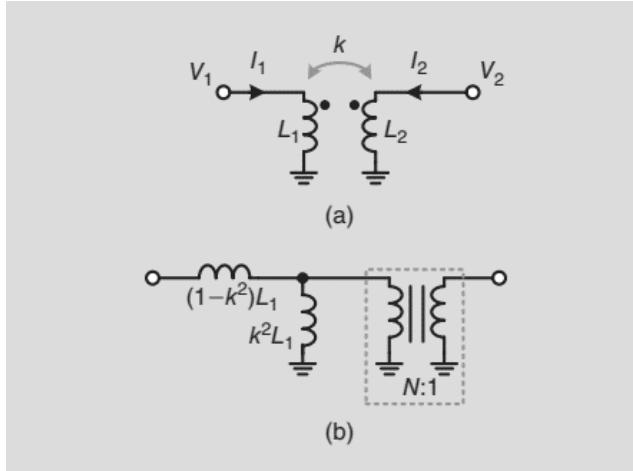


Fig. 9.12: The simplified model of transformers.

9.4.2.3. Cascading stages

Using multiple stages was necessary as at this frequency range using 65nm technology a single stage will either provide high gain or high output power.

A PA stage provided high power but very low gain hence it had very low PAE so it needed two driving stages with reasonable output power that doesn't degrade the total output power and increase the gain with a low penalty in the total power consumption to get good PAE. The chosen architecture for the PA is three stages of differential CS as shown in Fig. 9.13.

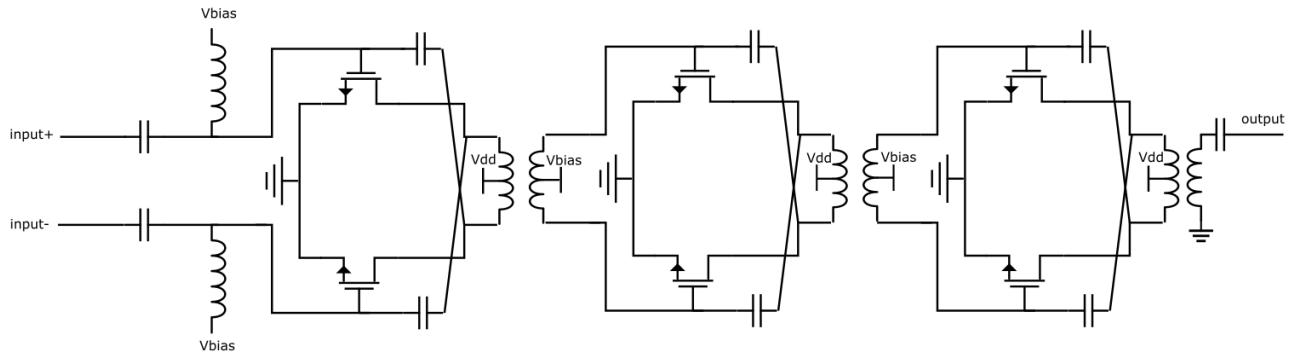


Fig. 9.13: Schematic of the proposed PA.

9.4.3. Design procedure

9.4.3.1. Current density selection

As the $OP1dB$ is maximized at the peak f_T current density J_{pft} , so it is the optimum current density for the PA stage but to avoid the thermal breakdown for the transistor the chosen current density is chosen less than J_{pft} with some margin.

9.4.3.2. PA stage transistors Sizing

The transistor sizing for the PA stage is chosen to set the $OP1dB$ to achieve a higher value than the required P1dB value to account for the degradation in the total P1dB after cascading the PA with the driver stages according to equation 9.7 and the losses of the balun. And at the same time have reasonable gain value to relax the output power requirement on the drivers.

$$\frac{1}{OP1dB_{total}} = \frac{1}{OP1dB_{PA}} + \frac{1}{OP1dB_{Driver1} \times Gain_{PA}} + \frac{1}{OP1dB_{Driver2} \times Gain_{PA} \times Gain_{Driver1}} \quad 9.8$$

9.4.3.3. Output matching

As explained in section 9.1 the matching at the output for power amplifiers is done to achieve the highest output power which may lead to a mismatch at the output causing low gain and bad values for S_{22} , for this design the optimum load impedance Z_{sopt} was very close to the conjugate match impedance so matching for power will not cause high degradation in S_{22} so the gain is also reasonable for this stage.

9.4.3.4. Driver stages transistor Sizing

The sizing for the drivers was done to have relatively high gain with reasonable output power values such that it does not degrade the total output power for the cascaded stages where the first driver for the PA has higher output power and lower gain compared to the second driver that drives the PA and the first driver. each driver has a higher PAE compared to the PA stage and the second driver has a higher PAE compared to the first one such that the total PAE increases with the cascaded stages, as the cascaded PAE for 2 stages is given in equation 9.9, it is notable that if the driver stage has higher PAE compared to the PA stage the total PAE will be greater than that of the PA.

$$PAE_{Two\ stages} = PAE_{PA} \left[1 - \frac{PAE_{PA} - PAE_{driver}}{PAE_2 + PAE_{driver} \times \frac{G_{driver}}{G_{driver} - 1} \times (G_{PA} - 1)} \right] \quad 9.9$$

9.4.3.5. Interstage matching

Inter-stage matching is done to maximize the gain of each driver stage to get a high value of PAE for each driver stage, transformer-based matching is used as explained in section 9.4.2.2, and the flatness of the gain in the required band is considered in the design of these matching networks by centering the matching at different frequency value for each of the matching network. The input match is done using an LC matching network as the required L value was reasonable to be manufactured, the input matching is also used in enhancing the gain flatness.

9.4.4. Layout

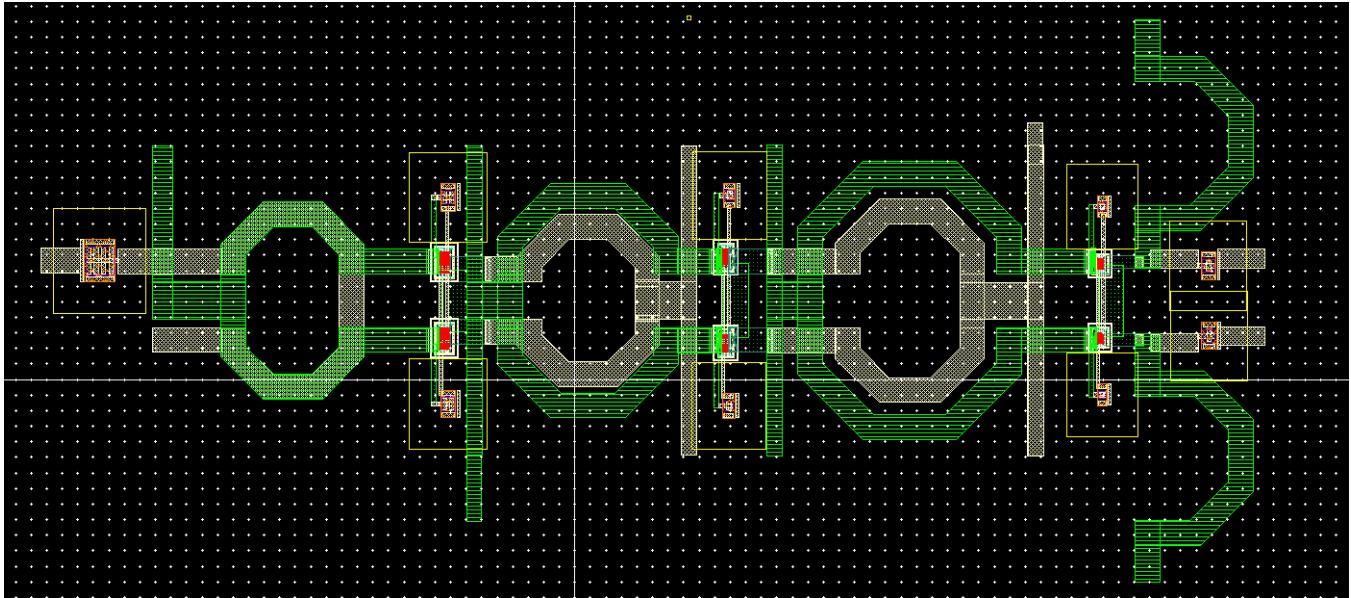


Fig. 9. 14: full layout of the PA

9.4.5. Post-layout simulation results

In this section, the results of PA are shown, after EM simulation and Parasitic extractions.

Fig. 9. 15 shows the gain of the PA (S_{21}), the achieved gain is larger than 21dB for the whole range with gain flatness of 0.5 dB.

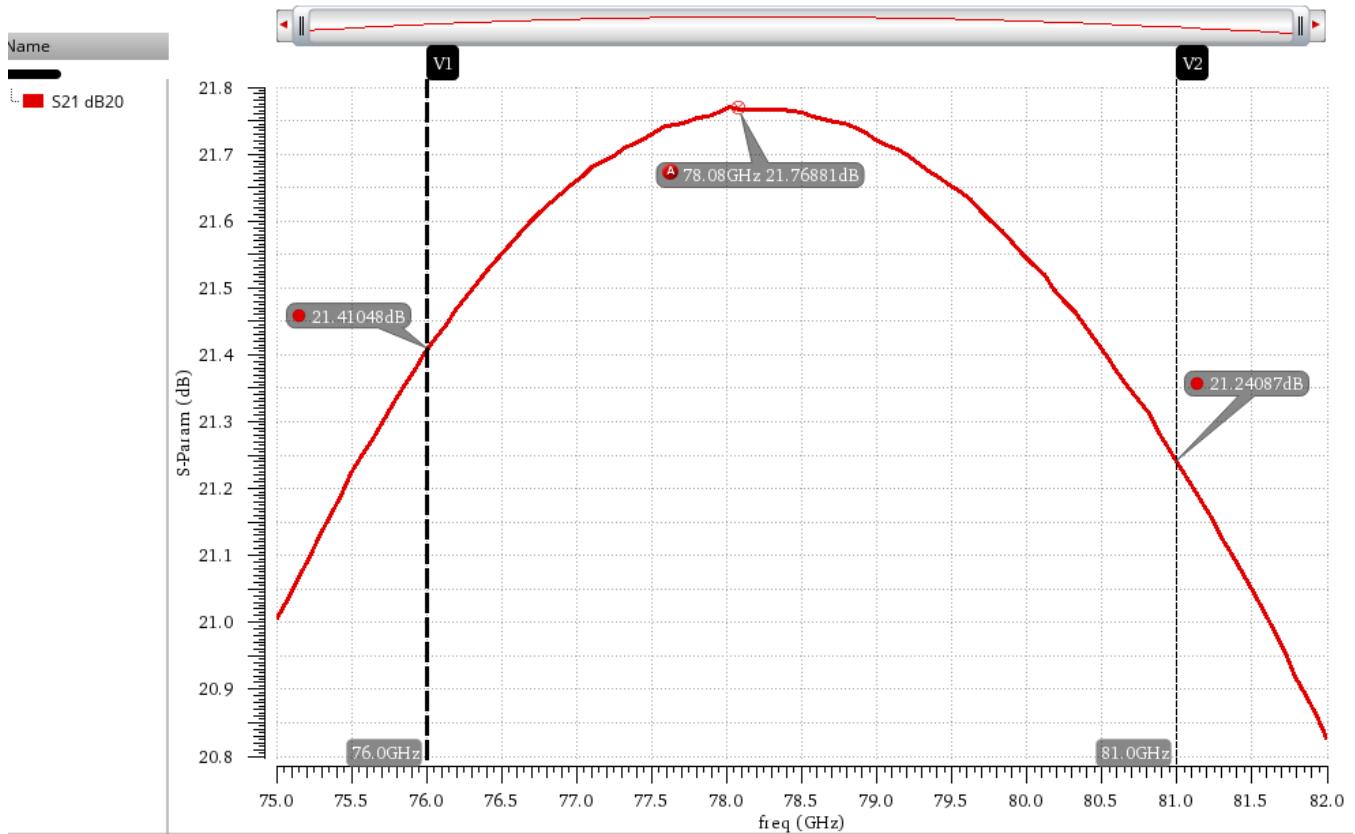


Fig. 9. 15: S_{21} vs frequency.

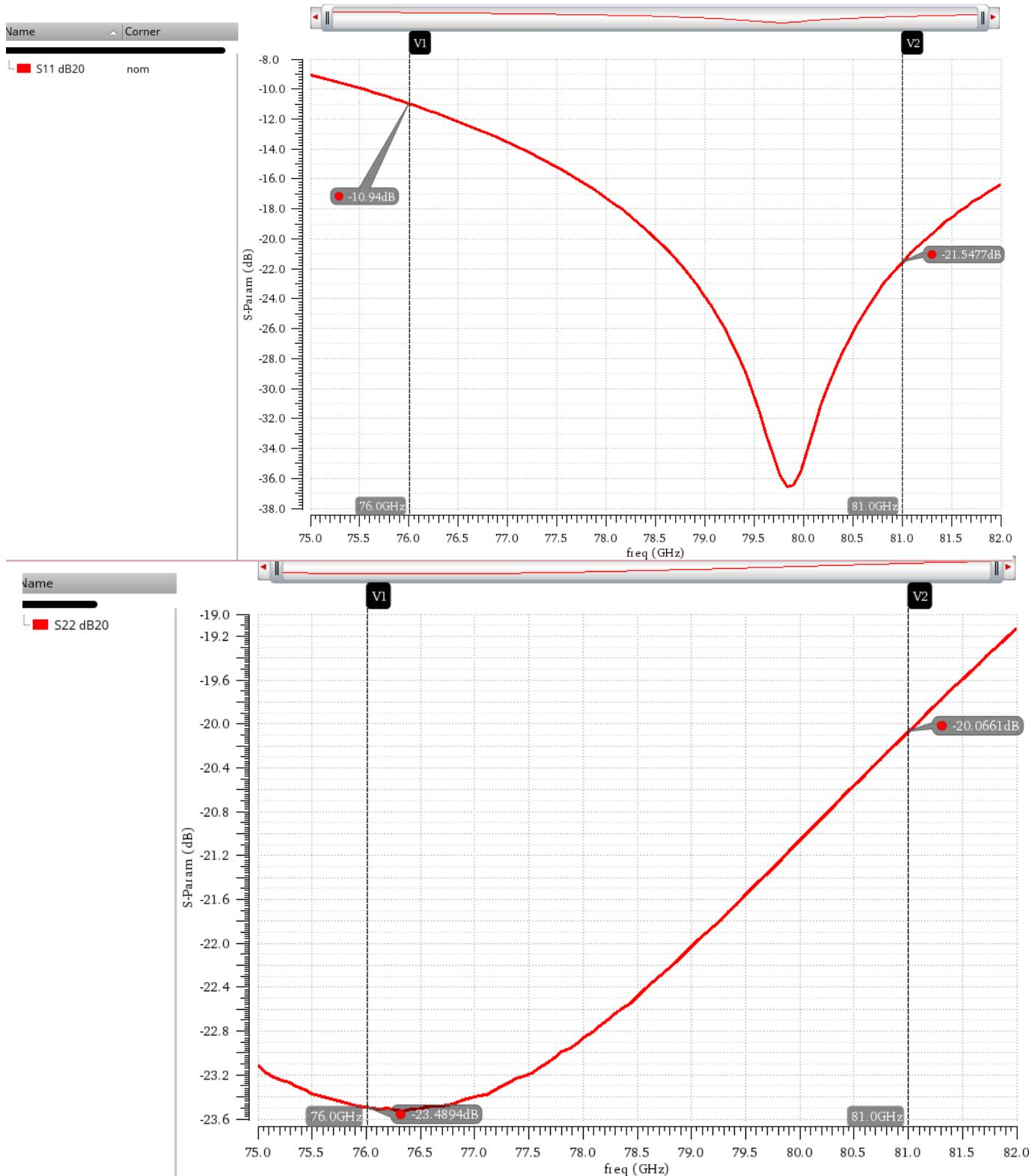


Fig. 9. 16: S_{22} vs frequency.

Fig. 9. 17 and Fig. 9. 16 show the S_{11} and S_{22} for the PA, where both are lower than -10dB for the whole range, note S_{11} is shifted toward high frequency to improve gain flatness and S_{22} is very good for the

whole range which is because the optimum load for power matching was very close to the conjugate match impedance, also as a part of improving the power-added efficiency S_{22} was improved while maintaining acceptable output power.

Fig. 9. 18 shows the NF of the PA which is less than 5.8dB, while Fig. 9. 20 shows the k-factor value for the PA which is larger than 1, and Fig. 9. 19 shows the B1-factor which is larger than for the whole range.

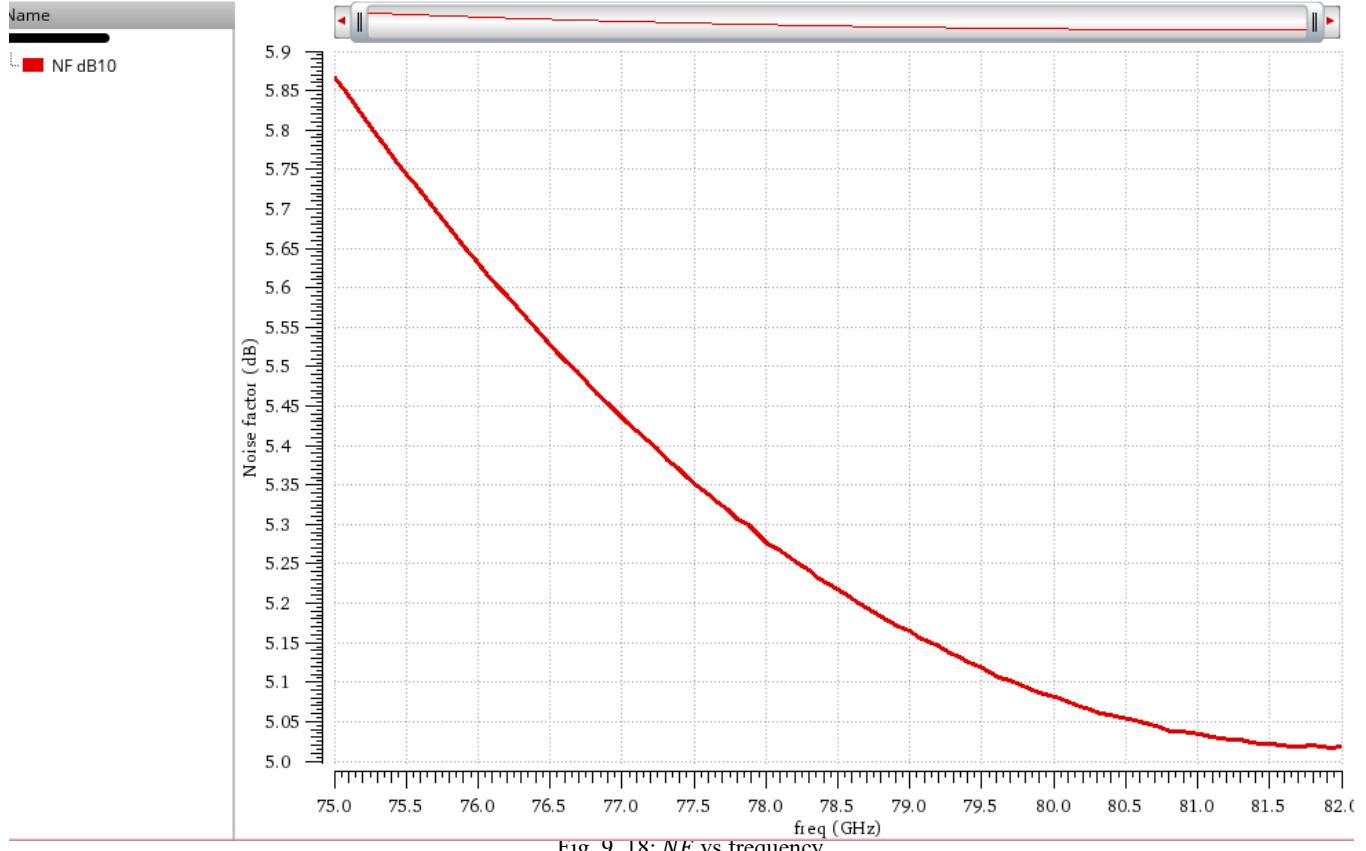


Fig. 9. 18: NF vs frequency.

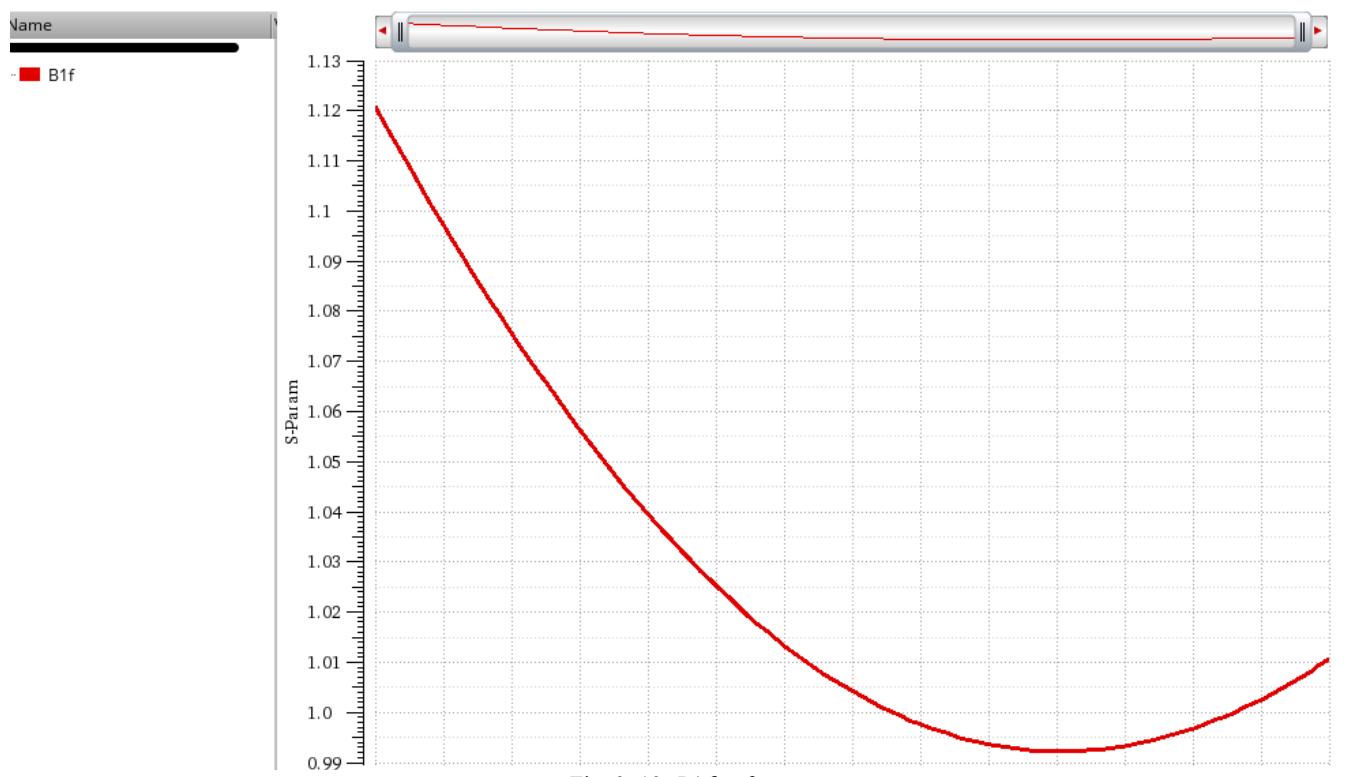
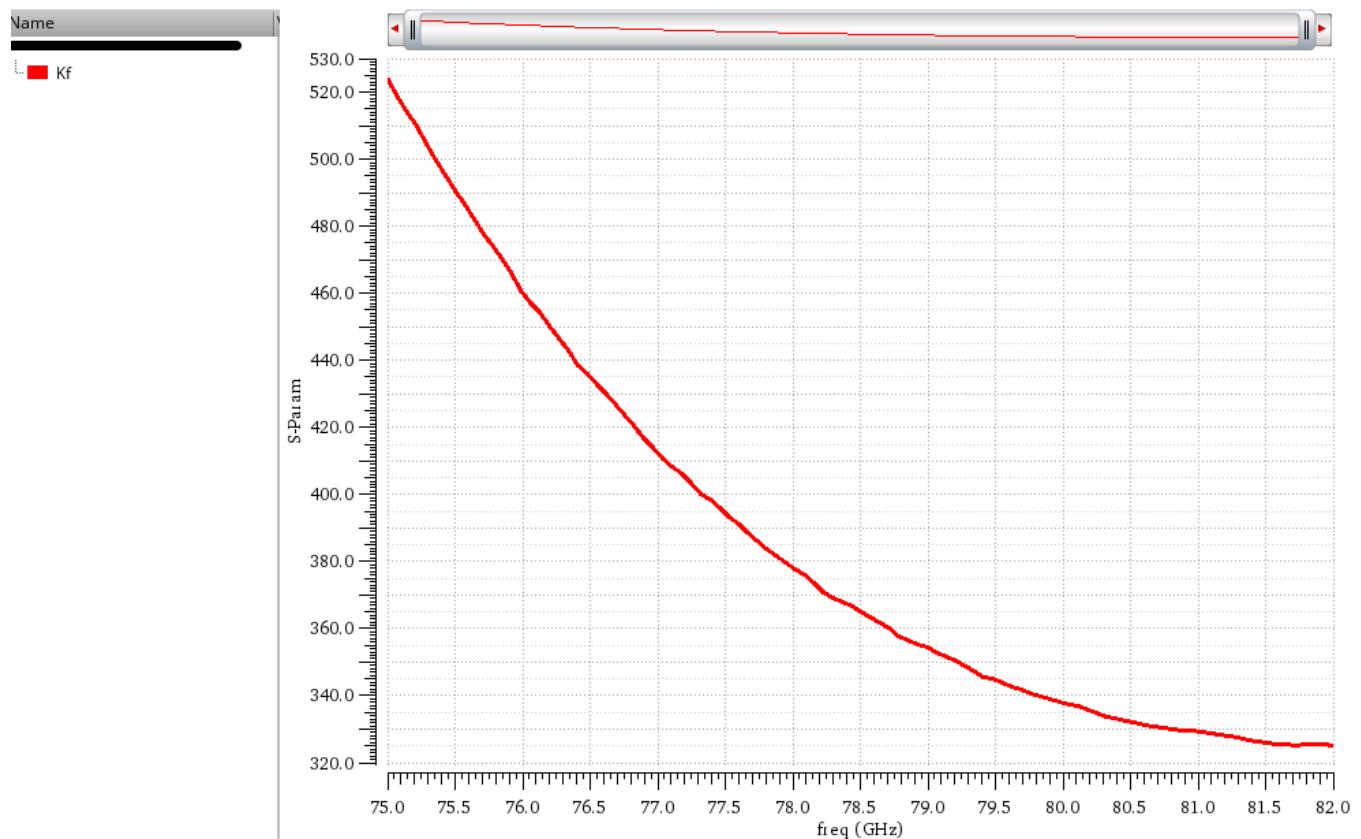


Fig. 9. 19: $B1f$ vs frequency.

Fig. 9.22 shows the output power of the PA versus input power for different frequency values (76G, 78.5G, and 81.5GHz) which shows the saturation power is almost constant across frequency, while Fig. 9.21 shows the PAE of the PA versus input power for different frequency values (76G, 78.5G, and 81.5GHz) which ranges from 15.28% at 76GHZ to 14.367% at 81GHz.

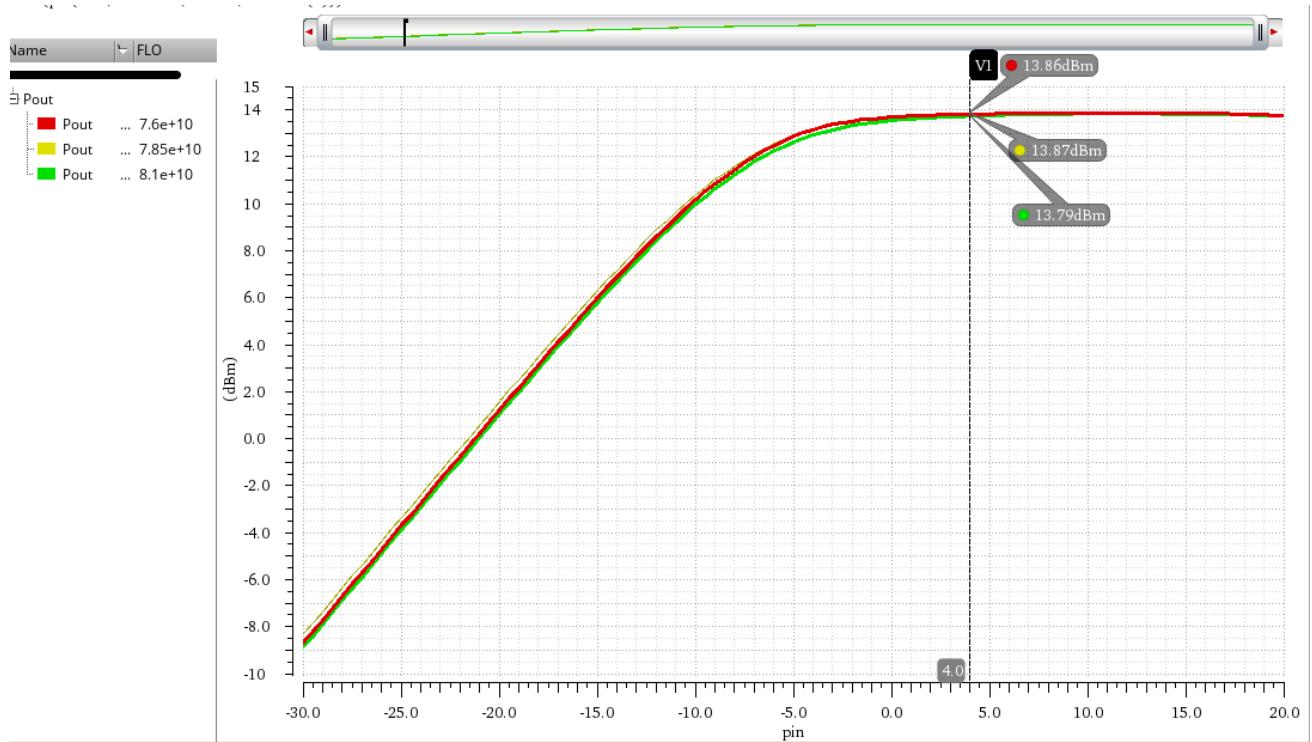


Fig. 9. 22: P_{out} vs P_{in} at 76G,78.5G and 81GHZ.

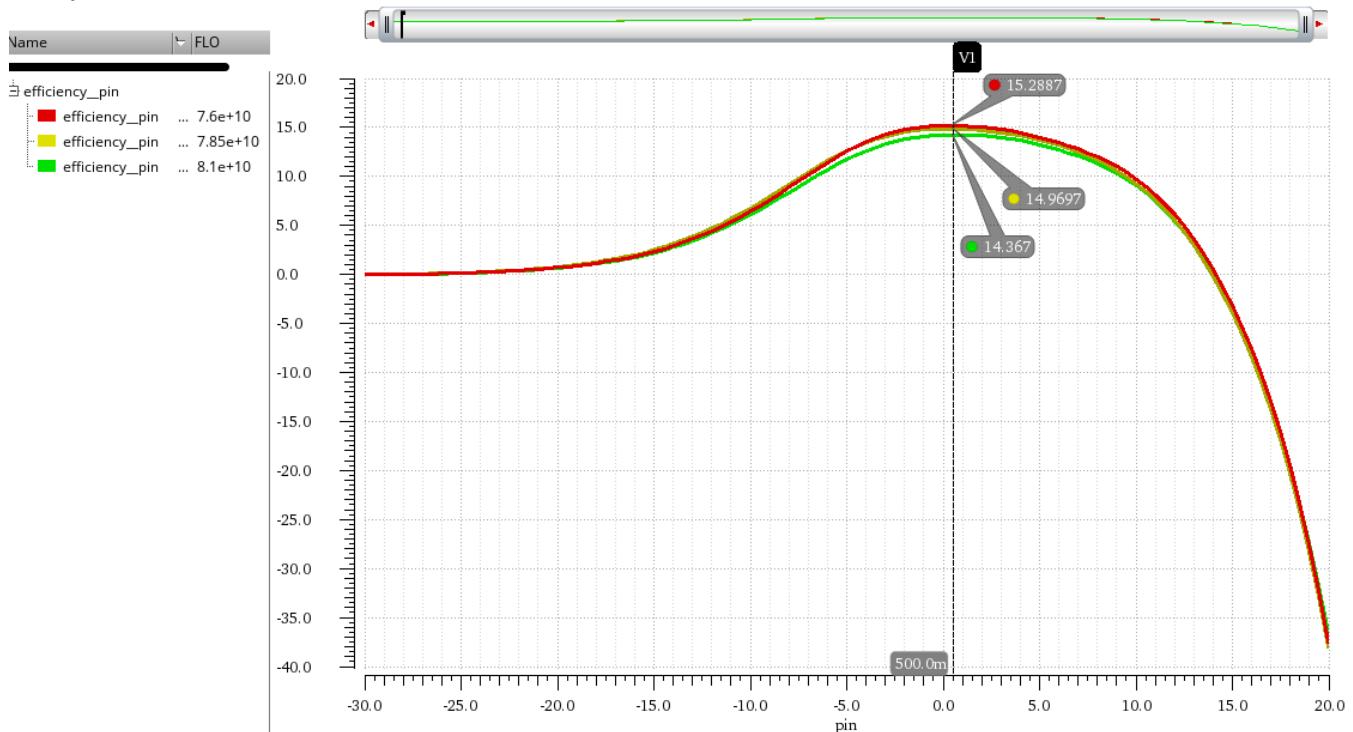


Fig. 9. 21: PAE vs P_{in} at 76G,78.5G and 81GHZ.

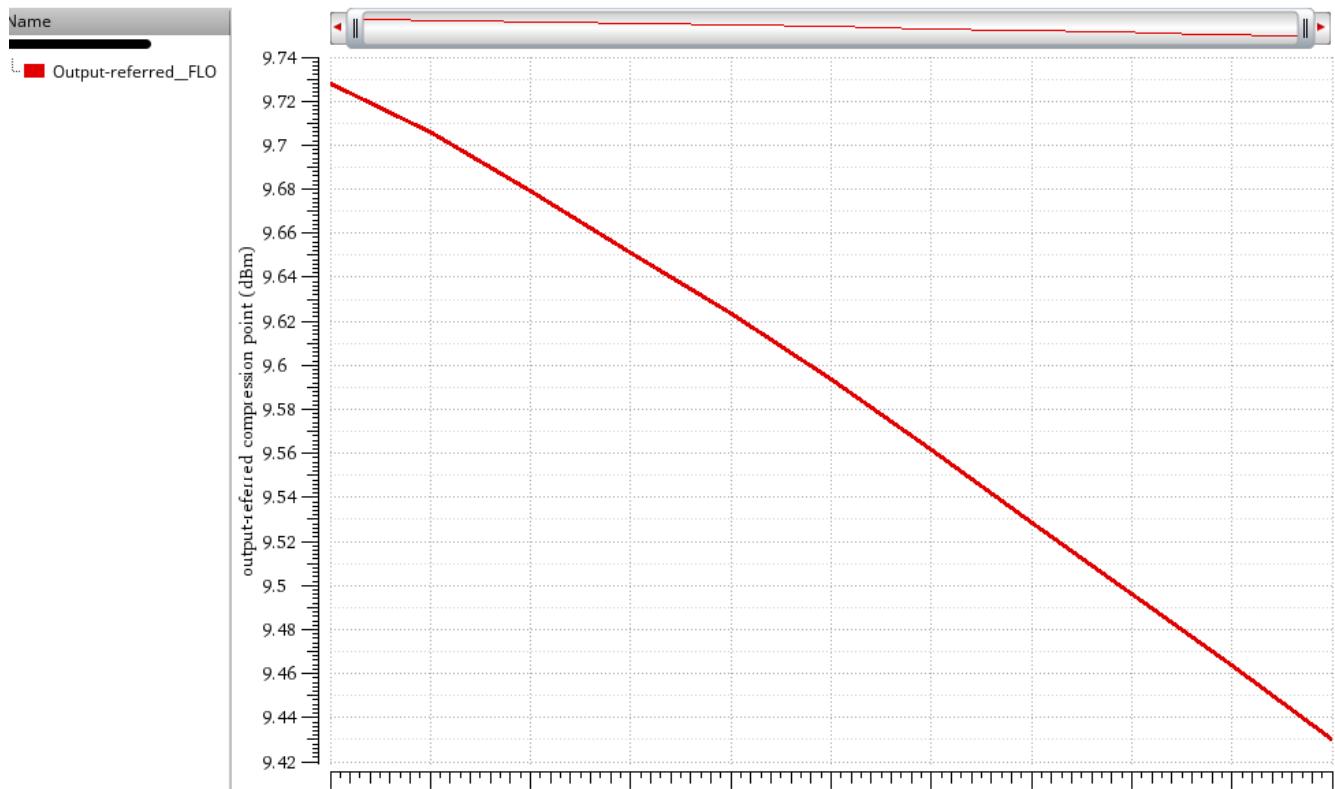


Fig. 9. 24: $OP1dB$ vs frequency.

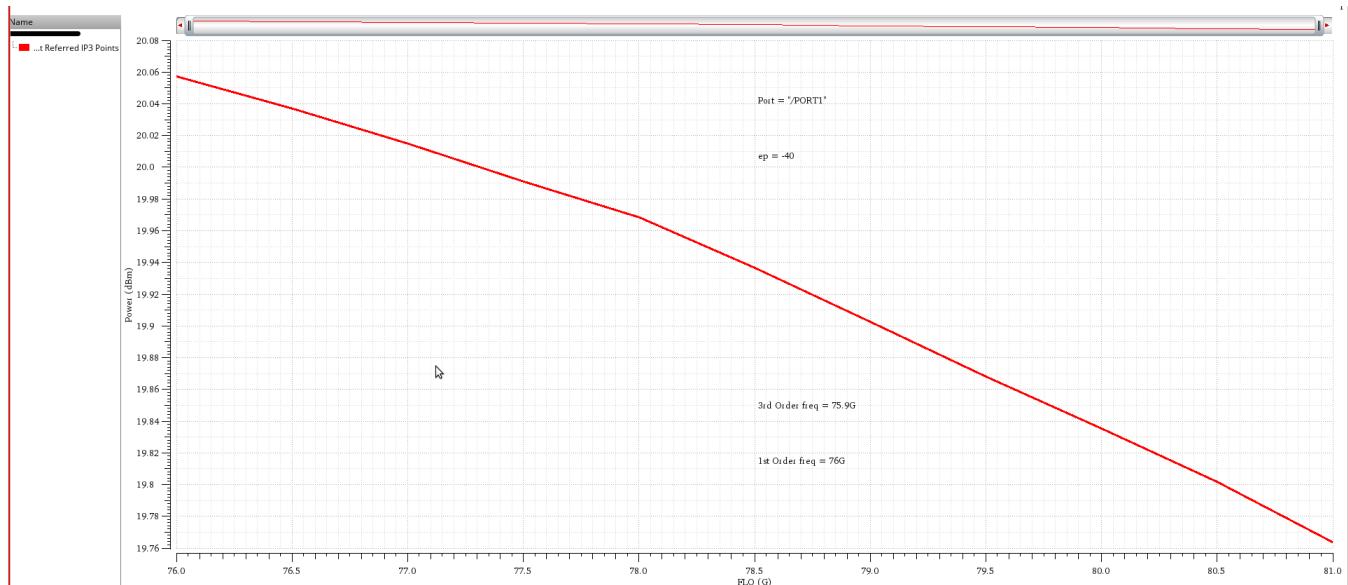


Fig. 9. 23: $OIP3$ vs frequency.

Fig. 9.23 shows the $OP1dB$ value for the PA which is ranges from 9.7dBm at 76GHz to 9.4dBm at 81GHz, while Fig. 9.24 the $OIP3$ value which ranges from 20.06dBm to 19.76dBm at 76GHz and 81GHz, respectively.

Spec	Target	Achieved
Gain (S21)	$> 20dB$	$21dB$
P_{sat}	$> 13dBm$	$13.7dB$
PAE	$> 13.5\%$	15%
$OP1dB$	-	$9.4dBm$
S11	$< -10dB$	$< -10.9dB$
S22	$< -10dB$	$< -19dB$
S12	-	$< -78dB$
NF	-	$< 5.9dB$
Power consumption	-	$136mW$

Table 9.2: required vs achieved specs of the PA.

9.4.6. Corners

The PA is tested across PVT variations where 2 corners are applied in addition to the nominal case, where the first corner is the slow-hot-low supply where the slow model for the transistors is used along with a supply of value equals 90% of the nominal case and temperature equals 85 degrees, while the second corner is fast-cold high supply where the fast model for the transistors is used along with a supply of value equals 110% of the nominal case and temperature equals -45 degrees.

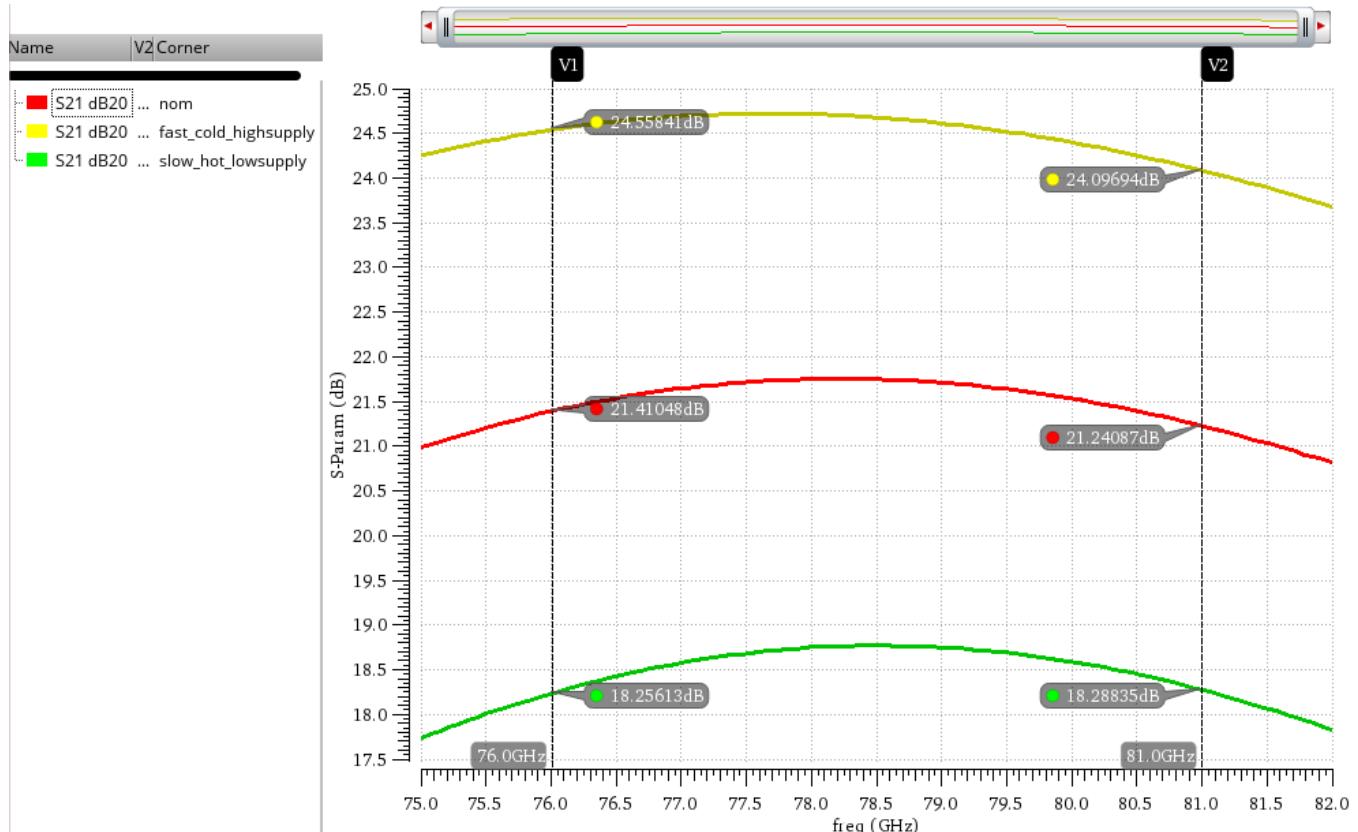


Fig. 9. 25: S_{21} vs frequency across different corners.

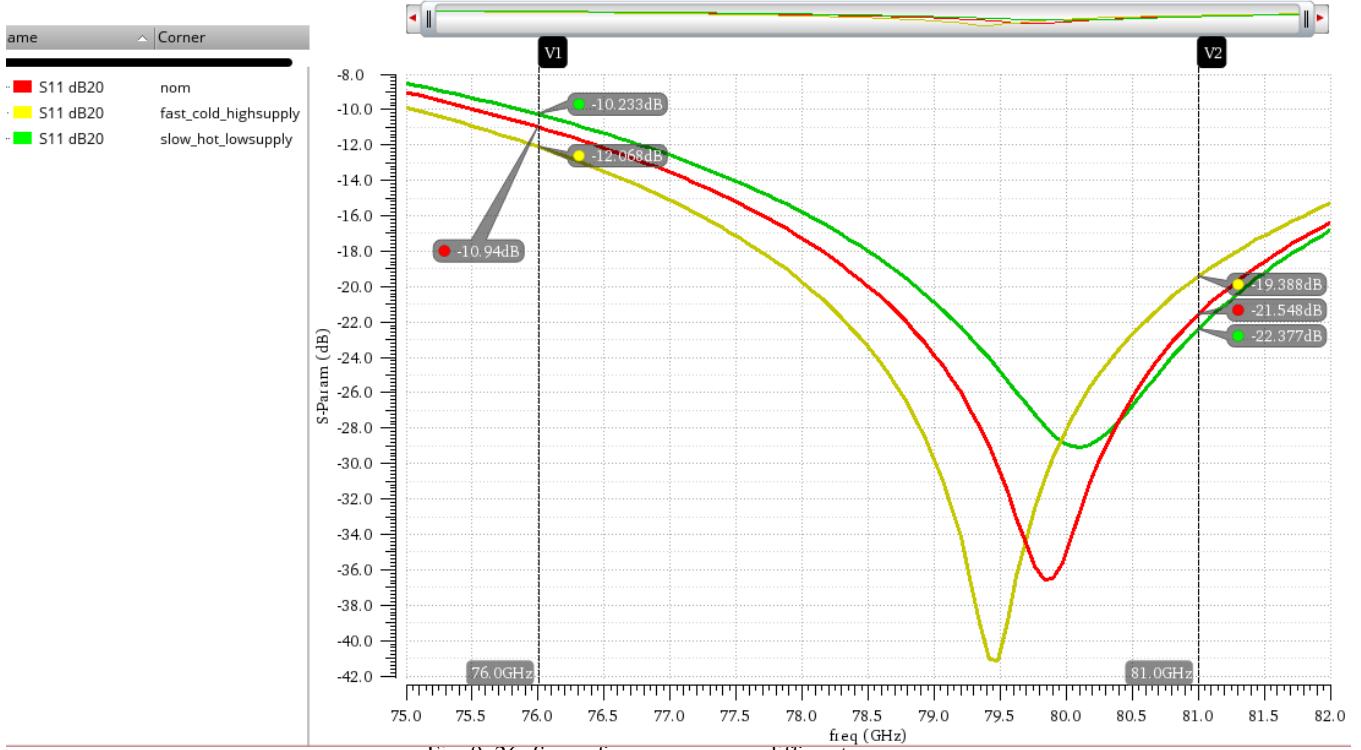


Fig. 9. 26: S_{11} vs frequency across different corners.

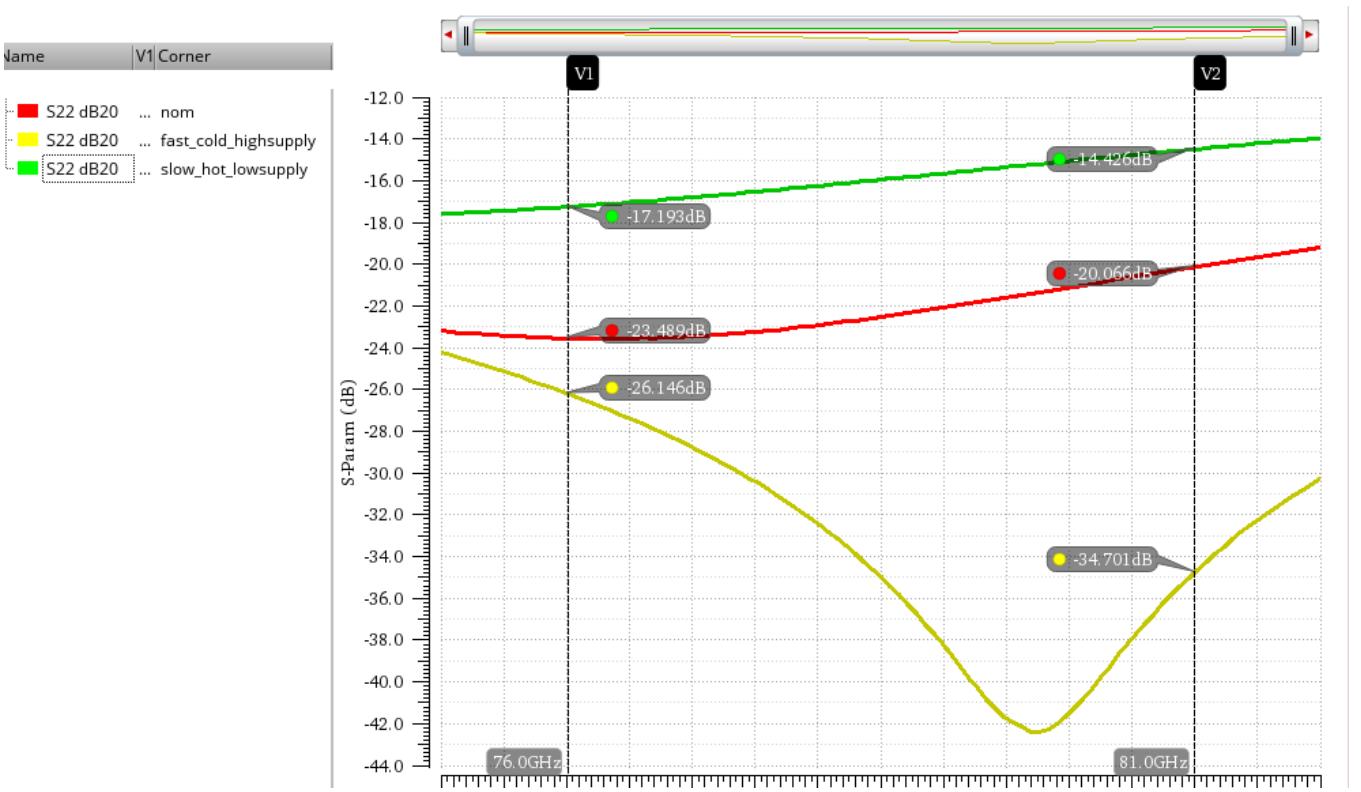


Fig. 9. 27: S_{22} vs frequency across different corners.

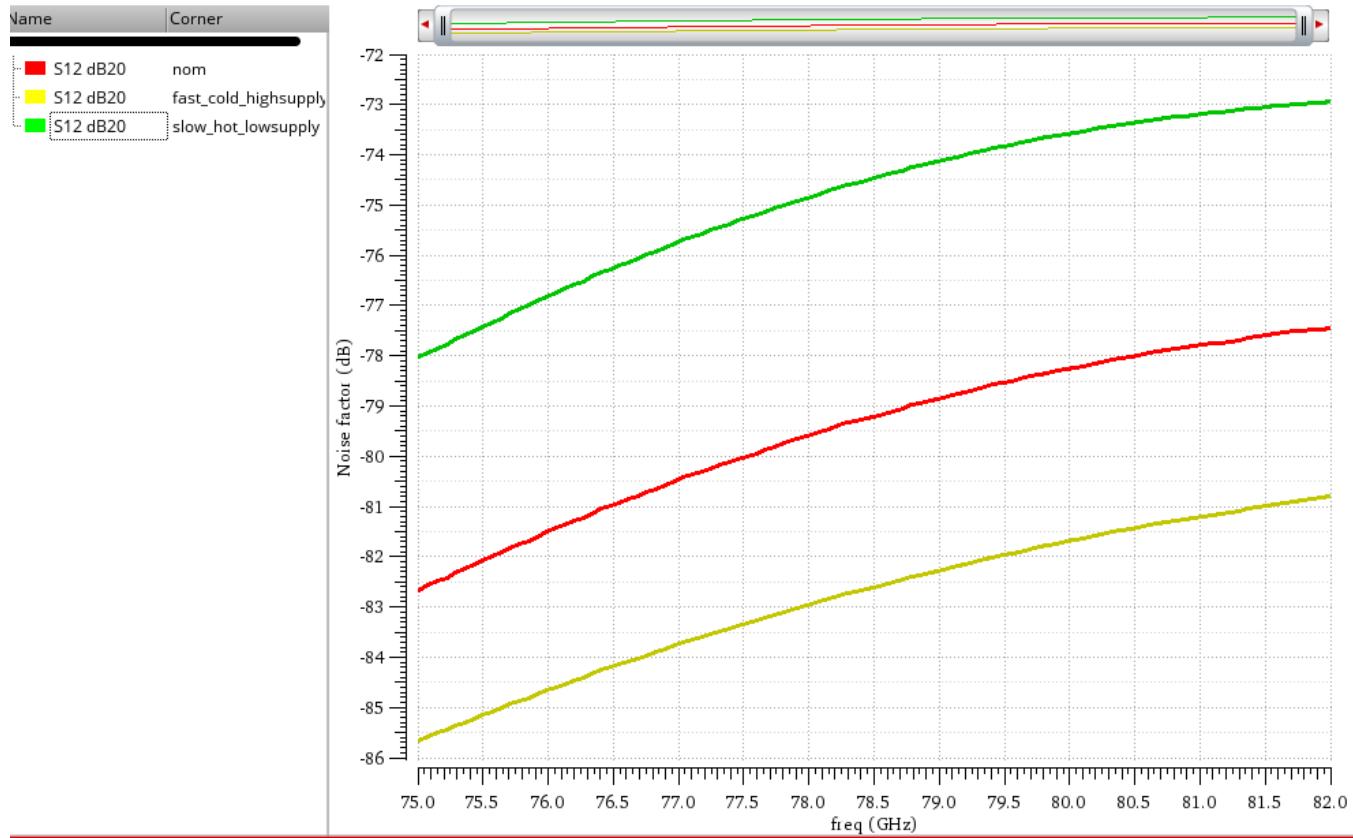
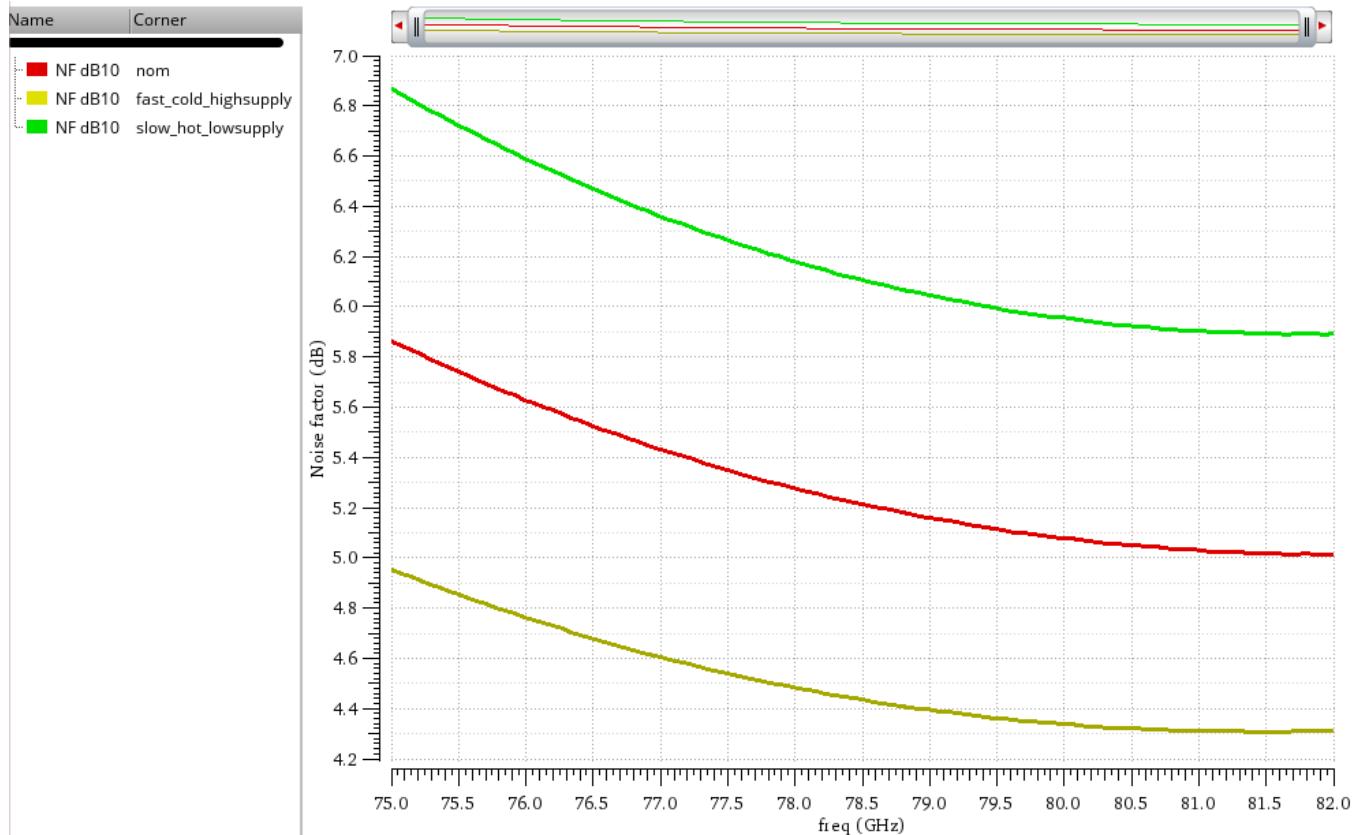
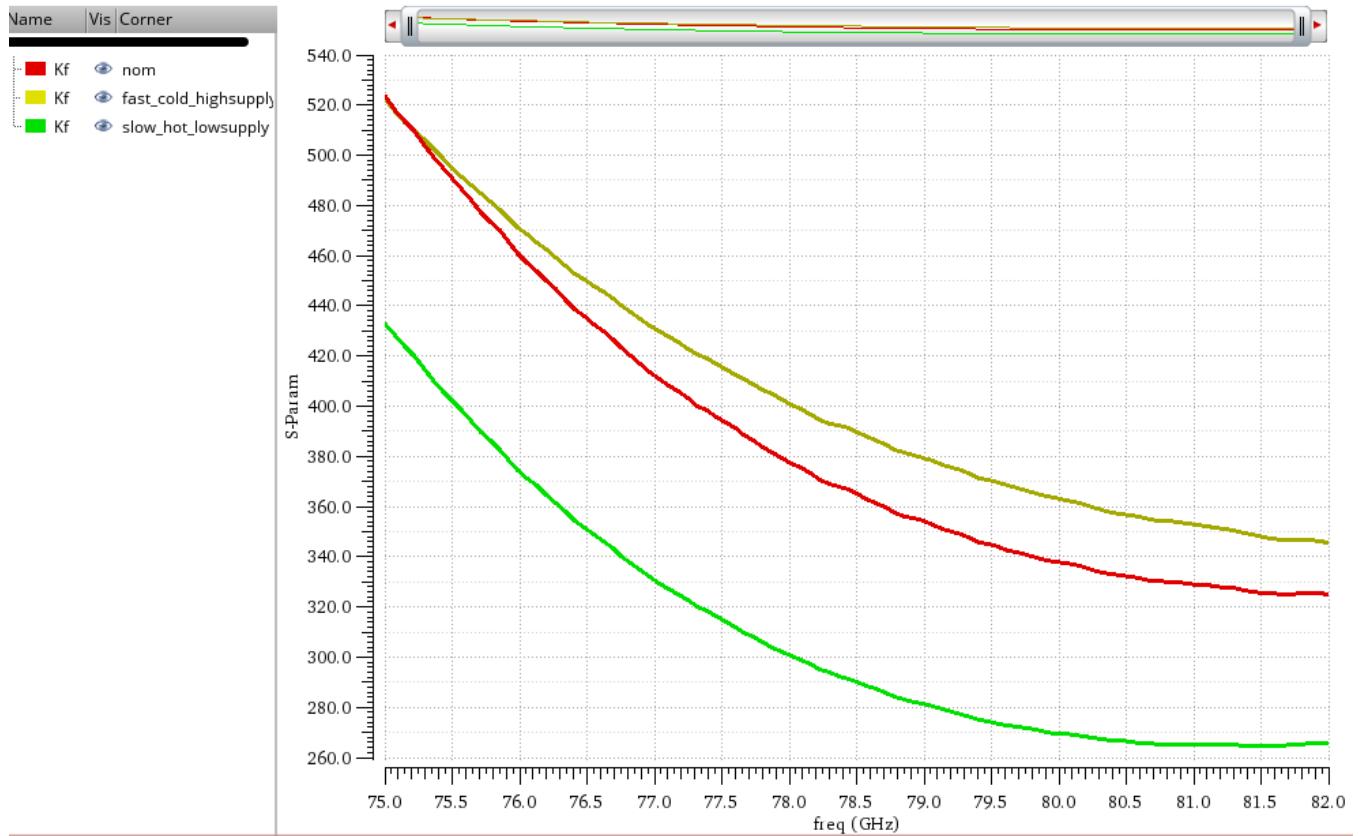


Fig. 9. 28: S_{12} vs frequency across different corners.





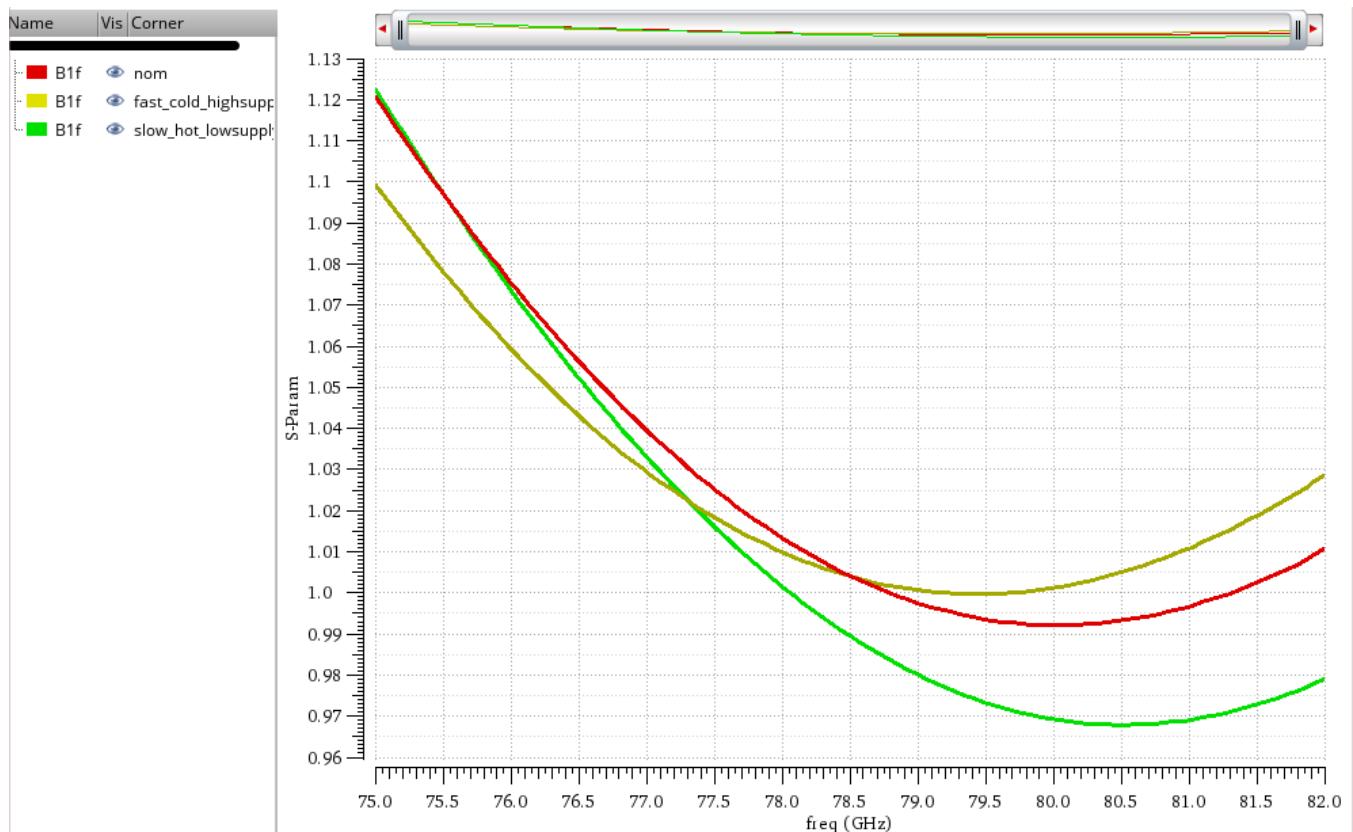


Fig. 9. 31: $B1f$ vs frequency across different corners.

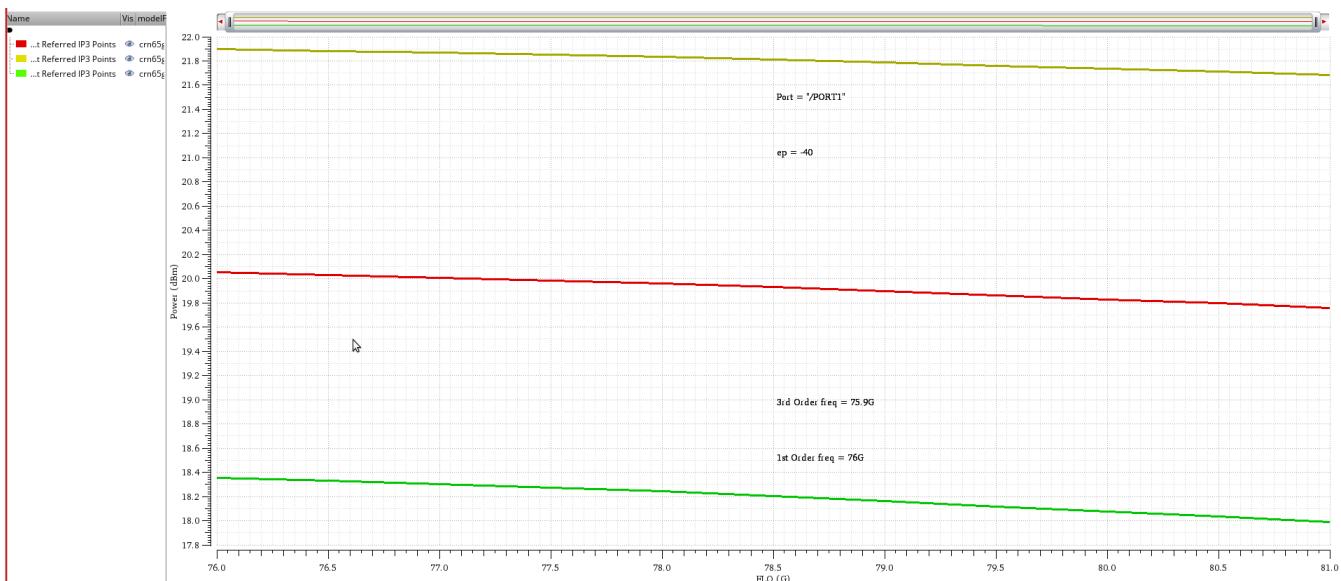


Fig. 9.32 OIP3 vs frequency across different corners.

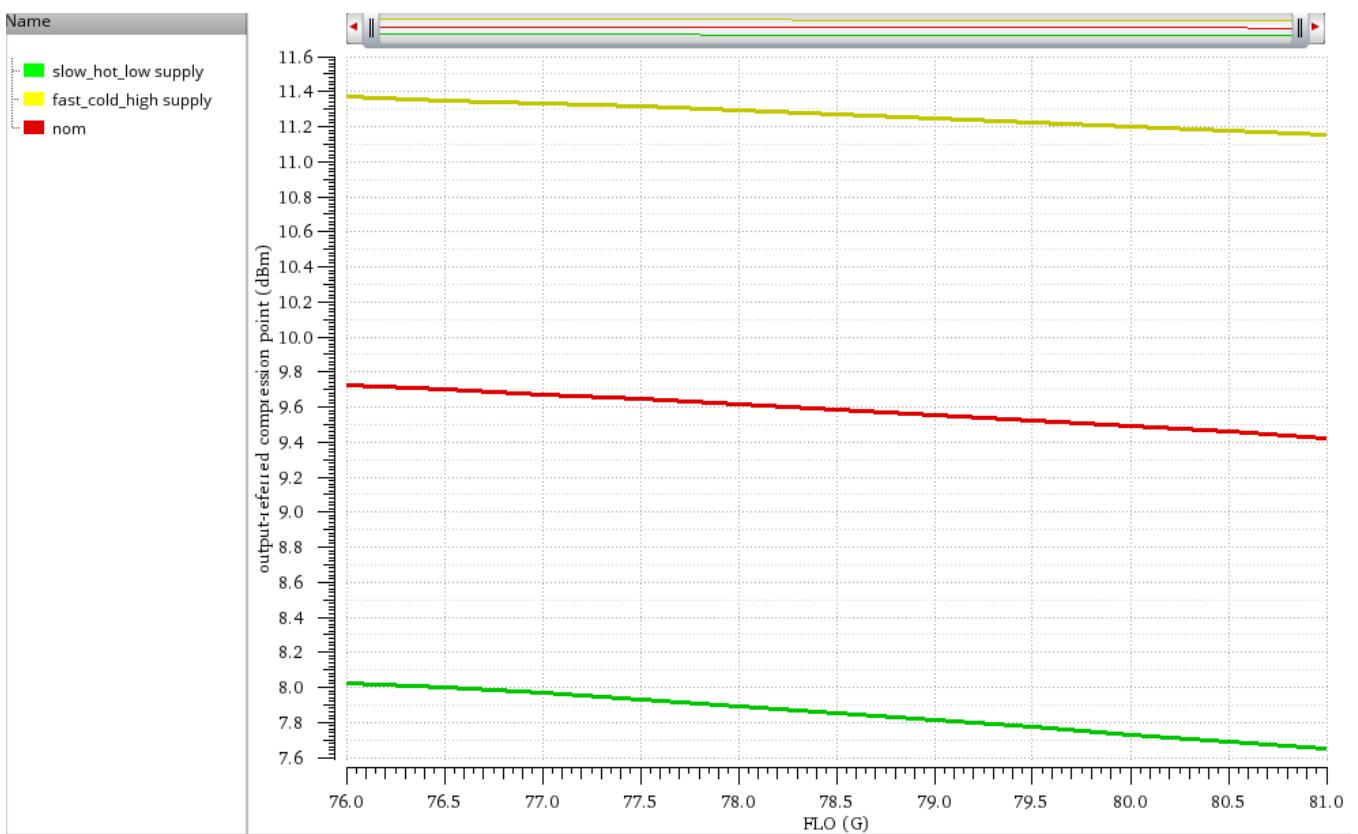


Fig. 9.33: OP1dB vs frequency across different corners.

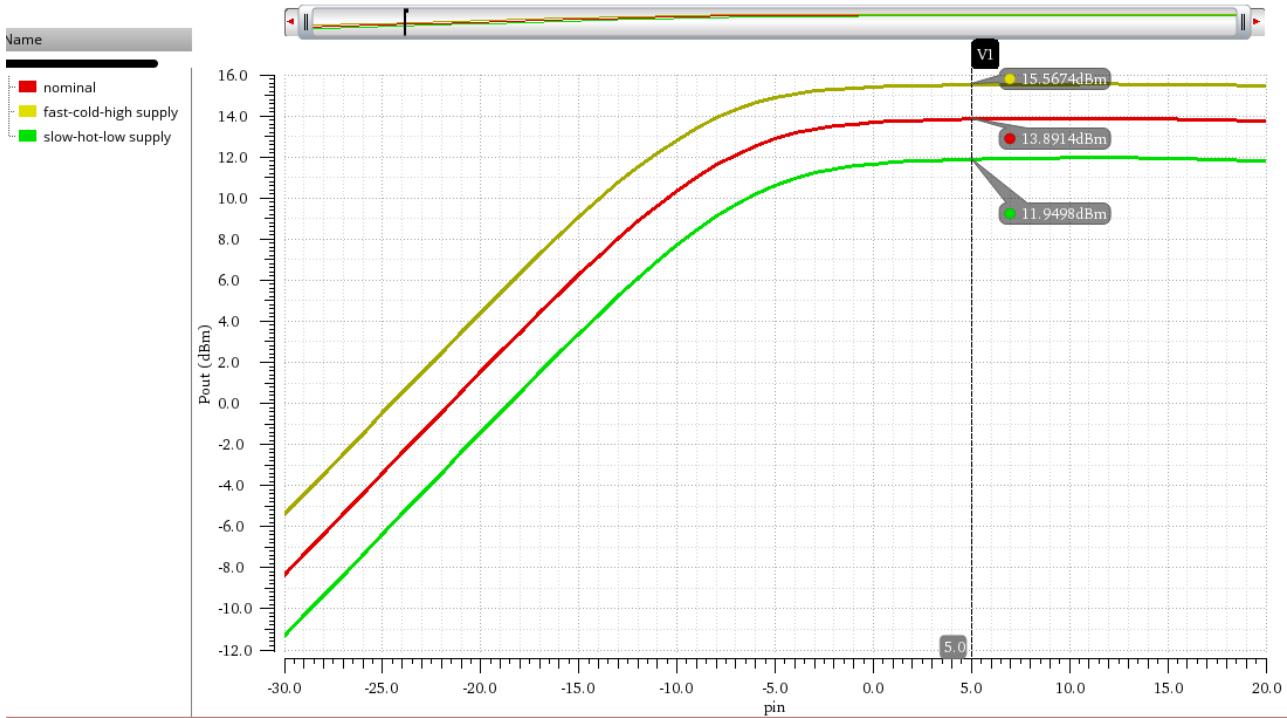


Fig. 9. 34: P_{out} vs P_{in} across corners at 78.5GHz.

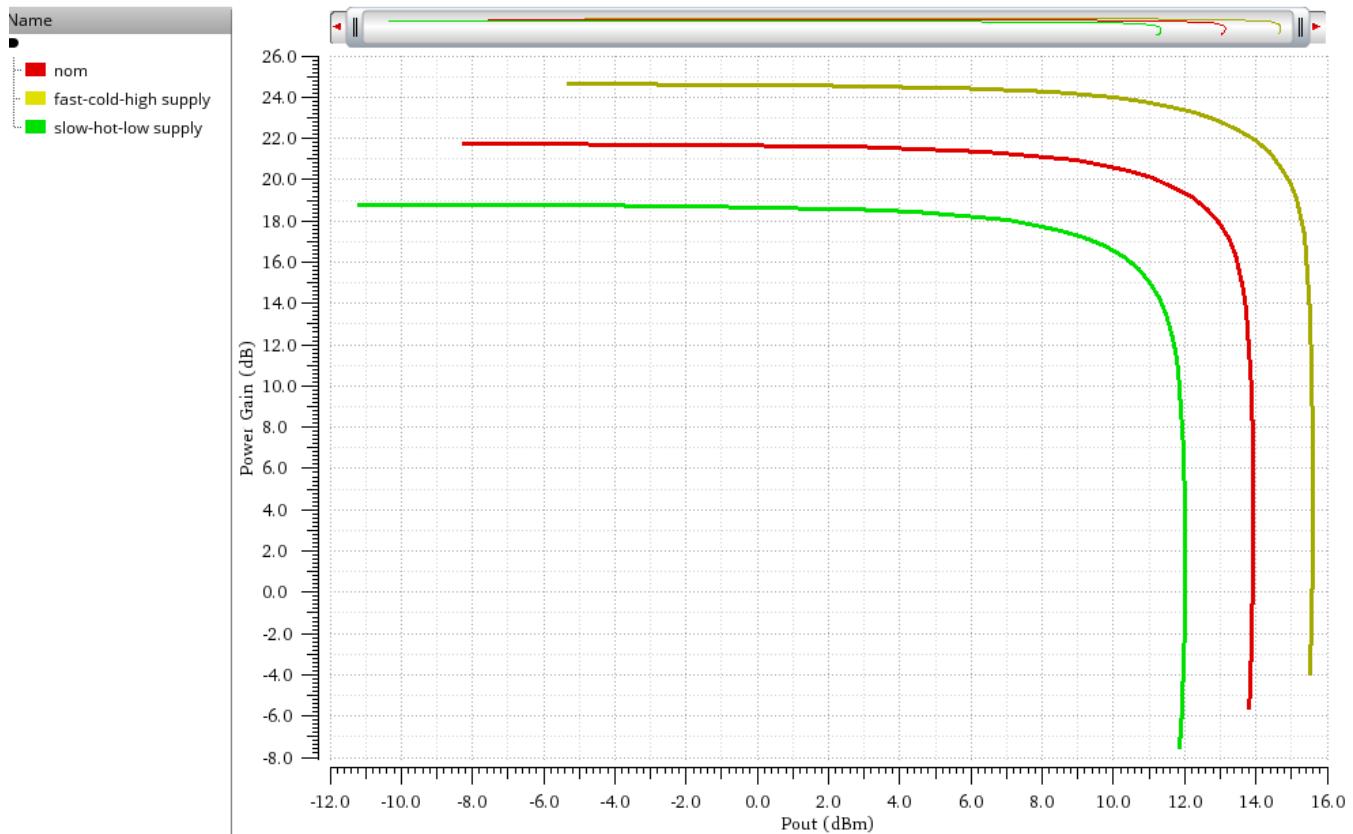


Fig. 9. 35: Power gain vs P_{out} across corners at 78.5GHz.

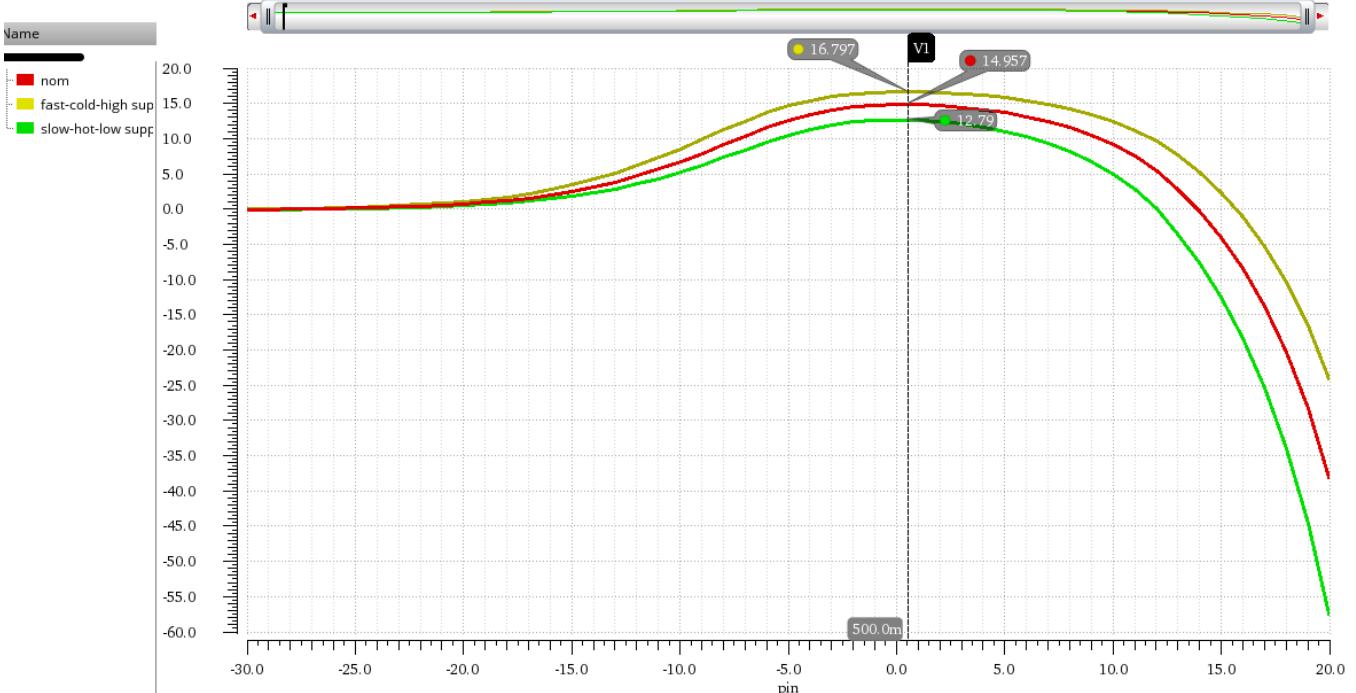


Fig. 9. 36: PAE vs P_{in} across corners at 78.5GHz.

9.4.7. Literature

Table 9.3 shows a comparison between the proposed design and the state of art.

	[3]	[4]	[5]	This work
Process	65-nm CMOS	65-nm CMOS	65-nm CMOS	65-nm CMOS
Frequency	76-81GHz	76-79GHz	77G	76-81GHz
Topology	Differential CS	Differential CS	Differential CS	Differential CS
Number of stages	3	3	2	3
Gain	-	-	8.5dB	21dB
P_{out}	13.4dBm	13.2dBm	9.6dBm	13.7dBm
OP1dB	-	-	8dBm	9.4dBm
Max PAE	13.5%	13.4%	6.45%	15%
Power Cons.	150mW	174.7mW	112mW	136mW

Table 9. 3: PA comparison with related design.

9.5. Cascode implementation

9.5.1. Architecture

The first step in design is to select the architecture. There are many points to be taken into our consideration during this step. These points are minimizing area, and power consumption, and trying to reduce trade-offs between two required specifications. This step includes two steps: -

- Structure (differential or single-ended, single-stage or multi-stage).
- Topology (common source or cascode).

9.5.2. Structure

A two-stage design was selected because the required gain was about 15dB, so it is not possible to get such gain from one stage. The last stage which is also called the output stage is to get the required P1dB because the linearity is dominated by the last block in a system.

$$\frac{1}{OP_{1dB_{Cascade}}} = \frac{1}{OP_{1dB2}} + \frac{1}{OP_{1dB1} * G_2} \quad 9.10$$

The rest of the gain is intended to be got by the first stage and it also helps to increase efficiency.

A differential stage was also selected because it gives us some advantages such as it is more immune to noise because it gets rid of common-mode noise. Moreover, it gives better linearity because even-order harmonics are rejected. If:

$$y_1(t) = \alpha_0 + \alpha_1 x_1(t) + \alpha_2 x_1^2(t) + \alpha_3 x_1^3(t) + \dots \quad 9.11$$

$$y_2(t) = \alpha_0 + \alpha_1 x_2(t) + \alpha_2 x_2^2(t) + \alpha_3 x_2^3(t) + \dots \quad 9.12$$

$$x_1(t) = -x_2(t) \quad 9.13$$

Then,

$$y_1(t) - y_2(t) = \alpha_1(x_1(t) + x_2(t)) + \alpha_3(x_1^3(t) + x_2^3(t)) + \dots \quad 9.14$$

This property is called odd-symmetry and it helps to increase linearity.

Differential stages increase P1dB by 3 dB compared to single-ended stages. This means that we could get higher P1dB with the same transistor sizes.

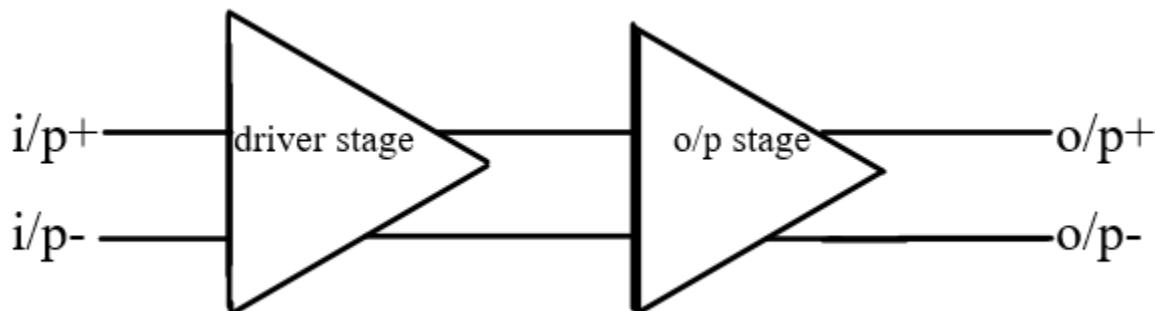


Fig. 9. 37: two-stage power amplifier structure)

9.5.3. Topology

A cascode topology was selected because it provides a high gain. It also experiences more stability, and its reverse isolation is high. We also double the supply voltage, and this would double the swing at the output.

9.5.4. Design methodology

1. We bias the transistor at $J_{c,opt}$, which refers to the current density at which maximum f_T occurs.
2. We sweep transistor sizes of the output stages to get the best P1dB.
3. We set L to the minimum value because this maximizes f_T , as it is inversely proportional to l^2 .
4. For the output stage: -
5. $W = 2\text{um} * 30$, $L = 60 \text{ nm}$ for the input transistor.
6. $W = 2\text{um} * 28$, $L = 60 \text{ nm}$ for the cascode transistor.
7. We bias the cascode transistor such that we make the voltage drop across each transistor the same and equal 1.2 V.
8. We make a power match, not a conjugate match at the output of the second stage to get maximum power at the output.

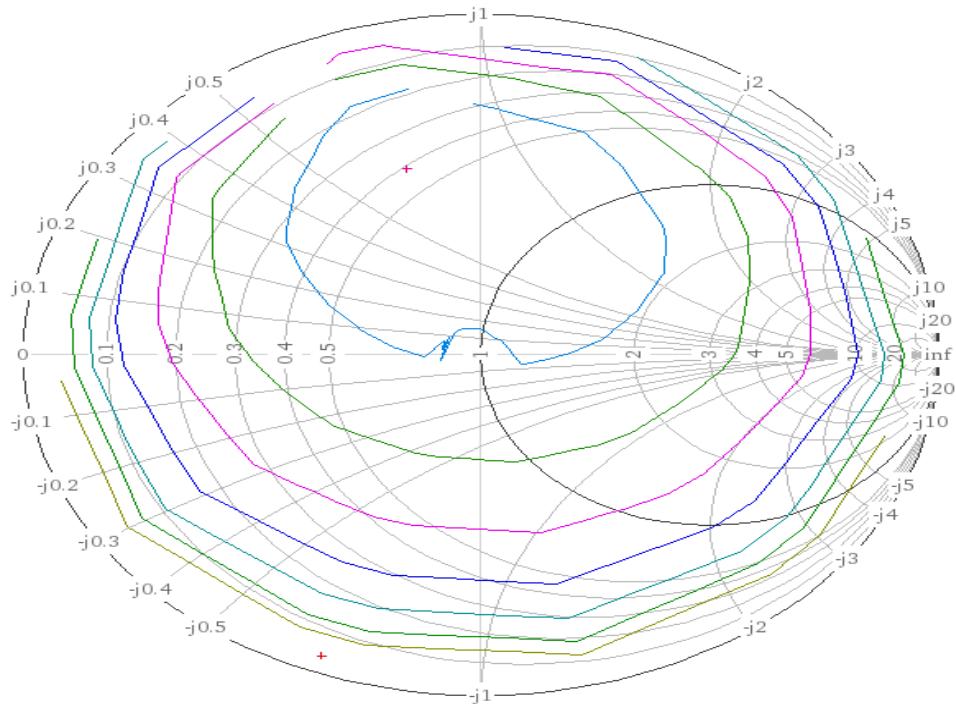


Fig. 9. 38: load-pull-analysis for the second stage

But if we select the best value, this will make S_{22} equals -4 dB. So, we select a value for output load such as it maximizes output power, but also keeps S_{22} below -10dB.

Then, we move to the first stage and choose transistor sizes such as they maximize gain.

$W = 2\text{um} * 11$, $L = 60 \text{ nm}$ for all transistors.

We make a conjugate match at the input of the first stage.

For the inter-stage matching, try to avoid making matching lossy by making the two impedances far on the Smith chart. It is right that if you make the transistor sizes of the first stage small, you will increase efficiency and decrease power consumption, but this also would make inter-stage matching very lossy and decrease efficiency.

9.5.4.1. Power amplifier schematic

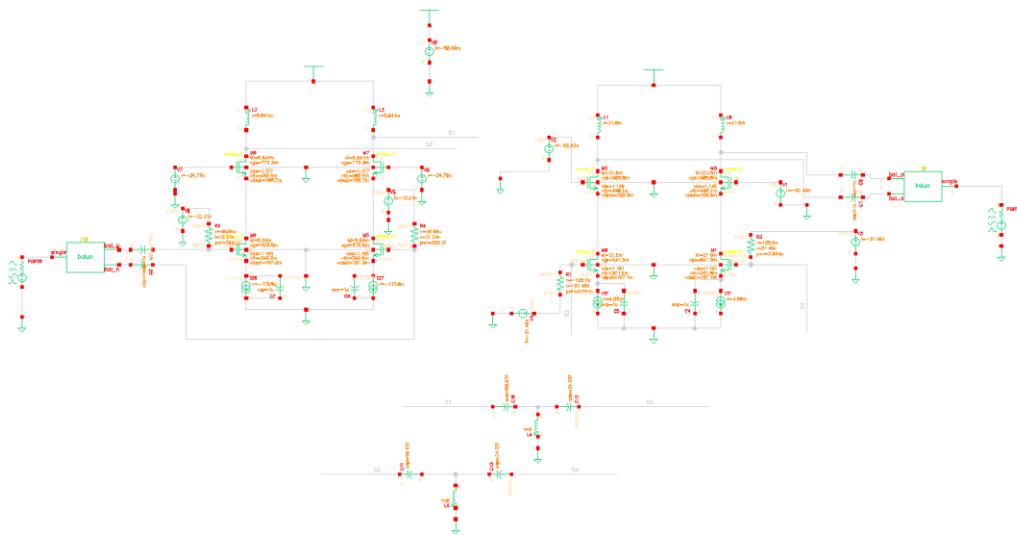


Fig. 9. 39: circuit schematic

This design consumes 8.64 mA per branch in the first stage and 21.6 mA per branch in the second stage.

9.5.4.2. Ideal Results

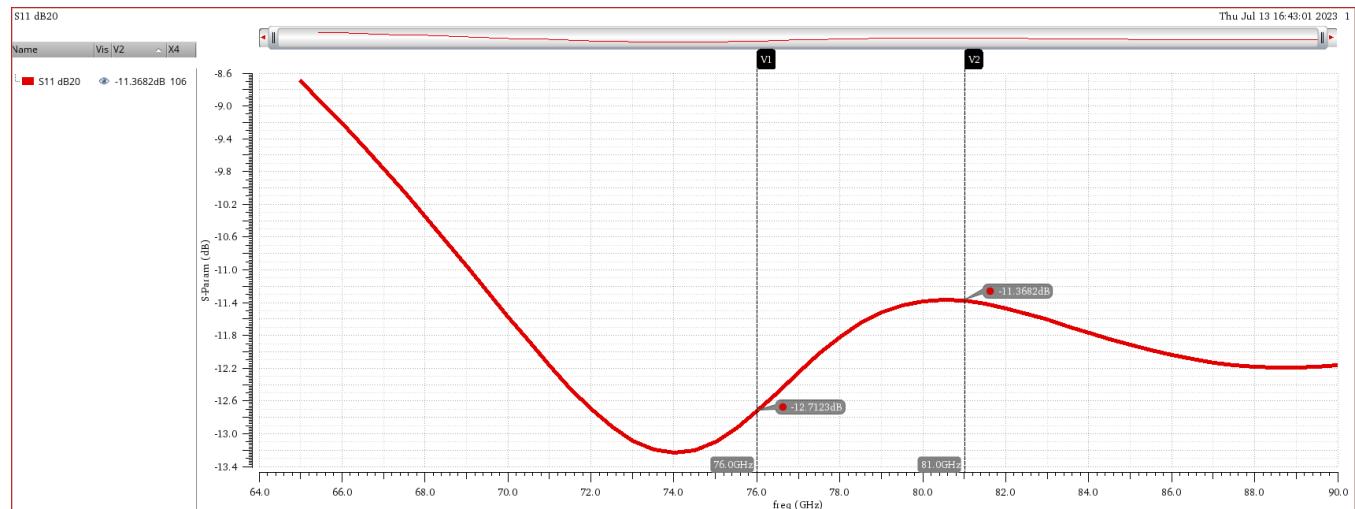


Fig. 9. 40: S_{11} vs frequency

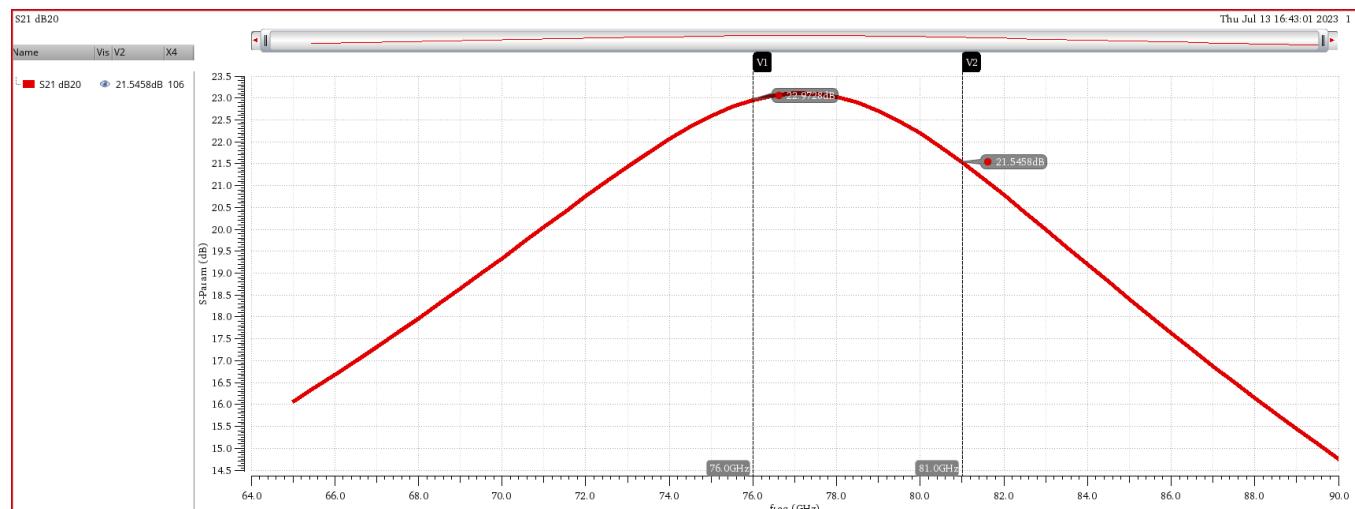


Fig. 9.41: S_{21} vs frequency

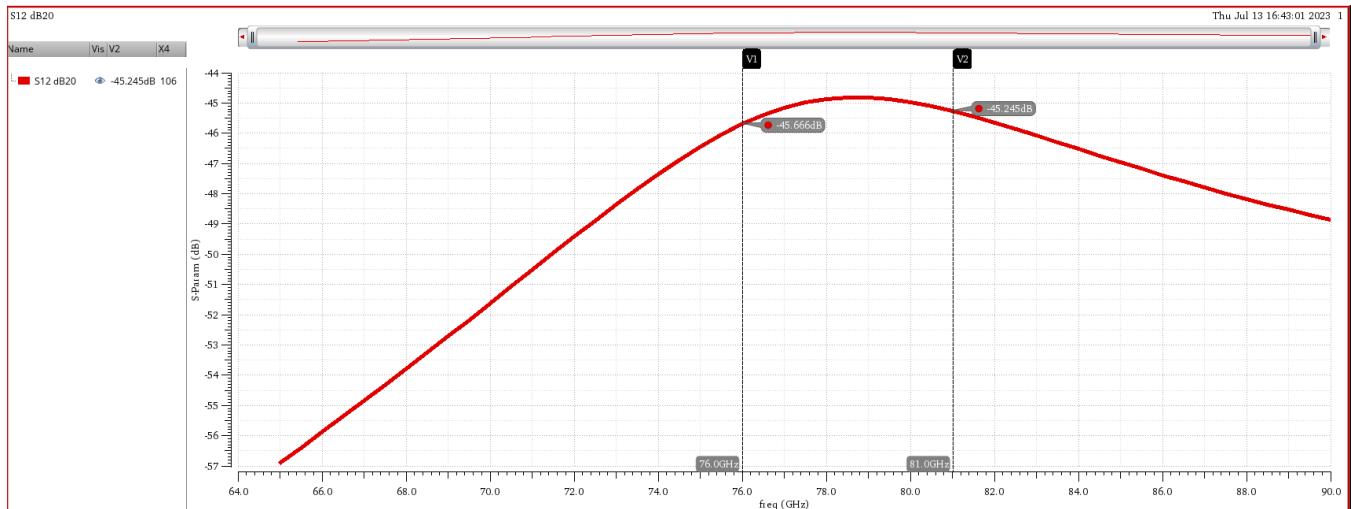


Fig. 9.42: S_{12} vs frequency.

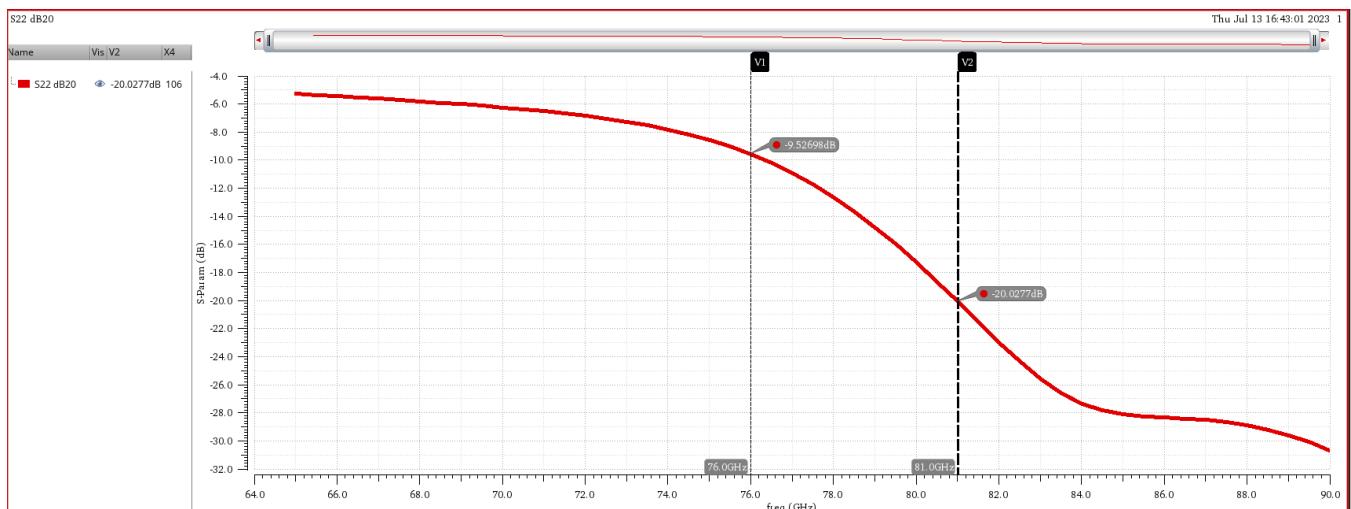


Fig. 9.43: S_{22} vs frequency.

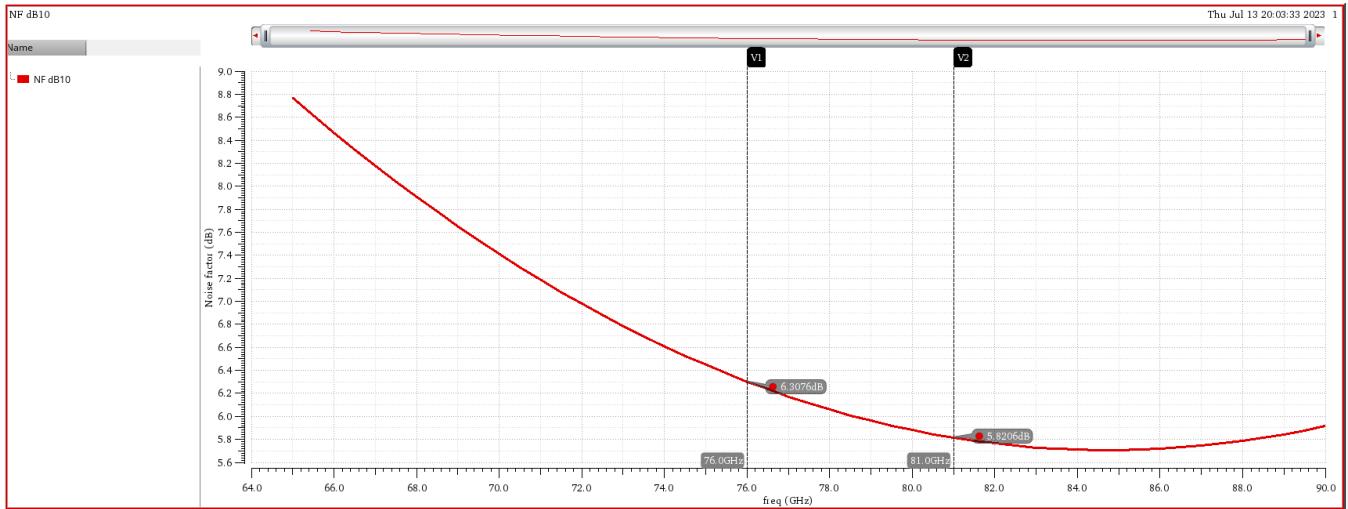


Fig. 9. 44: NF vs frequency.

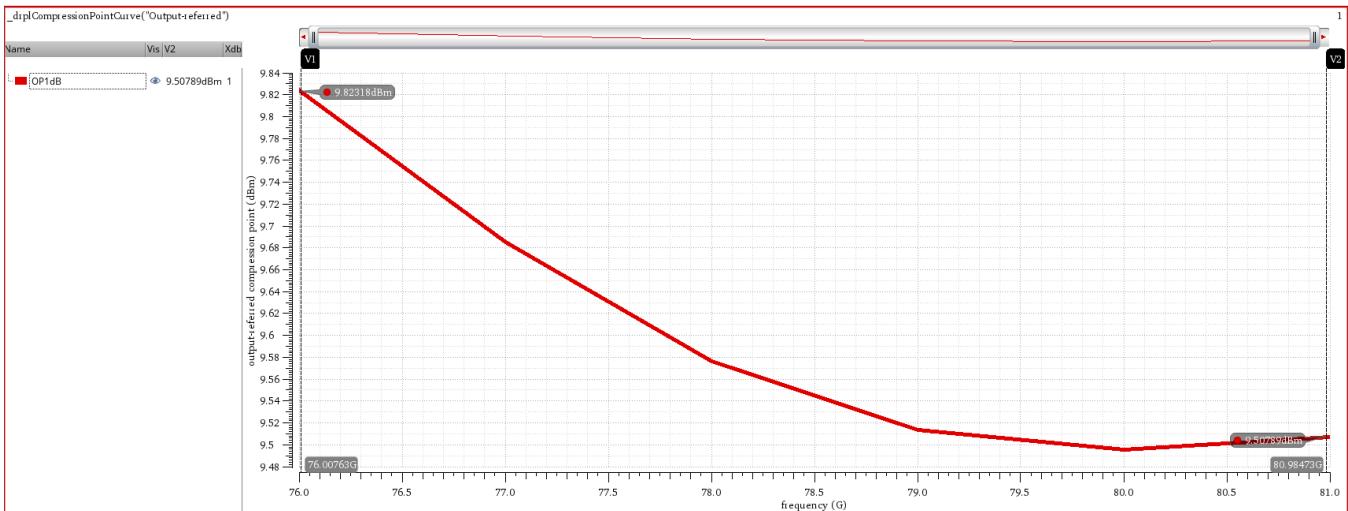


Fig. 9. 45: Op1dB vs frequency.

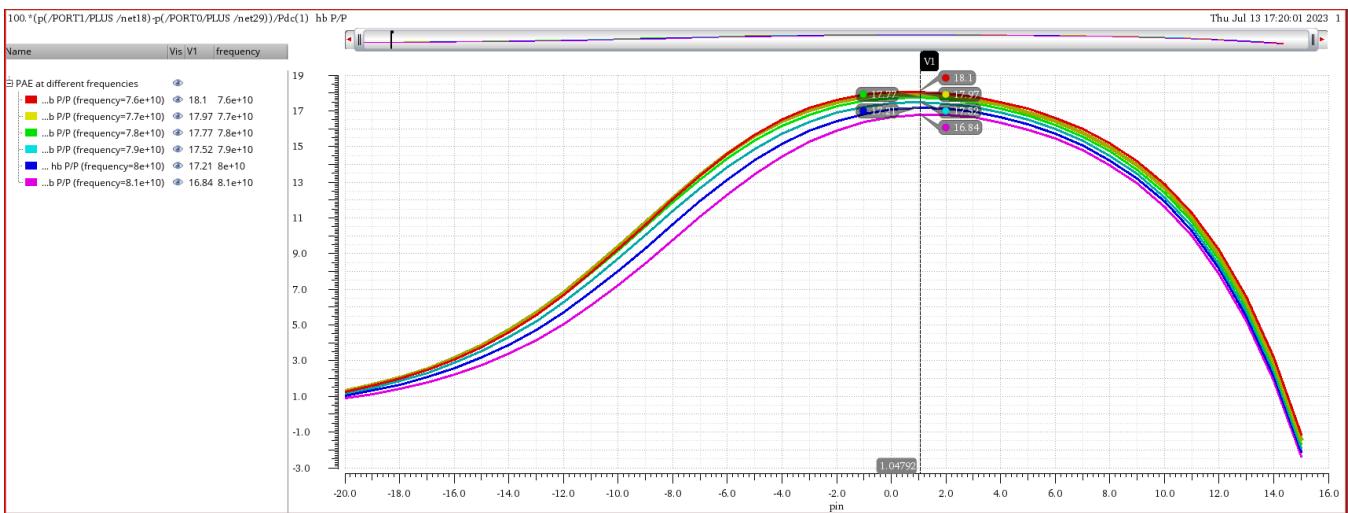


Fig. 9. 46: PAE vs P_{in} at different frequencies.

9.5.5. PA layout

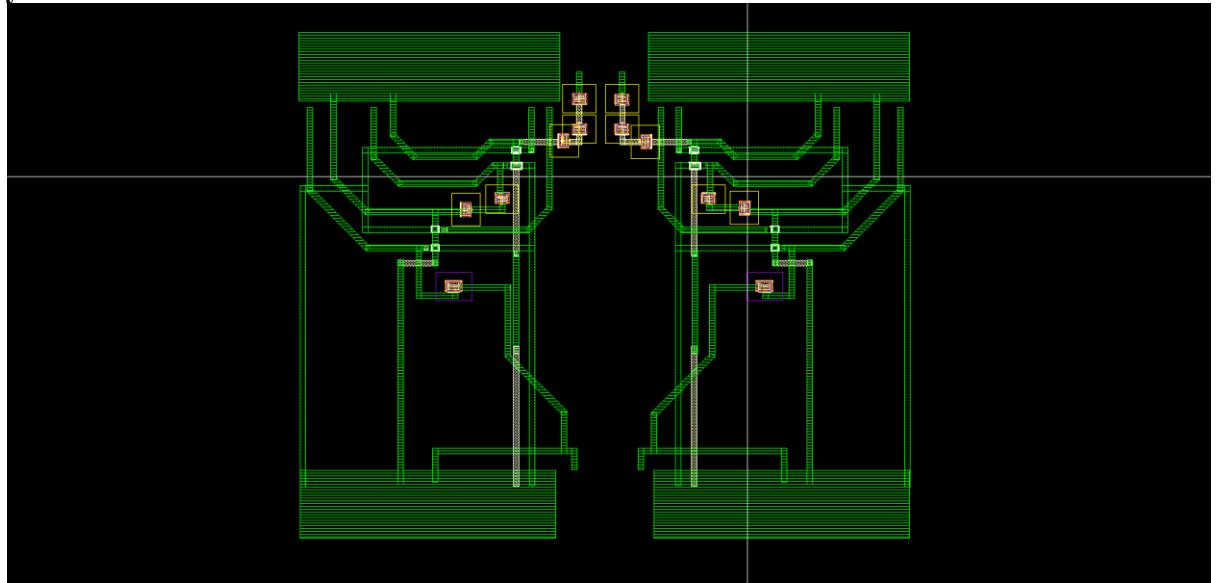


Fig. 9. 47: Power amplifier layout

Layout considerations

1. Same orientation: - all devices should be in the same orientation. So, if any error in dimensions occurs during fabrication, it occurs in all devices with the same amount.
2. Same neighbors: - the devices that need to be matched should see the same surroundings.
3. IR drops: - As $R = \frac{\rho L}{A}$, wires in both branches in the differential structure should be of the same length to get the same node voltages in both branches.

9.5.6. Post-layout simulations: -

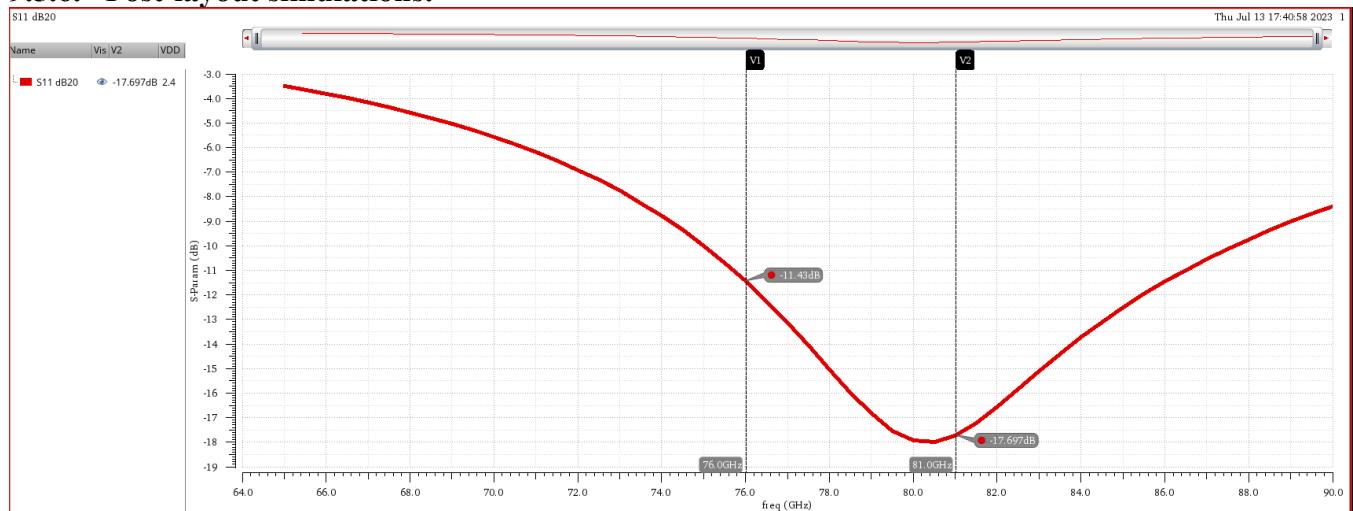


Fig. 9. 48: S_{11} vs frequency.

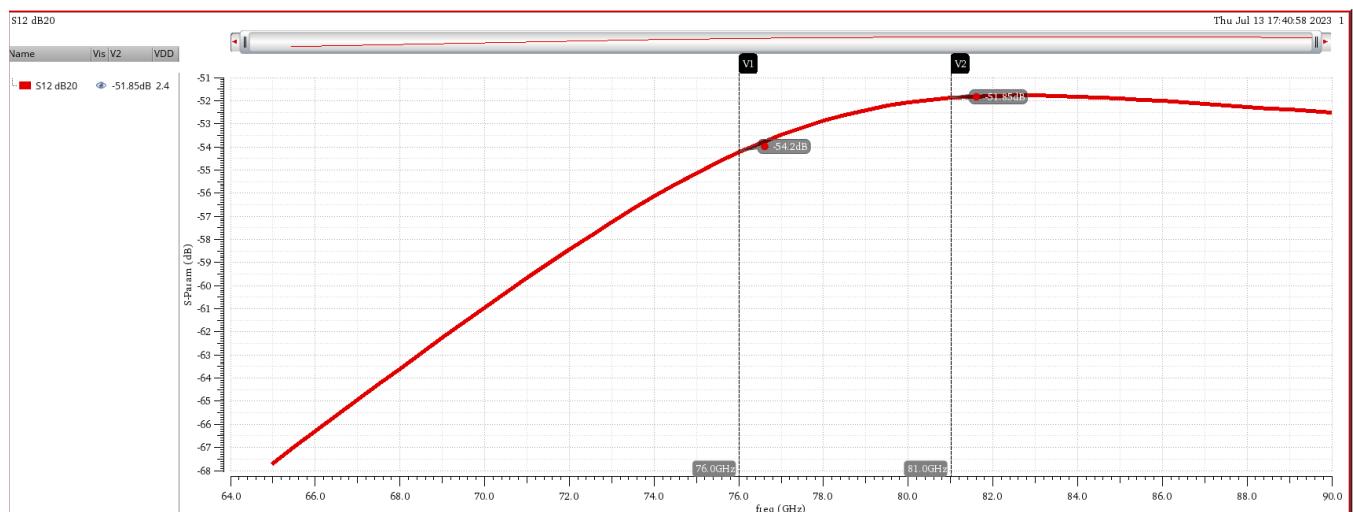


Fig. 9. 49: S_{12} vs frequency.

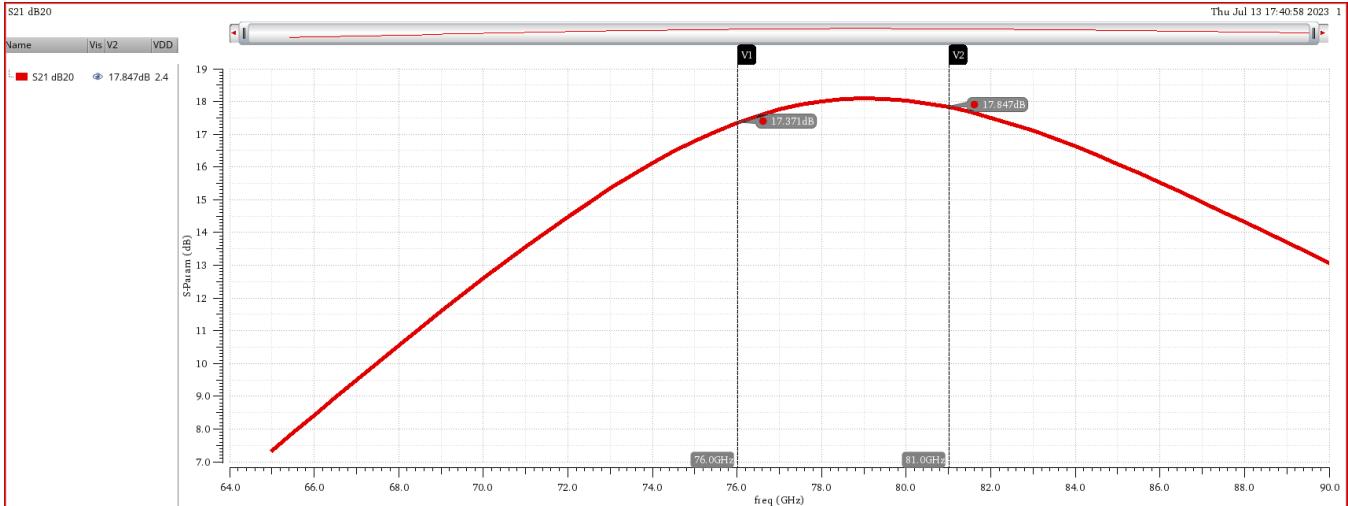


Fig. 9. 50: S_{21} vs frequency.

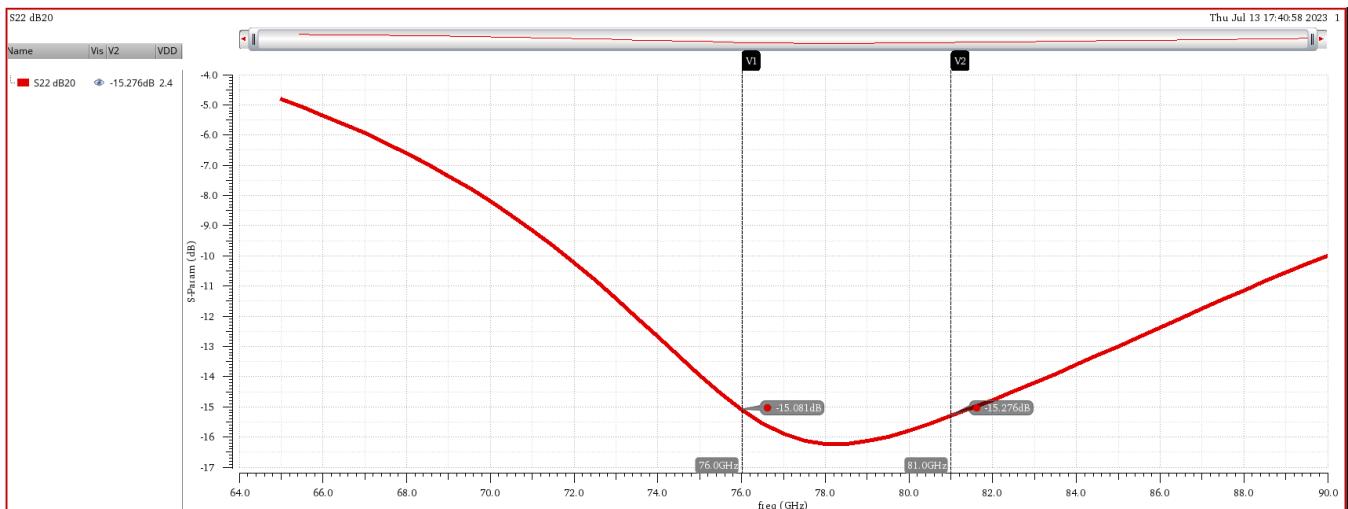


Fig. 9. 51: S_{22} vs frequency

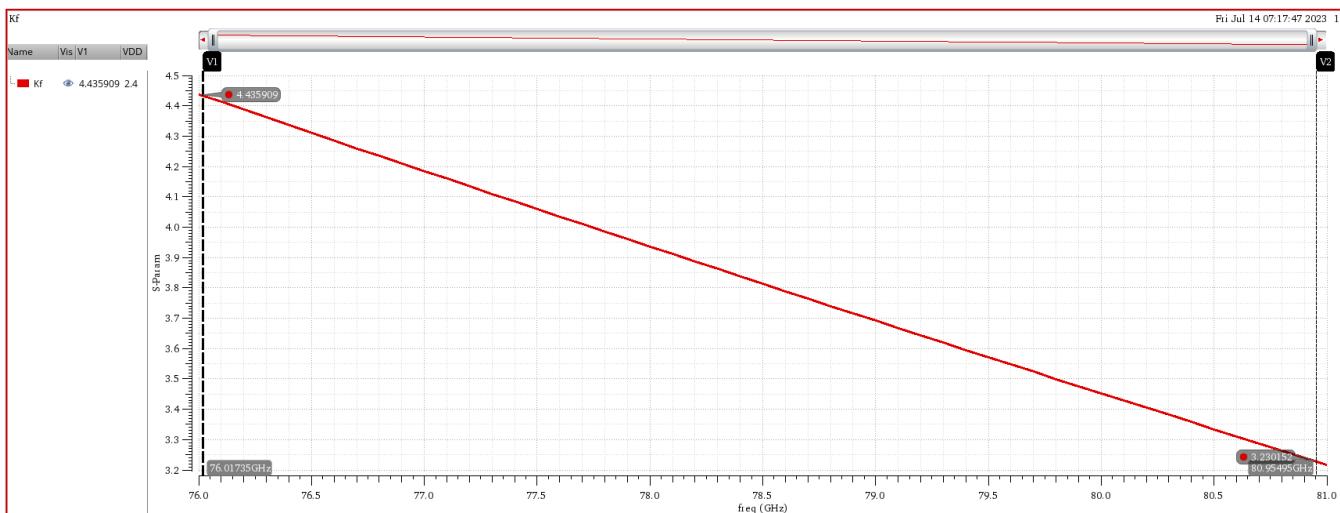


Fig. 9. 52: K_f for first stage vs frequency

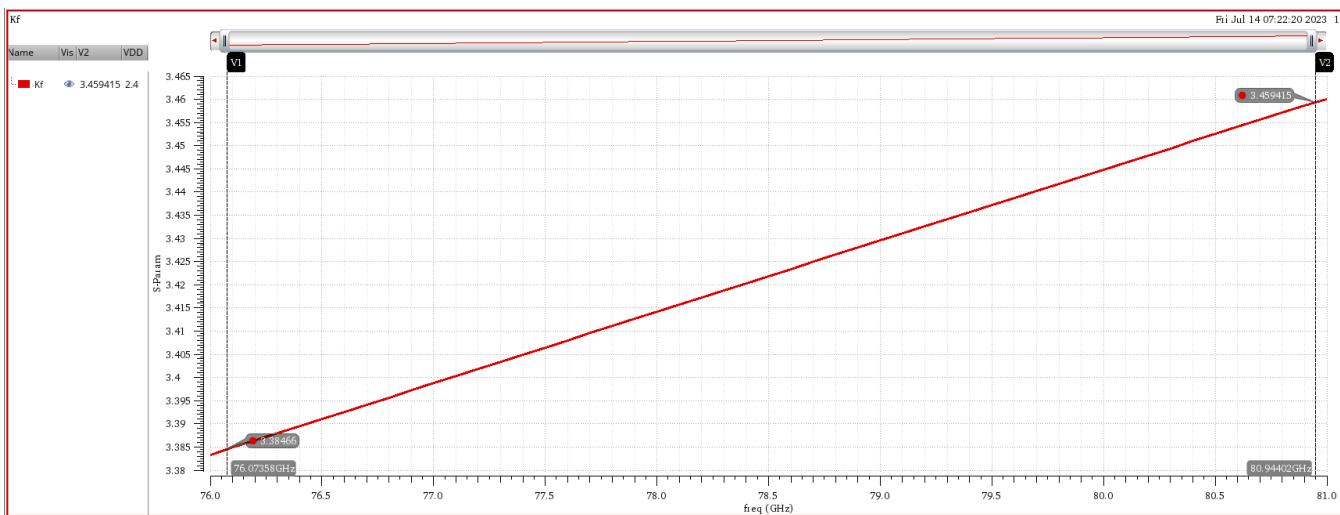


Fig. 9. 53: K_f for second stage vs frequency

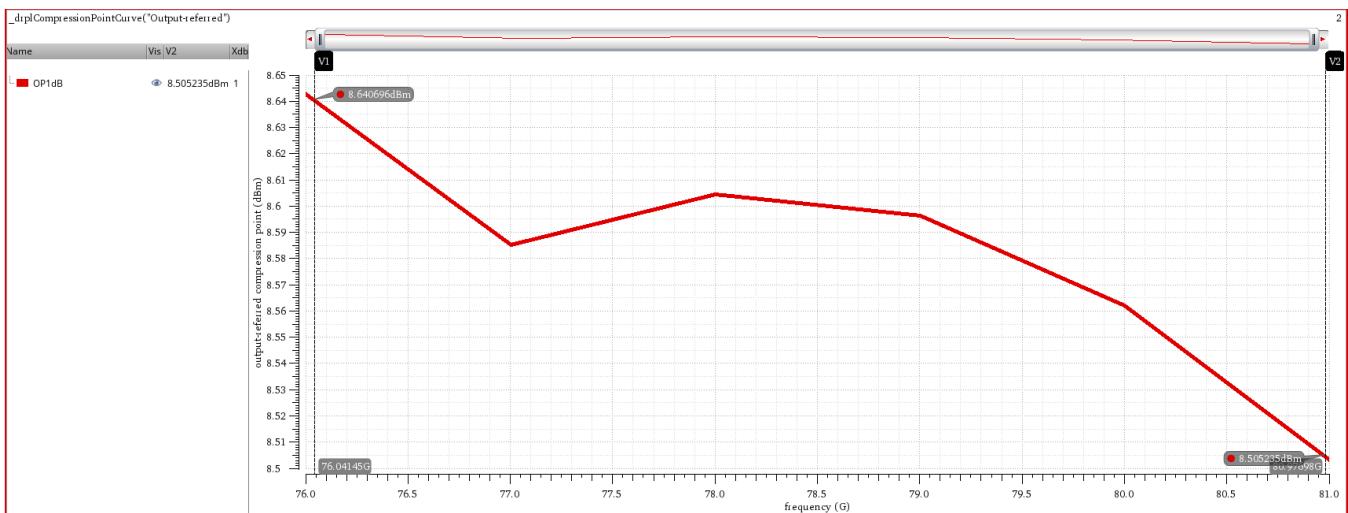


Fig. 9. 54: $OP1dB$ vs frequency

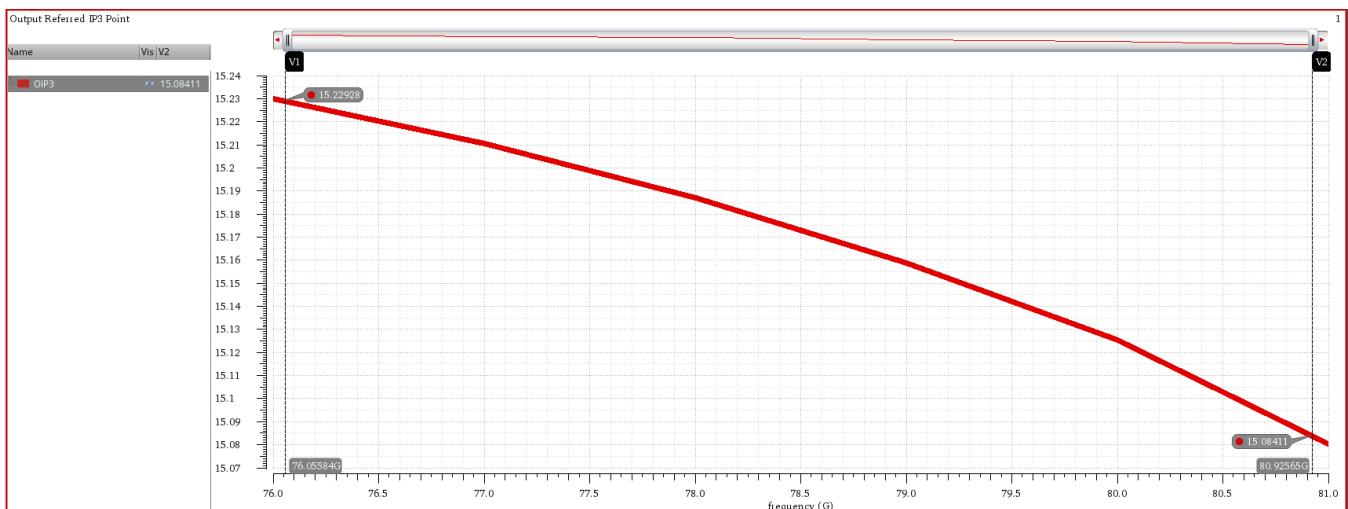


Fig. 9. 55: $OIP3$ vs frequency.

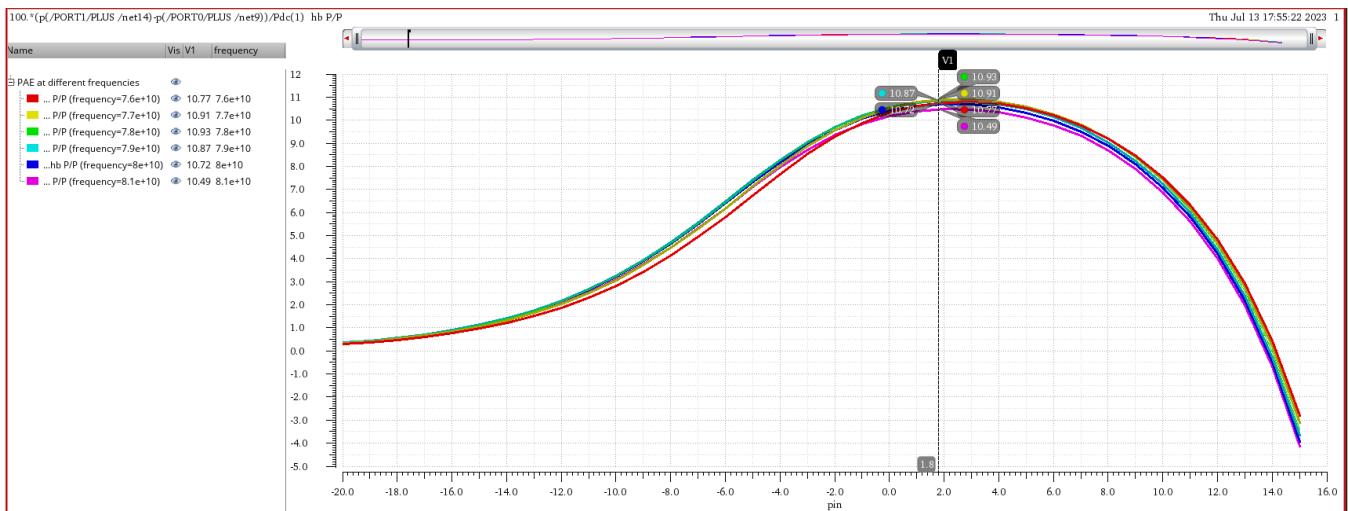


Fig. 9. 56: *PAE* at different frequencies.

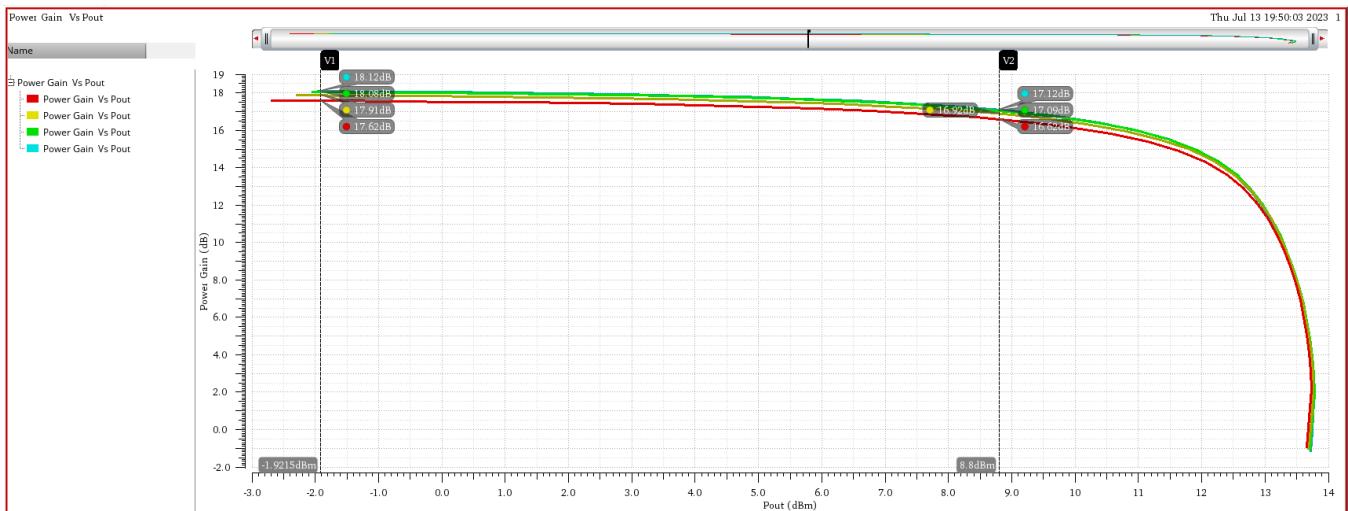


Fig. 9. 57: *gain VS P_{out}* at different frequencies.

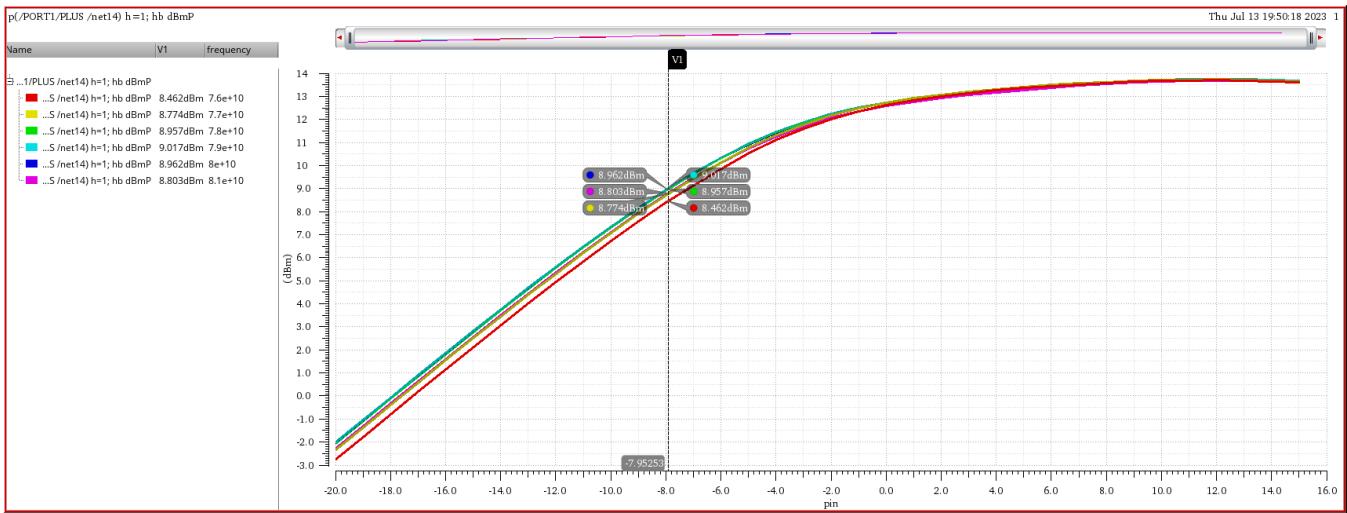


Fig. 9. 58: P_{out} vs P_{in} at different frequencies.

9.5.7. Corners: -

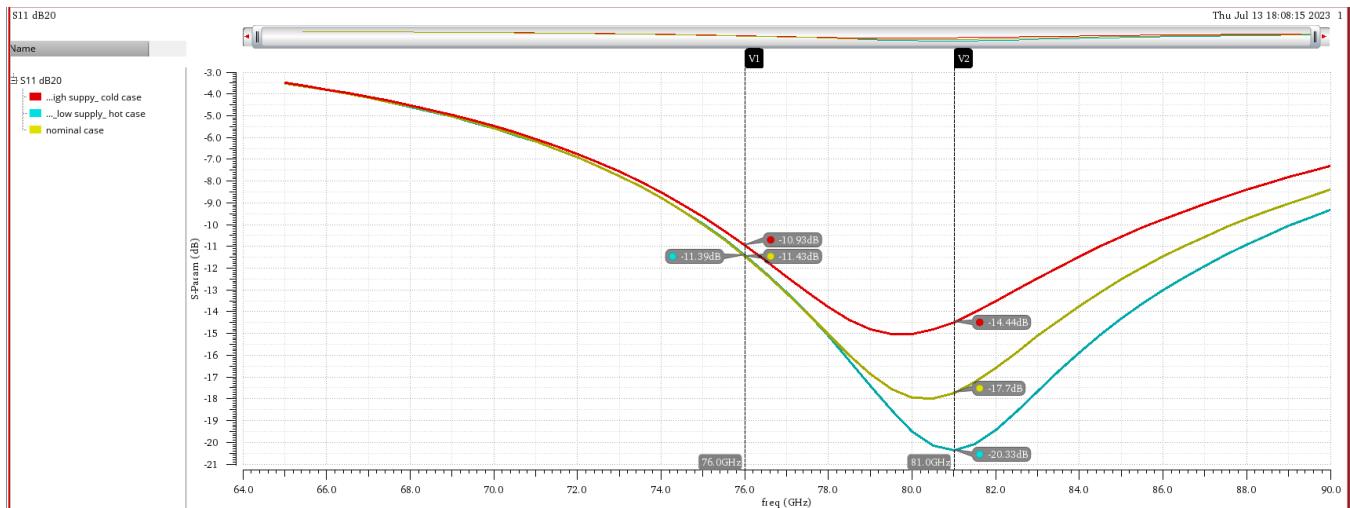


Fig. 9. 59: S_{11} corners VS frequency

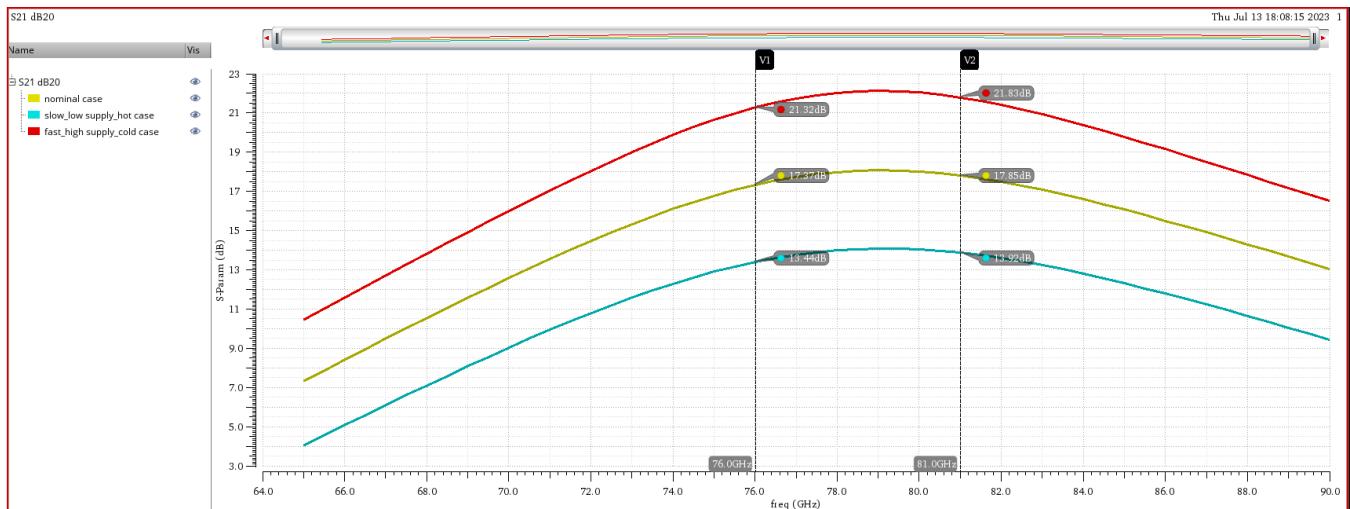


Fig. 9. 60: S_{21} corners VS frequency

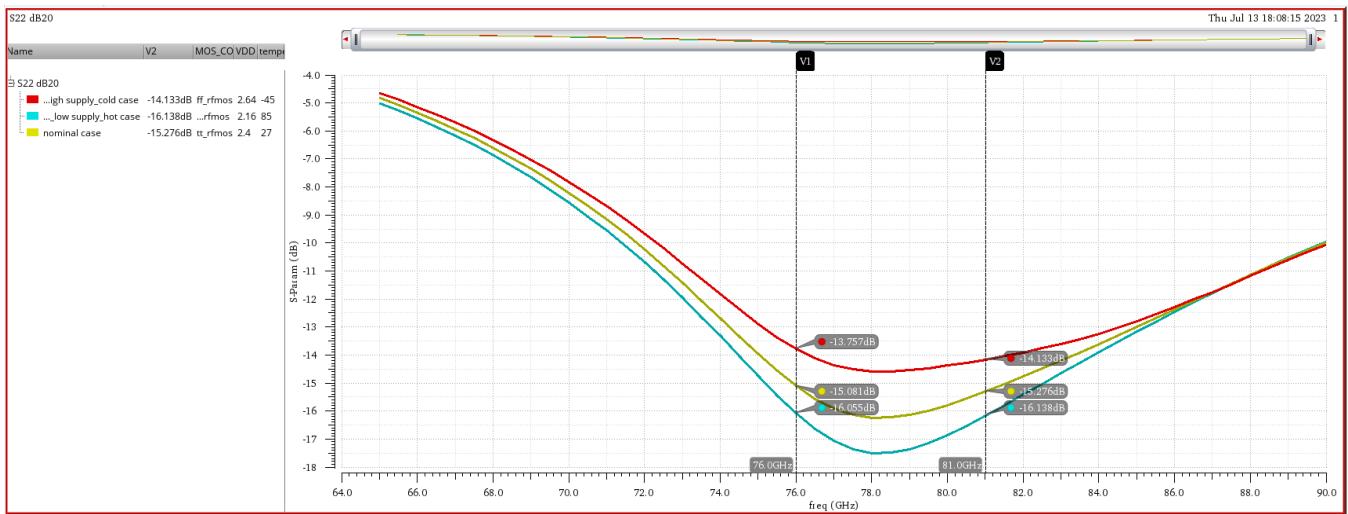


Fig. 9.61: S_{22} corners VS frequency

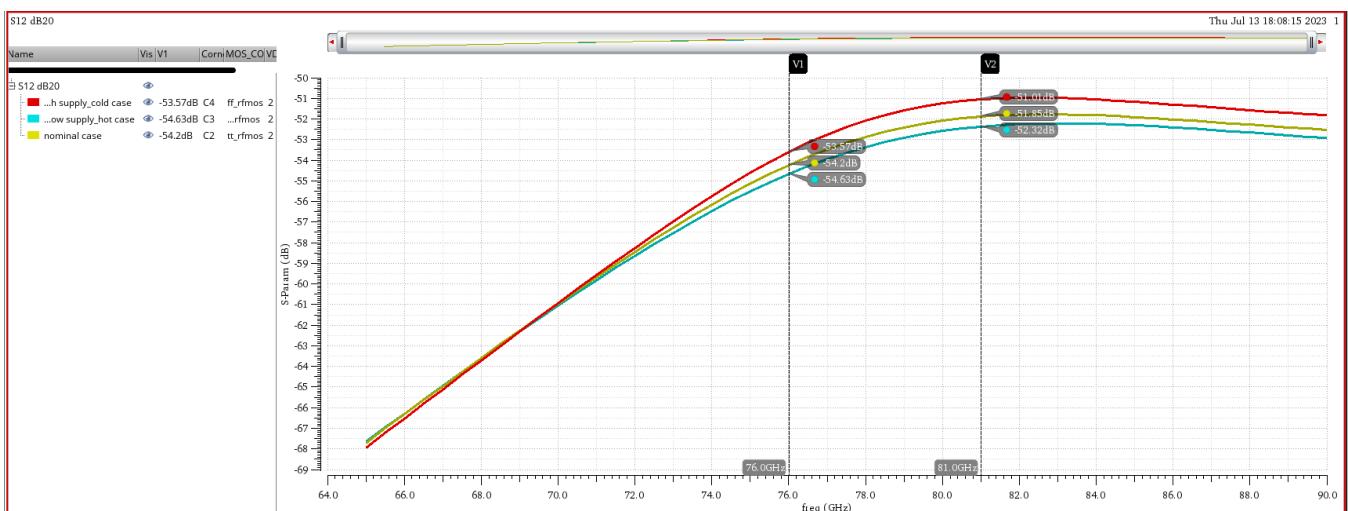


Fig. 9.62: S_{12} corners VS frequency

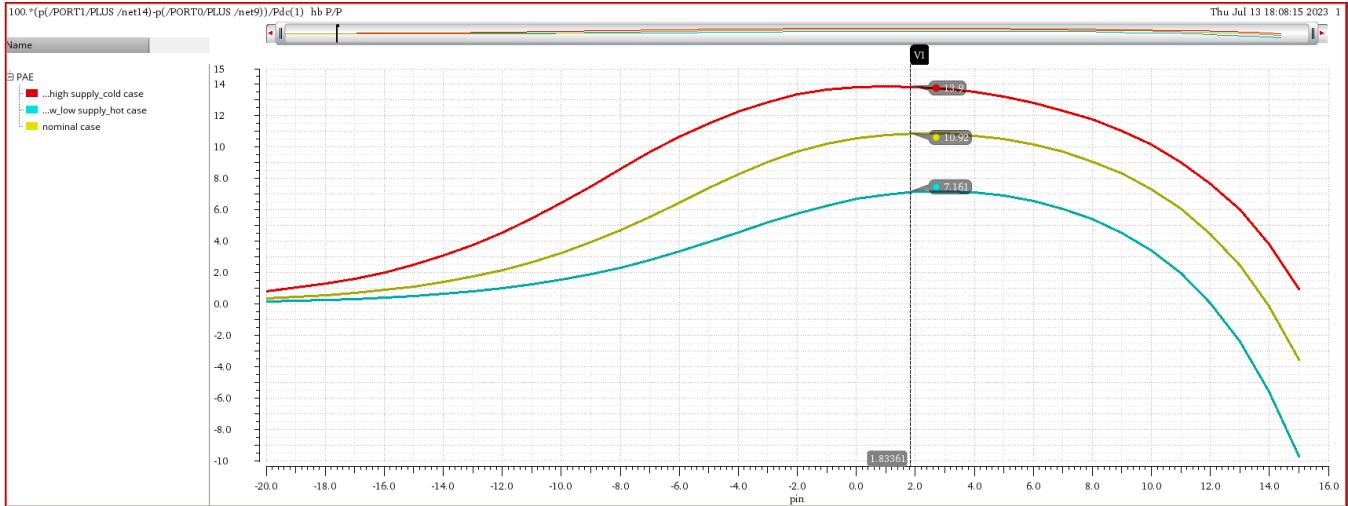


Fig. 9.63: *PAE corners VS P_{in} at 78.5 GHz.*

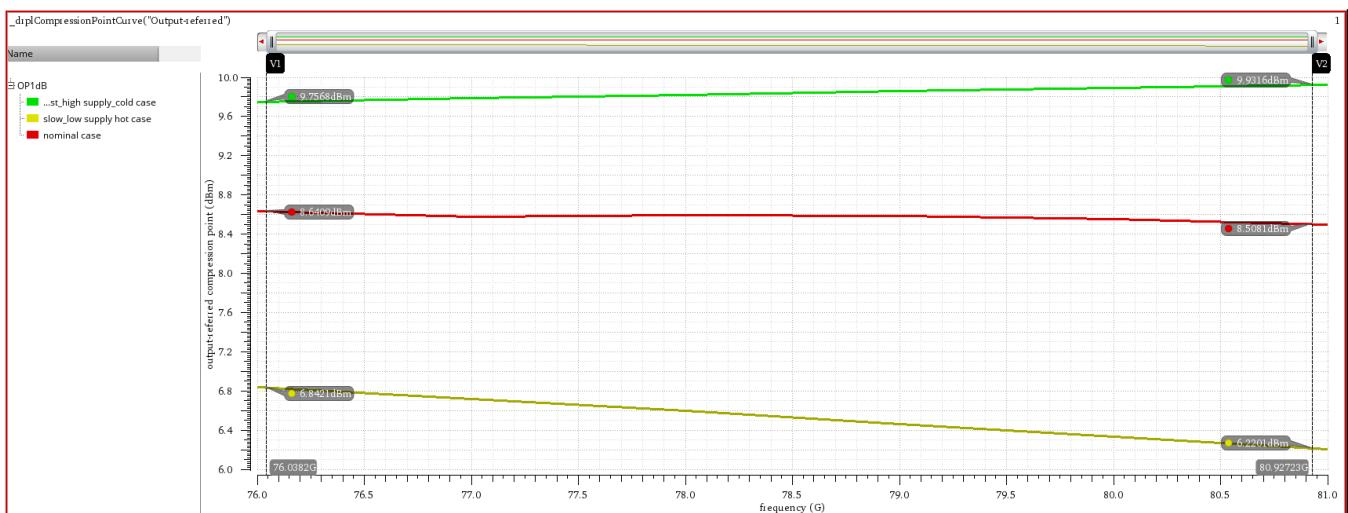


Fig. 9.64: *OP1dB corners VS frequency.*

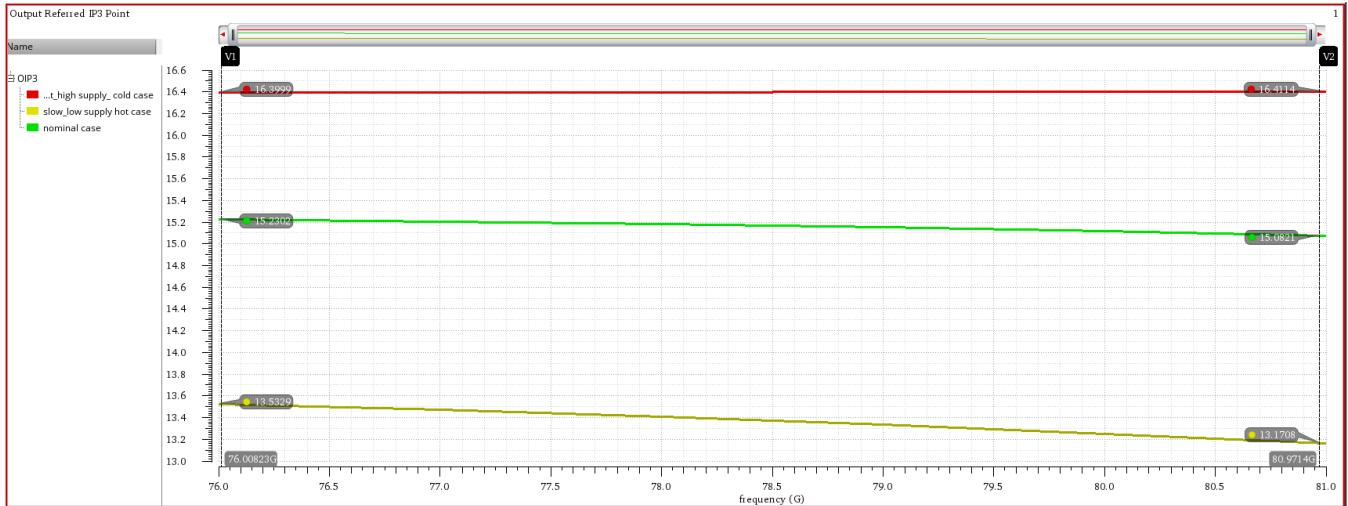


Fig. 9. 65: OIP3 corners VS frequency.

It is observed that the worst-case corner when $V_{DD} = 2.16$ v, temperature = 125° C and ss transistors because of lower V_{DD} makes $OP1dB$ worse. Higher temperature causes higher losses and hence efficiency drops. Slow transistors also reduce linearity.

	[9]	[8]	This work
Process	65-nm CMOS	65-nm CMOS	65-nm CMOS
Frequency	76-81GHz	77GHz	76-81GHz
Topology	Differential Cascode	Differential Cascode	Differential Cascode
Number of stages	3	-	2
Gain	15 dB	15 dB	18.11 dB
P_{out}	11.45 dBm	11 dBm	13.6 dBm
OP1dB	7 dBm	7 dBm	8.6 dBm
PAE	6.8% at output power	10% (maximum)	10.9%
Power Cons.	187 mW	115 mW	145.152 mW

Table 9.4 : PA comparison with related design.

	Cascode topology	CS topology
Process	65-nm CMOS	65-nm CMOS
Frequency	76-81GHz	76-81GHz
Topology	Differential cascode	Differential CS
Number of stages	2	3
Gain	18.11 dB	21 dB
P_{out}	13.6dBm	13.7dBm
OP1dB	8.6 dBm	9.4dBm
Max PAE	10.9%	15%
Power Cons.	145.152mW	136mW

Table 9.5: PA two designs comparison.

References

- [1] S. Cripps, RF Power Amplifiers for Wireless Communications, Norwood, MA: Artech House, 1999
- [2] B. Razavi, *RF Microelectronics*. Inde: Dorling Kindersley, 2014.
- [3] C. Cui *et al.*, “A 77-ghz FMCW radar system using on-chip waveguide feeders in 65-NM CMOS,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 11, pp. 3736–3746, 2015. doi:10.1109/tmtt.2015.2477343
- [4] T. Ma *et al.*, “A CMOS 76–81-ghz 2-TX 3-RX FMCW radar transceiver based on mixed-mode PLL Chirp Generator,” *IEEE Journal of Solid-State Circuits*, vol. 55, no. 2, pp. 233–248, 2020. doi:10.1109/jssc.2019.2950184
- [5] H. Jia *et al.*, “A 77 GHz frequency doubling two-path phased-array FMCW transceiver for Automotive Radar,” *IEEE Journal of Solid-State Circuits*, vol. 51, no. 10, pp. 2299–2311, 2016. doi:10.1109/jssc.2016.2580599
- [6] L. Lancon, H. Vallee, G. Montoriol, F. Brunelli, and T. Taris, “Integrated-transformer-based impedance matching method: Impedance matching with Transformers,” *IEEE Microwave Magazine*, vol. 23, no. 9, pp. 40–56, 2022. doi:10.1109/mmm.2022.3158033
- [7] E. Bloch and E. Socher, “Beyond the smith chart: A universal graphical tool for impedance matching using transformers,” *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 100–109, Nov. 2014, doi: 10.1109/ MMM.2014.2356114.
- [8] A Fully Integrated 77-GHz FMCW Radar Transceiver in 65-nm CMOS Technology Jri Lee, Member, *IEEE*, *Yi-An Li, Meng-Hsiung Hung, and Shih-Jou Huang*.
- [9] A 77-GHz CMOS Automotive Radar Transceiver with Anti-Interference Function Tang-Nian Luo, Student Member, *IEEE*, *Chi-Hung Evelyn Wu, and Yi-Jan Emery Chen, Senior Member, IEEE*.
- [10] High-frequency integrated circuits (The Cambridge RF and microwave series) by Sorin.

10

Receiver Integration Results

This Chapter shows the simulated results and the layout for the integrated receiver chain, the chain is simulated at a constant IF frequency of 1MHz, while sweeping both RF and LO frequencies, the integration results are done across PVT corners, the simulated results show the following:

- Conversion gain, S11, S22 and NF vs frequency.
- IIP3 and OIP3 vs frequency.

10.1. Simulation Results

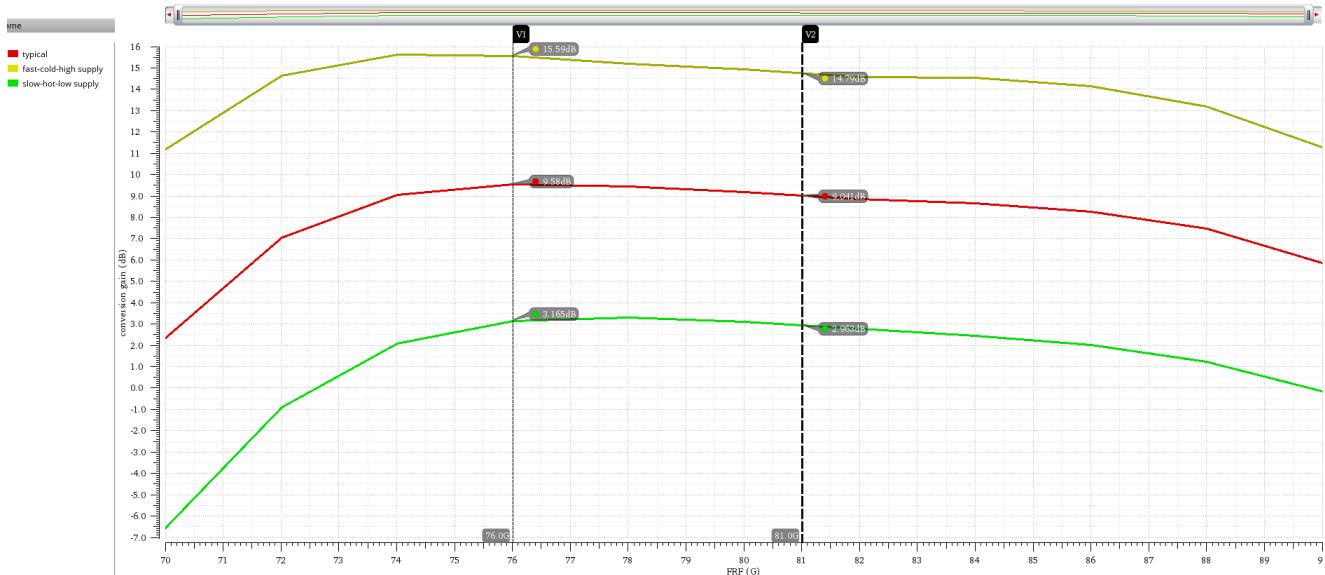


Fig. 10.1: conversion gain vs frequency across PVT corners.

Fig. 10.1: conversion gain vs frequency across PVT corners. Fig. 10.1 shows the overall conversion gain of the receiver chain, the gain is around 9dB in the nominal corner while for the fast and slow corners the gain is around 15dB and 3dB respectively.

Fig. 10.1: conversion gain vs frequency across PVT corners. conversion gain vs frequency across PVT corners Fig. 10.3 and Fig. 10.2 show S_{11} and S_{22} of the receiver chain across corners versus frequency and it can be seen that the both input and output matching are good (<-10dB) across all corners.

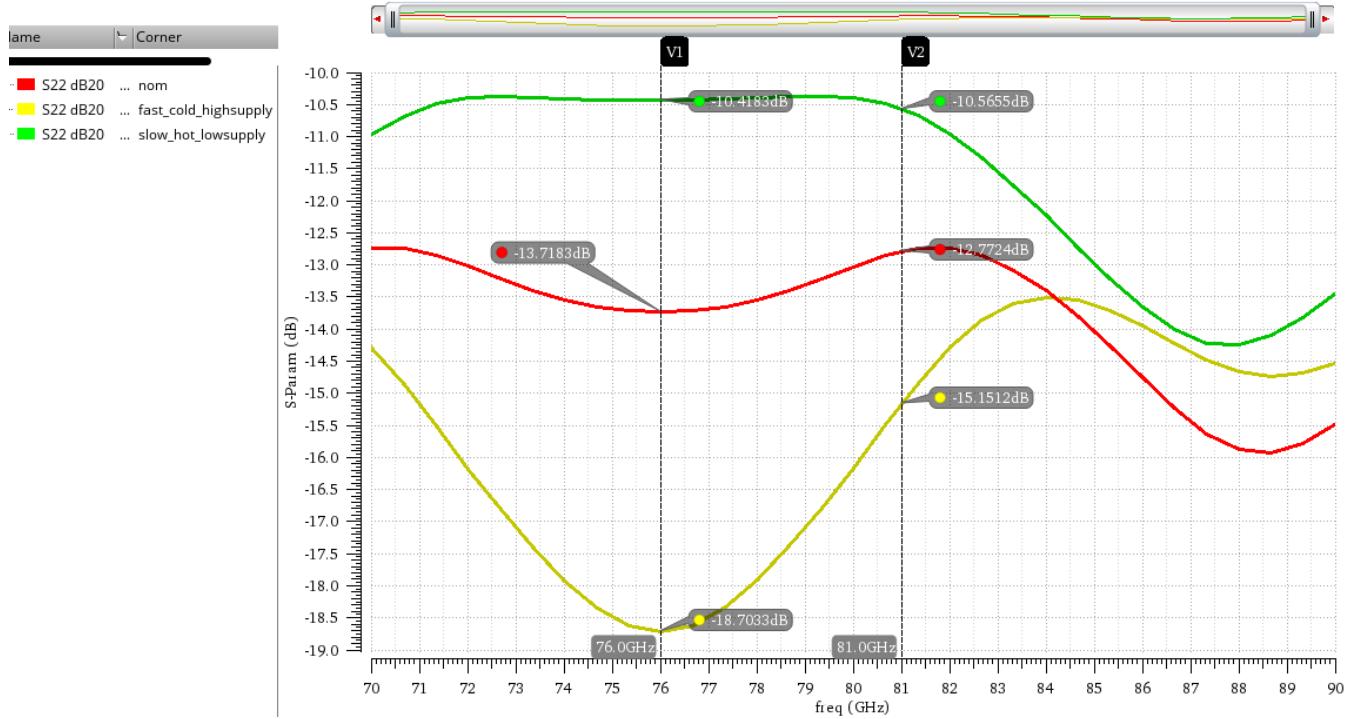


Fig. 10.2: S_{22} vs frequency across PVT corners.

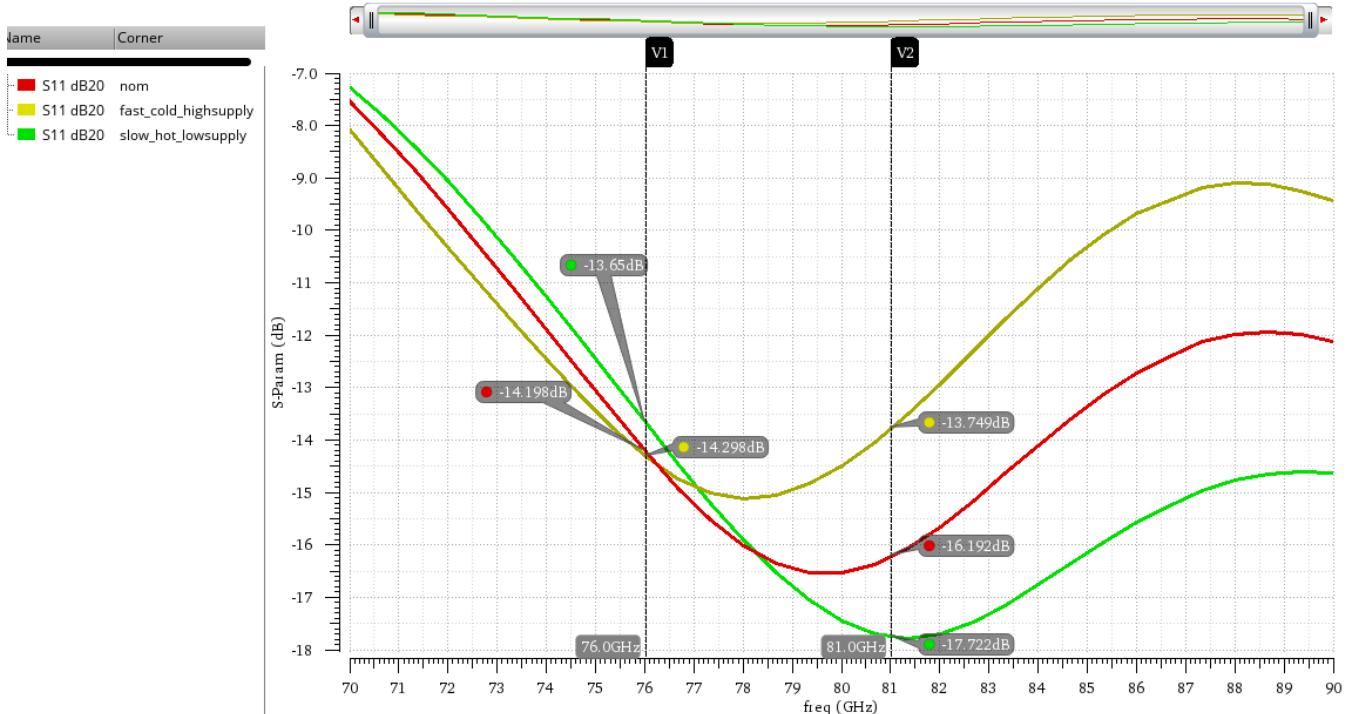


Fig. 10.4 shows the total *NF* of the receiver chain across corners versus frequency where the *NF* is less than 9.5dB in the nominal corner, less than 11.9dB for the slow corner and less than 7.5dB for the fast corner. While Fig. 10.5 shows the total *IIP3* value across corners versus frequency where the *IIP3* is larger than -11.4dBm for both nominal and slow corners and larger than -15dBm for the fast corner.

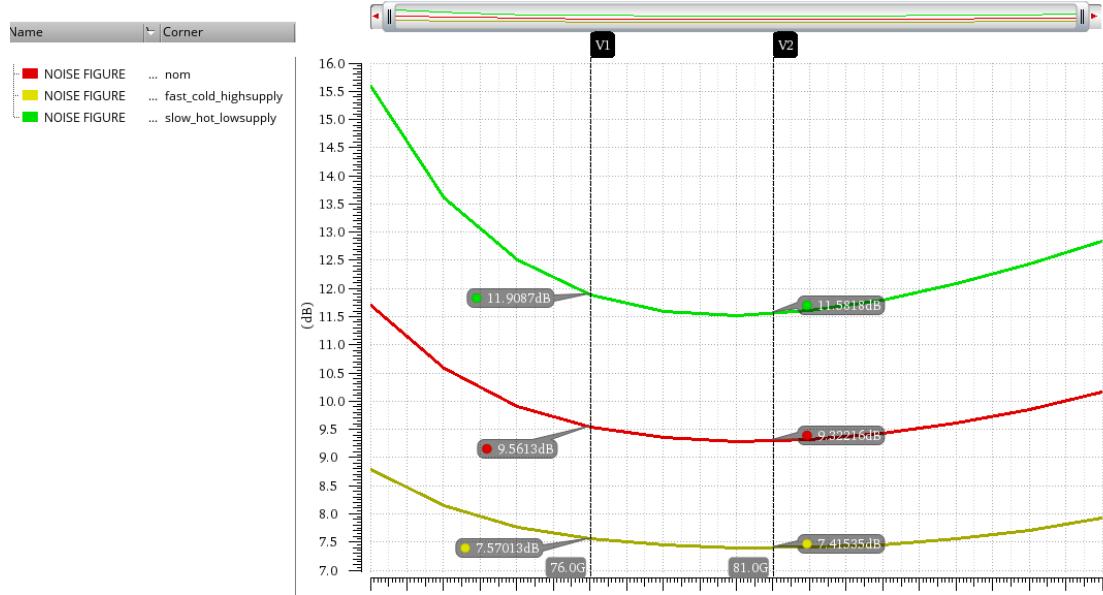


Fig. 10.4: *NF* vs frequency across PVT corners.

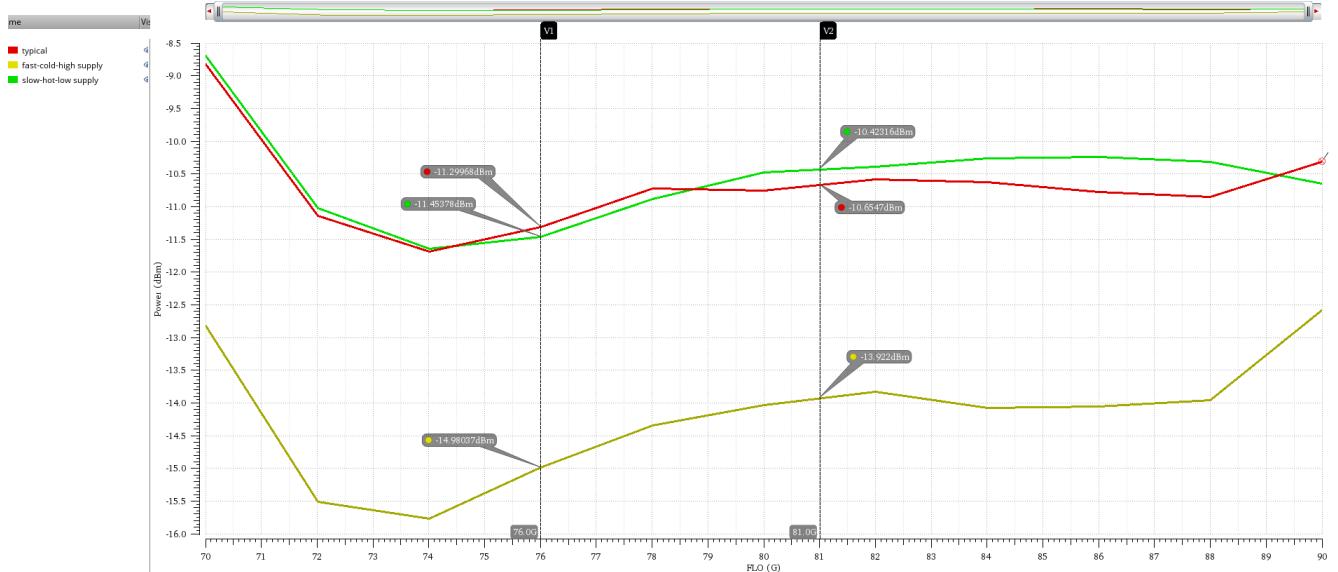


Fig. 10.5: *IIP3* vs frequency across PVT corners.

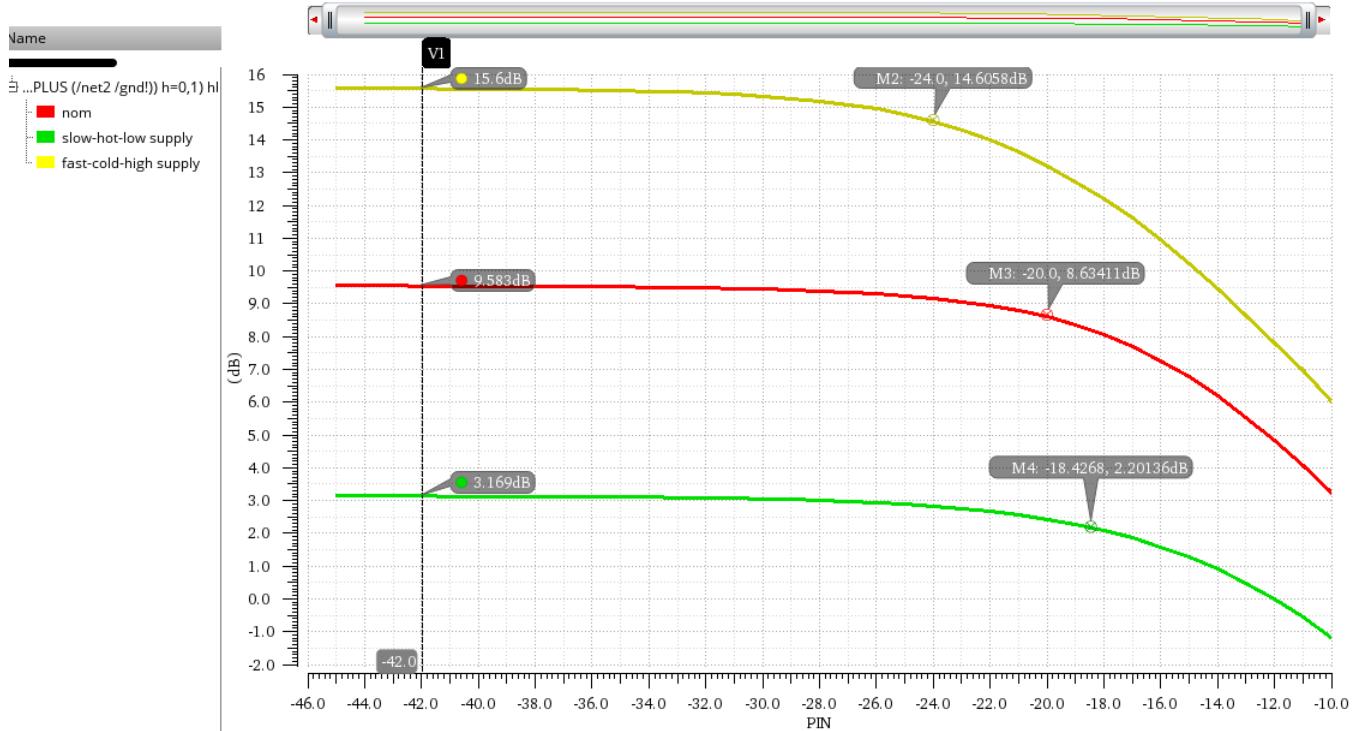


Fig. 10.6: power gain vs input power at 76GHz across corners.

Fig. 10.6 shows the power gain vs input power at 76GHz across corners, the IP_{1dB} point is shown in the graph where for the IP_{1dB} equals -20dBm nominal corner, -24dBm for the fast corner and -18.4dBm for the slow corner.

10.2. Transceiver layout

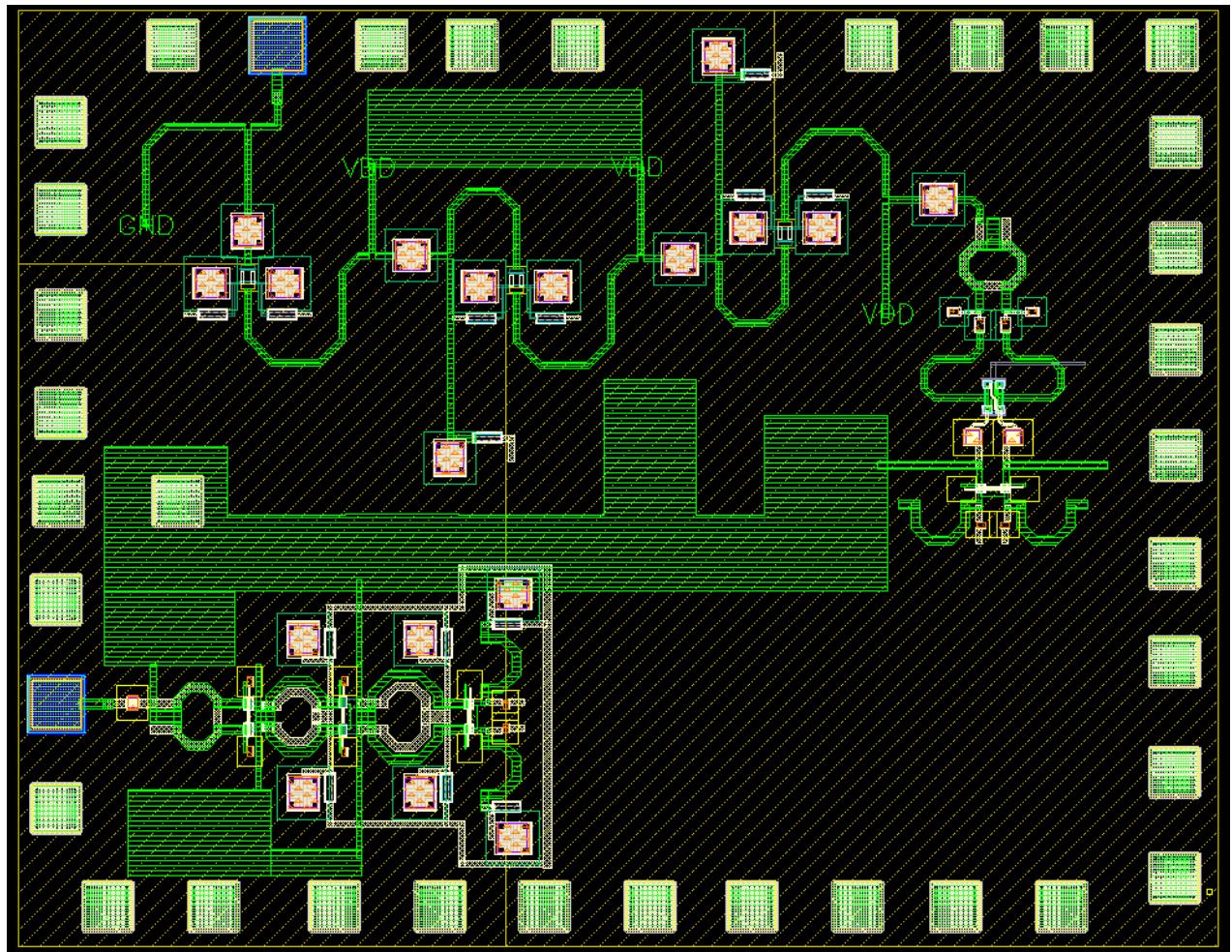


Fig. 10.7: Transceiver Layout.

10.3. Literature

	[1]	[2]	[3]	This work
Process	65-nm CMOS	65-nm CMOS	65-nm CMOS	65-nm CMOS
Frequency (GHz)	68-81	76-76.7	75-82	76-81
Gain (dB)	9	16	31.6*	9.5
NF (dB)	7.1	13	8.8	9.5
IP1dB (dBm)	-16.8	-20	-37	-20
Power Cons. (mW)	18	28.5	9.5	-

* Includes an IF amplifier.

References

- [1] H. Jia *et al.*, “A 77 GHz frequency doubling two-path phased-array FMCW transceiver for Automotive Radar,” *IEEE Journal of Solid-State Circuits*, vol. 51, no. 10, pp. 2299–2311, 2016. doi:10.1109/jssc.2016.2580599.
- [2] Le, V.H., Duong, H.T., Huynh, A.T., et al.: ‘A CMOS 77-GHz receiver front-end for automotive radar’, *IEEE Trans. Microw. Theory Tech.*, 2013, 61, (10), pp. 3783–3793.
- [3] Luo, T., Wu, C.E., Chen, Y.E.: ‘A 77-GHz CMOS automotive radar transceiver with anti-interference function’, *IEEE Trans. Circuits Syst. I*, 2013, 60, (12), pp. 3247–3255.