Jonas Piccinotti

Università degli Studi di Napoli Federico II

Ph.D itee Information Technology Electrical Engineering

The purpose of this PhD research was to develop and optimize a fast numeric algorithm able to compute monostatic and bistatic RCS predictions obtaining an accuracy comparable to what commercially available from well-known electromagnetic CADs, but requiring unprecedented computational times. This was realized employing asymptotic approximated methods to solve the scattering problem, namely the Geometrical Optics (GO) and the Physical Optics (PO) theories, and exploiting advanced algorithmical concepts and cutting-edge computing technology to drastically speed-up the computation.

The First Chapter focuses on an historical and operational overview of the concept of Radar Cross Section (RCS), with specific reference to aeronautical and maritime platforms. How geometries and materials influence RCS is also described.

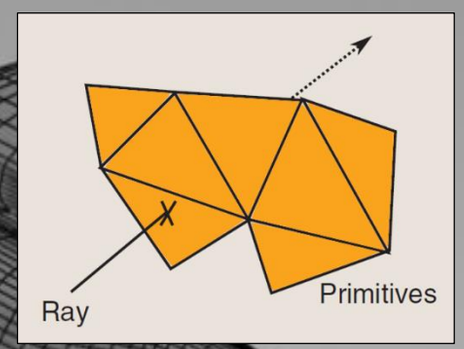
The Second Chapter is dedicated to the first phase of the algorithm: the electromagnetic field transport phase, where the GO theory is applied to implement the “ray tracing”. In this Chapter the first advanced algorithmical concept which was adopted is described: the Bounding Volume Hierarchy (BVH) data structure. Two different BVH approaches and their combination are described and compared.

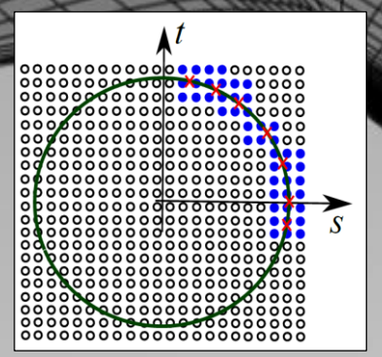
The Third Chapter is dedicated to the second phase of the calculation: the radiation integral, based on the PO theory, and its numerical optimization. Firstly the Type-3 Non-Uniform Fast Fourier Transform (NUFFT) is presented as the second advanced algorithmical tool that was used and it was indeed the foundation of the calculation of the radiation integral. Then, to improve the performance but also to make the application of the approach feasible in case of electrically large objects, the NUFFT was further optimized using a “pruning” technique, which is a stratagem used to save memory and computational time by avoiding calculating points of the transformed domain that are not of interest.

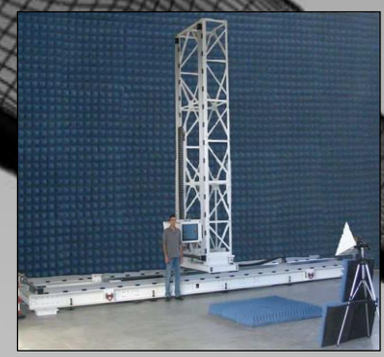
To validate the algorithm, a preliminary measurement campaign was held at the headquarter of the Ingegneria Dei Sistemi (IDS) Company, located in Pisa. The measurements, performed on canonical scatterers using a Synthetic Aperture Radar (SAR) imaging equipment set up on a planar scanner inside a semi-anechoic chamber, are discussed.

# Fast GO/PO RCS calculation:

## A GO/PO parallel algorithm implemented on GPU and accelerated using a BVH data structure and the Type 3 Non-Uniform FFT



Jonas Piccinotti

Ultra-Fast RCS prediction algorithm

2017 - XXIX

Ph.D. in Information Technology and Electrical Engineering

UNIVERSITÀ DEGLI STUDI DI NAPOLI
FEDERICO II


# UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

## PH.D. THESIS IN

### INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING

# FAST GO/PO RCS CALCULATION

## A GO/PO PARALLEL ALGORITHM IMPLEMENTED ON GPU AND ACCELERATED USING A BVH DATA STRUCTURE AND THE TYPE 3 PRUNED NON-UNIFORM FFT


## JONAS PICCINOTTI

TUTORS: AMEDEO CAPOZZOLI
ANGELO LISENO
CLAUDIO CURCIO


XXIX CICLO


SCUOLA POLITECNICA E DELLE SCIENZE DI BASE
DIPARTIMENTO DI INGEGNERIA ELETTRICA E TECNOLOGIE DELL'INFORMAZIONE

*I would like to thank my Professors Amedeo Capozzoli, Angelo Liseno and Claudio Curcio for guiding me along the arduous path that started with the Bachelor's Degree, continued with the Master Degree, and it is now completed by this Ph.D. work. They offered me unparalleled competence and constant support throughout all the adverse logistic situations encountered during the developing of this research.*

*I would also like to thank the IDS - Ingegneria Dei Sistemi - Company for the precious resources provided. Specifically, sincere gratitude is due to Stefano Sensani, Antonio Sarri, Luca Fiori, and Giacomo De Mauro, who accompanied me during the measurement campaign held at IDS headquarter in Pisa. In fact, they are now working with me and the Italian Air Force for projects that encompass the problems approached in this research and this is a proof of their value and vast knowledge in this framework.*

*Heartfelt appreciation goes to my Squadron, the Reparto Sperimentale Volo (Flight Test Center) of the Italian Air Force. All my Superior Officers were always sensitive to the importance of the choice I made committing myself to this demanding activity: not only they were comprehensive, but also offered me the possibility to refine my studies by means of the technical-operational courses offered by the Italian Air Force.*

*Finally, I have to thank my parents for the enduring education they provide me: no matter how specialized I have become, they can always give me new and valuable lessons. I would not have reached this goal without them.*

*Rome, 26th September 2017*

*Jonas Piccinotti*

Table of Contents

# Introduction

Having good situational awareness on enemy forces has always been a major concern of any military operation. In the aeronautical and maritime contexts this is accomplished by means of several tools, among these, the radar technology definitely plays a key role. Radars can be ground based, airborne or mounted on vessels, but in any case the basic working principle stays the same. The probability for a radar to detect and track an enemy asset is function of several variables: among these variables, the Radar Cross Section (RCS) of the target is a fundamental parameter. The RCS can intuitively be regarded as the electromagnetic visibility of an object, or, in other words, the ability of this object to re-direct the electromagnetic power density that impinges on it towards the direction from which that power density came from (monostatic RCS case), or another specific direction of interest (bistatic RCS case).

Calculating the RCS of an object of interest, an aircraft for instance, may be accomplished either with an experimental approach (i.e. measurement campaign) or with a simulation approach (i.e. prediction algorithms). Both ways involve great complexity and difficulties, especially because the bands of interest range from the L band to the X band, covering an interval from roughly 1 to 12 GHz, and probably even lower and above, if one wants to be sure to include every radar threat.

The experimental approach would require to perform many measurements, illuminating and observing the target from a great combination of angles in order to obtain a good RCS characterization. Handling an aircraft and positioning it so that it can be illuminated from above and below may be extremely onerous. Both in the case of an open field test range or an anechoic chamber, the dimensions of the set-ups would be massive.

The simulation approach may nowadays appear more feasible, considering that very powerful computing machines are easily affordable. Several electromagnetic prediction tools are also commercially available and, in theory, once a good 3D model of the target is realized, the RCS may be calculated by "simply" numerically solving Maxwell's equations. Unfortunately though, due to the physical dimensions of the objects of interest (i.e. aircrafts and vessels reach hundreds or even thousands of wavelengths in size) and the frequencies of interest (i.e. roughly L-S-C-X bands), solving Maxwell's equations by means of a numeric method without any approximating assumption would require a massive computational power and memory storage. Workstations that can provide such resources exist, but are less easily available and affordable and, anyways, such brute force approach would definitely be not a smart way to proceed. In fact, despite certain computing servers allow solving problems which were not even approached in the past because of their computational burden, the focus of the electromagnetic community in such matters have lately been put on how to efficiently use the available resources so that, even with compact commercial workstations, repeated RCS simulations can be accomplished in reasonable times with a satisfying accuracy in order to perform an iterative optimization process when designing a new platform (e.g. optimization of a stealth planform [1] or antenna siting process).

Within this framework, the purpose of this PhD research work was to develop and optimize a fast numeric algorithm able to compute monostatic and bistatic RCS predictions obtaining an accuracy comparable to what commercially available from well-known electromagnetic CADs, but requiring unprecedented computational times. This was realized employing asymptotic approximated methods to solve the scattering problem, namely the Geometrical Optics (GO)

1 "Planform" is a term used in the aeronautical world to indicate the overall geometry of an aircraft, specifically used with reference to the "God's eye" view

and the Physical Optics (PO) theories, and exploiting advanced algorithmical concepts and cutting-edge computing technology to drastically speed-up the computation. The algorithm was composed of two main phases: the electromagnetic field transport from the source to the target, and the calculation of the scattered Far Field from the target by means of a radiation integral.

The First Chapter of this thesis focuses on an historical and operational overview of the concept of RCS, with specific reference to aeronautical and maritime platforms. How geometries and materials influence RCS is also described.

The Second Chapter is dedicated to the first phase of the algorithm: the EM field transport phase, where the GO theory is applied to implement the "ray tracing". In this Chapter the first advanced algorithmical concept which was adopted is described: the Bounding Volume Hierarchy (BVH) data structure. The BVH is a technique used to speed-up the process where a GO ray is tested against a meshgrid-discretized geometry to detect where the ray hit the surface of the target body. Two different BVH approaches and their combination are described and compared.

The Third Chapter is dedicated the second phase of the calculation: the radiation integral, based on the PO theory, and its numerical optimization. The Type-3 Non-Uniform Fast Fourier Transform (NUFFT) is the second advanced algorithmical tool that was used and it was indeed the foundation of the calculation of the radiation integral. Then, to improve the performance but also to make the application of the approach feasible in case of electrically large objects, the NUFFT was further optimized resorting to the "pruning" technique, which is a stratagem used to save memory and time by avoiding calculating points of the transformed domain that are not of interest. In the effort to contain the computational complexity, the "Domain Decomposition" technique, which is a way to partition the global problem into easier and smaller sub-problems before executing the NUFFT calculation, was

planned to be used as well, but in fact it will be implemented during future development.

The cutting-edge computing technology used to speed-up the algorithm was the parallelization of the code and its implementation on Graphical Processing Unit (GPU). In particular, the NUFFT routine required an extensive work to be parallelized and translated into the appropriate coding language. It was firstly coded using MatLab, a high-level intuitive programming environment which allowed a low-effort implementation in one and two dimensions. Then, the 1D and 2D MatLab NUFFTs were converted in C++, used as intermediate language before proceeding to the implementation in CUDA, which is the proprietary NVidia programming language used to realize parallelized GPU routines. Once all the 1D and 2D codes (MatLab, C++, and CUDA) gave the same exact results, a 3D version was realized using the same process until reaching the final product, namely a 3D CUDA code.

Finally, in order to validate the hybrid GO-PO algorithm, a short measurement campaign was held at the headquarter of the Ingegneria Dei Sistemi (IDS) Company, located in Pisa. The measurements, accomplished using a Synthetic Aperture Radar (SAR) imaging equipment set up on a planar scanner inside a semi-anechoic chamber, validated the algorithm predictions for canonical scatterers, such as the square plate, the cylinder, the sphere, and the corner reflector, all entirely made of aluminum.

# Chapter 1

# Radar Cross Section: a key parameter for the modern stealth military airborne assets

*Stealth* is defined as secret, clandestine, or surreptitious. Since the beginning of recorder warfare, armies have sought an advantage by the use of secret, clandestine, or surreptitious means (Alexander, 2004).

Starting with these broad premises, the purpose of the present chapter is to investigate the role of the stealth technology in military modern aviation, recalling the great improvements achieved in the past 30 years in this area by the world leading aeronautical companies. However, some relevant maritime platforms will be studied, as they constitute interesting examples of stealth mechanisms.

## 1.1. Framework of stealth technology: the Electronic Warfare

A frame of definitions is given in the following, to allow the correct placement of the stealth technology with respect to the vast variety of key technologies applied in the military scenario.

Information Warfare (IW) is the appellation applied to conducting warfare-like actions against an adversary's information systems or protecting one's own information systems from such activities. IW can be applied to both military e non-military scenarios. When applied to the military context it is usually called Command and Control Warfare (C2W) and it comprises five pillars: physical destruction of information

systems, psychological operations (PSYOPS), deception, operational security (OPSEC), Electronic Warfare (EW).

EW focuses on attacking information systems by withdrawing from, or imparting energy into enemy communication systems, so that the intended transport of information is either intercepted, denied, or both. EW is usually limited to the radio frequency (RF) part of the electromagnetic (EM) spectrum that starts about at 500 kHz and extends up to hundreds of GHz. However, the focus of this thesis work is primarily on the radar bands, shown in Figure 1, with specific interest on the S, C, and X bands.

| | |
|---|---|
| L | 1 – 2 GHz |
| S | 2 – 4 GHz |
| C | 4 – 8 GHz |
| X | 8 – 12 GHz |
| $K_u$ | 12 – 18 GHz |
| K | 18 – 26 GHz |
| $K_a$ | 26 – 40 GHz |

Figure 1: EM bands nomenclature as per the IEEE STD-521-2002

It is generally accepted that Electronic Warfare (EW) has three distinct components: Electronic Support (ES), Electronic Attack (EA), and Electronic Protect (EP). ES is comprised of those measures taken to collect information about an adversary by intercepting radiated emissions. EA refers to attempting to deny adversaries access to their information by radiating energy into their receivers. EP involves activities undertaken to prevent an adversary from successfully conducting ES or EA on friendly forces (Poisel, 2002).

Within the frame established by the previous definitions, the stealth technology, applied to different military platforms, can be considered as a form of EP, as well as the radar and infra-red (IR) passive decoys, commonly known with the terms "chaffs" and "flares", which are artifices purposely released by military platforms when being tracked by unfriendly radars or IR tracking devices.

A brief historical overview is provided in the following, to address the origins of stealth technology. Stealth is a broad concept which encompasses several characteristics (i.e. radar bands, IR bands, visible band, acoustic band) all pertaining the "detectability" of a platform, and, potentially, a concept that applies to very different combatant units, ranging from aircrafts, to vessels, to tanks and wheeled machines, and even to the single foot mobile soldiers.

## 1.2. Historical overview of stealth technology and concept of RCS

As the radar became an operational military tool by the end of the '40, it was already evident that warfare had entered a new era, in which the ability to see and hear enemy forces without being seen or heard, was as integral to the order of battle as tanks, planes, bombs, bullets, or troops. In 1942, with the ambitious plan of developing a low-drag long-range intercontinental bomber to strike the United States, the Reich scientists and engineers started developing an aircraft employing the "flying wing" airframe concept, which was only later discovered as an inherently stealthy profile. After two years and several prototypes, in early 1944 the Horten brothers came up with the Ho IX, also known as the Ho 229, which can be considered as the first stealth aircraft in history.

Despite the stealth features were not a real objective in the design of the aircraft, an astounding discovery was made during the first flight tests: the aircraft failed to show up accurately on radars. Once realized

such brilliant property, the Horten brothers, who were familiar with the experiments involving a carbon compound used by the German Navy to seal the hulls of their vessels, tested the compound on the fuselage of the Ho 229 obtaining promising results in terms of radar return: they had just employed a primitive form of Radar Absorbent Material (RAM).

Although a great discovery had just been made, some major drawbacks were apparent at the same time: the huge jet engines mounted on the Ho 229 featured significant compressor blades, which acted as good radar reflector. Additionally, the hot exhaust produced by the engines was easily detectable by infrared imaging devices, which were in experimental stages at that time, and later soon became an operational reality. These two key points address a fundamental characteristic of the stealth concept: to be properly covert, a military aircraft has to minimize its detectability in different bands of the electromagnetic spectrum. Returns originating from the radar band, from the infra-red (IR) band, and possibly from the visible band, have to be prevented, masked or disguised somehow.

Through the decades, techniques have been developed to mitigate the IR signature of airborne platforms: hot exhaust gases can be mixed with fresh air to cool down the resulting flow; additionally, the shape of the terminal section of the exhaust ducts plays a significant role in the IR visibility to thermal imagers. The visible band instead poses an intrinsic obstacle to stealthness: so called “low-visibility” paintings or camouflage patterns are largely used on almost every military platform.

However, nowadays, the optical visibility concern is becoming less and less important since the detection ranges of the cutting edge technology both in the radar and IR bands are drastically beyond the optical visual range. Although it constitutes a challenging scientific problem, discussing the IR signature of a modern military aircraft and the reduction techniques to achieve an IR stealth platform is outside the

scope of this work, which, as outlined in the preceding, will focus only on the radar band (Alexander, 2004).

## 1.3. Stealth technology applied to aeronautics: aircraft features affecting the RCS

The majority of single-seat and twin-seat strike-fighter aircrafts, which represent the category of main interest within the scope of this work, have Fire Control Radar installed on the nose of the aircraft, protected by a radome aerodynamically shaped and optimized. Great improvements have been achieved in the past 40 years in both the hardware and software characteristics of these complex airborne systems. However, operative frequencies appear to be bracketed in the X band [8-12 GHz], since usually this type of radars operates around the 10 GHz operative frequency.

A radar detects and tracks a target, and, if able to achieve high azimuth and range resolution, it can classify and even identify a target, thanks to algorithms elaborating distinct characteristics of the returns. It is, therefore, crucial in the design and operation of radars to be able to quantify or otherwise describe the echo, especially in terms of such target characteristics as size, shape, and orientation. For that purpose, the target is ascribed an effective area called the Radar Cross Section (RCS). Figure 2 shows the polar plots of the RCS measured at 1GHz for a WWII B-26 bomber aircraft, which has become a well-known example in literature dedicated to scattering problems, and of the simulated RCS at 300 MHz of a generic non-stealthy aircraft. In both cases, a complete 360-degree azimuthal scanning is studied.

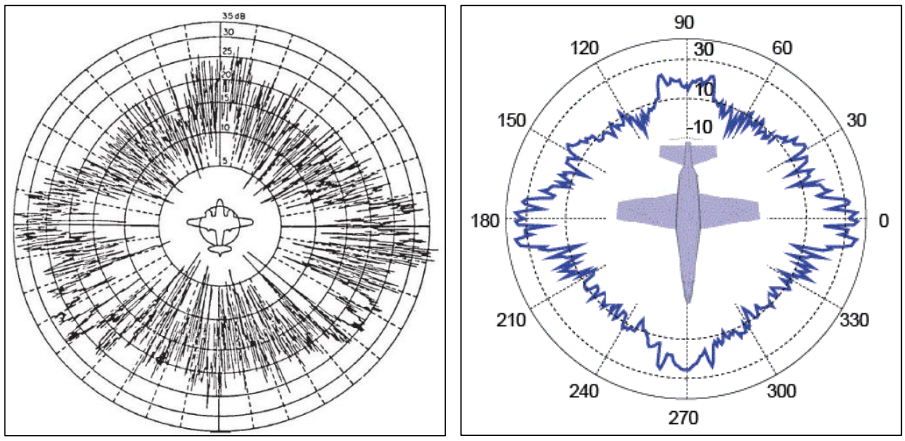



Figure 2: Measured (left) and simulated (right) RCS polar plots

An object exposed to an electromagnetic wave (usually referred as "illuminated" by the EM field) disperses incident energy in all directions. This phenomenon is called "scattering" and the object itself is called "scatterer". The energy scattered back to the source of the wave is called "backscattering" and constitutes the radar echo (also radar "return") of the object. The RCS is the projected area of a metal sphere that would return the same echo signal as the target, had the sphere been substituted for it. Unlike the echo of the sphere, however, which is independent from the direction of illumination, the echoes of all but the simplest targets vary significantly with the direction of illumination and the direction of observation, which may not be the same in a bistatic case. As will be shown later, this variation can be quite rapid, especially for targets many wavelengths in size.

The echo characteristics depend in strong measure on the size and nature of the target surface exposed to the radar beam. The variation is small for electrically small targets (targets less than a wavelength or so in size) because the wavelength of the incident field is too long to resolve target details. On the other hand, the flat, singly curved and doubly curved surfaces of electrically large targets all give rise to different echo characteristics. Reentrant structures such as the engine

intakes and exhausts generally have large echoes, and even the trailing edge of airfoils can be significant echo sources.

Before proceeding with a formal definition of the RCS and pointing out the reduction techniques relying on the RCS properties, an accounting of the main scattering mechanism affecting typical military platforms is provided in the following.

The overall radar "observability" of a typical small/medium-size fast-mover airborne asset (i.e. pure fighter, strike-fighter, UCAV [2], ICBM [3]) can be ascribed to seven basic scattering mechanism. Figure 3 schematically illustrates such mechanisms, portraying a simplified typical airborne platform, which can be representative of a conventional single-engine aft-swept wing aircraft or a missile with significant control surfaces.

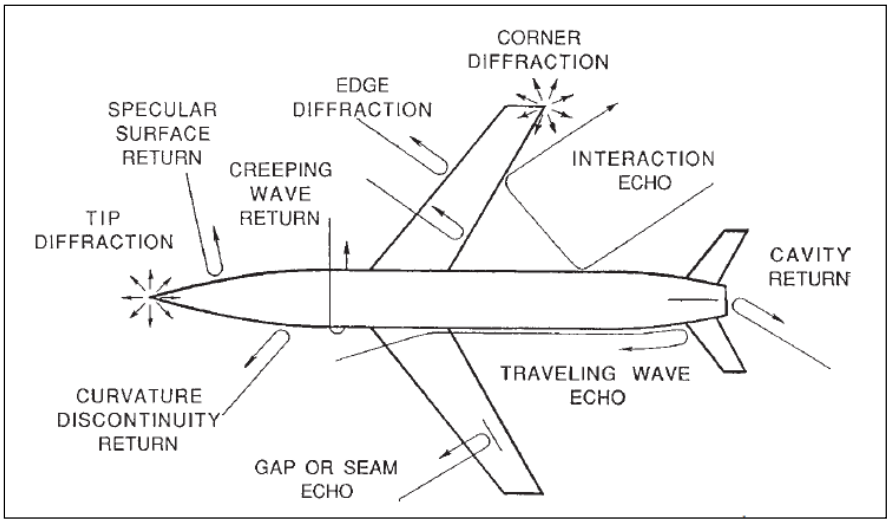



Figure 3: Schematization of the seven basic scattering mechanisms (Skolnik, 2008)

2 UCAV: Unmanned Combat Air Vehicle.

3 ICBM: Inter-Continental Ballistic Missile.

All the described phenomena greatly depend on the target aspect angle as seen from the illuminating radar. Some are dominant whereas other are weak. Additionally, not all the mechanism depicted in Figure 3 applies at the same manner to different kinds of typical military platforms, such as different aerial vehicles, warships or military ground vehicles, due to evident overall design differences. The seven mechanisms are briefly discussed in order of significance (Skolnik, 2008).

### 1.3.1. Cavities

Intakes (JEM), cockpit and exhaust. Despite Figure 3 only depicts a hypothetical exhaust duct at the rear of the aircraft (Figure 4), engine intake ducts (Figure 5 left) and cockpit cavity (Figure 7) constitute the main reentrant structures seen from a front side view, which operationally is the most relevant, since it's the aspect angle seen in long-range detection either when the detection is performed by enemy aerial platforms or by ground-based air-surveillance radars.

Figure 4: Exhaust ducts

The scattering caused by such cavities largely impacts on the overall RCS and tends to persist over aspect "Angles off Nose [4]" as wide as 60°. Most of the internal ducts surfaces (i.e. compressor stages and turbine faces) are metallic so that radar waves that enters such portions of the aircraft will likely find a way back out towards the radar. Additionally, the rotating compressors stages situated at the end of the inlet ducts (Figure 5 right) cause an interesting phenomenon exploited by radar algorithms to attempt the classification or even identification of enemy targets: the Jet Engine Modulation (JEM).

Figure 5: Engine intakes on a stealthy F-22A (left) and a visible jet engine compressor at the end of the inlet duct on a F-16 (right)

This phenomenon has been observed at angles as great as 60° AoN. Since the compressor and blade assembly are in rotational periodic motion with respect to the airframe of the target, they impart a periodic modulation on the signal scattered which has two significant consequences on the backscattered radiation: generation of noise in the received signal, clearly a drawback, and the generation of a radar custom signature that can be useful for target identification, a great advantage. To avoid or at least minimize these effects, specific geometries have been developed for the intake ducts. Figure 6 shows

4 "Angle off Nose" (AoN) is a common nomenclature in the aeronautical world used to indicate angles measured on the azimuthal plane starting from the longitudinal axis of the aircraft.

engine intakes and the internal inlet duct section reaching the compressors, as realized in a very low observable aircraft (Skolnik, 2008).

Figure 6: Inlet duct shaping aimed to mask the compressor

An analogous scattering effect is generated when the radiation enters the cockpit passing through the cover, called “canopy”, since the cockpit constitute is an actual cavity hosting the pilot(s). A countermeasure applied to limit the amount of radiated energy entering the canopy is the insertion of an indium-tin-oxide layer and a gold tint applied within the polycarbonate transparent material. Unmanned vehicles have obviously solved this specific source of scattering by completely removing the cockpit, usually replacing it with smooth homogenous surfaces and a central, stealthy, single engine intake. The UCAV “nEUROn”, realized by an all-European six-nation joint venture, is a great example of this geometry and is shown in a following paragraph. The UCAV was developed as a technology demonstrator for low observability technologies both in radar and IR bandwidths and it was the object of a test campaign executed in collaboration with the Italian Air Force during the second year of the present PhD work (Marchetto, Mercurio, Migliozzi, Piccinotti, & Risoldi, 2015). More details will follow. Figure 7 depicts a stealthy “golden” canopy.

Figure 7: Canopy metallization

### 1.3.2. Specular scatterers

A specular scatterer is any target surface oriented perpendicular to the line of sight of the radar and so perpendicular to the impinging radiation. Flat surfaces offer particularly large returns in the specular direction, but the return intensity decreases significantly away from the perpendicular direction. Instead, the returns coming from singly and doubly curved surfaces (i.e. cylindrical and spheroidal surfaces) are lower than those from flat surface, but are more persistent with changes of aspect angle from the perpendicular direction, as schematized in Figure 8.

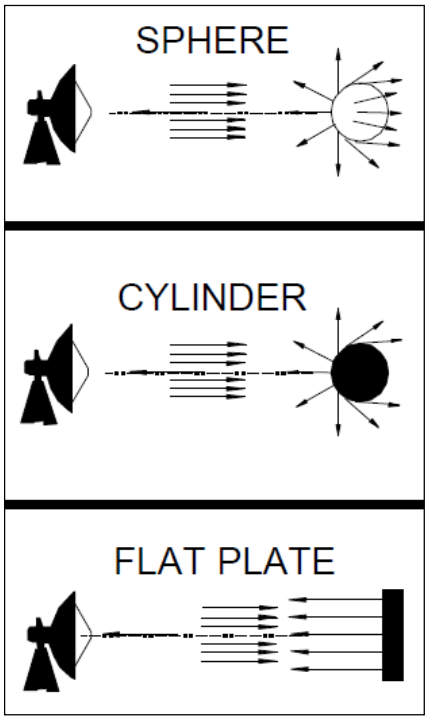



Figure 8: Simplified scattering mechanism from a plate, a cylinder and a sphere

The detrimental effects of this kind of reflections have to be addressed also for those surfaces which just temporarily act as scattering sources, as, for instance, the panels sealing the weapons bay in the modern stealth fighters. Figure 9 shows the F-22 fighter weapon bay in the open configuration, where an angled solution was adopted for such panels, as mitigation for these specular scatterers.

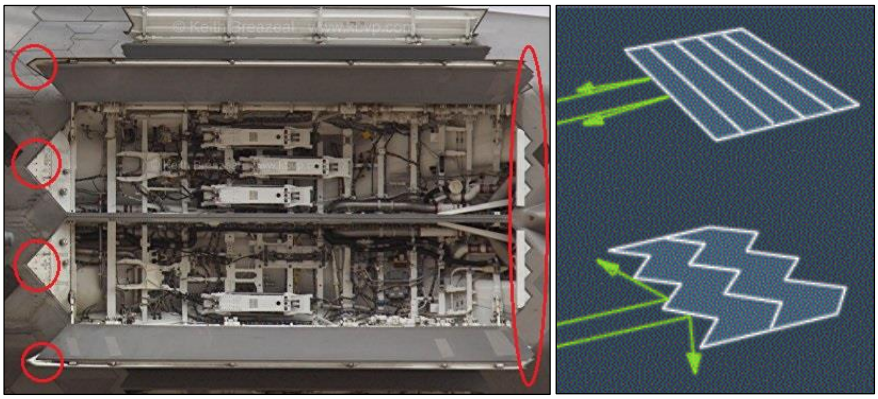


Figure 9: F-22 weapon bay featuring angle-shaped panels

### 1.3.3. Tips, edges, and corners diffraction

Scattering from tips, edges, and corners is less significant than specular returns, so this level of detail is addressed only when the main scattering source have been suppressed (i.e. overall shape of the airframe). The echoes from tips and corners are localized and tend to increase with the square of the wavelength and not with the size of any surface. Thus, they become progressively less important as the radar frequency rises.

Figure 10 shows several examples of these type of reflectors: four IFF [5] antennas just in front of the F-16 canopy (left), canards control surfaces on the sides of the nose of the Eurofighter Typhoon (right), pylon used to carry stores on the underside of the wings (right), Pitot tubes found on every aerial platform (left and right).

5 IFF: Identification Friend or Foe. A transponder system used on civilian and military aircraft to transmit ownship data making oneself visible to other aircrafts flying in the vicinity and to Air Traffic Control services.

Figure 10: Examples of tips, edges and corners on a F-16 (left) and A on Typhoon (right)

### 1.3.4. Surface discontinuities

Most airframes have slots and gaps all along the external skin used as small intakes for cooling of avionics bays or heat exchangers (Figure 11).

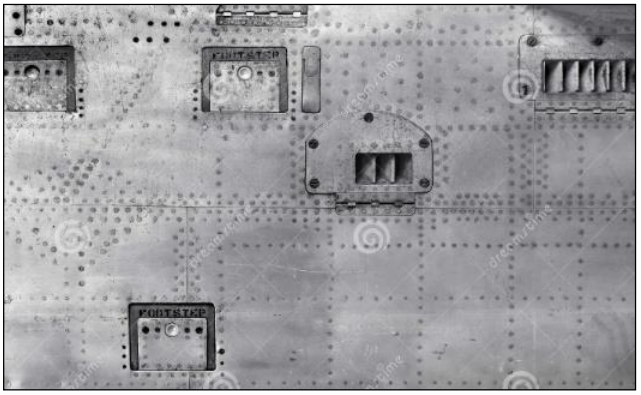


Figure 11: slots, latches and rivet heads on the external skin of an old generation military aircraft

Additionally, where control surfaces (e.g. ailerons, stabilizers, canards, flaps) meet the stationary airframe. Slots, gaps, and even rivet heads can reflect detectable amount of energy back to the illuminating

radar. To minimize this type of returns, modern stealth aircrafts use extremely smooth surfaces, avoiding discontinuities and resorting to specific dielectric compounds and sealants to fill in gaps and slots and to treat those portions of the external skin where an imperfection could result in a scattering hot spot. A great example of this modern treatments is observable on the skin of the nEUROn UCAV depicted in Figure 12.

Figure 12: Incredibly smooth surface of the UCAV “nEUROn” treated with sealants and compounds

However, these scattering sources are secondary if compared to the one described previously so that it is not easy to isolate and characterize them.

### 1.3.5. Interactions

Relatively strong returns can occur when a pair of target surfaces are oriented such that the impinging waves can bounce from one surface to another and then back to the radar, as in the interactions between the fuselage and the trailing edge of the right wing shown in Figure 3. Similar interactions occur for ship targets when bulkheads, railings, masts, and other topside features become mirrored in the mean sea

surface as shown in the following discussing of military vessels (Skolnik, 2008).

## 1.4. RCS formal definition

An object exposed to an electromagnetic wave (usually referred as “illuminated” by the EM field) disperses incident energy in all directions ($4\pi$ steradians). This phenomenon is called “scattering” and the object itself is called “scatterer”. The energy scattered back to the source of the wave is called “backscattering” and constitutes the radar echo (also radar “return”) of the object: this specific condition of the direction of illumination being the same direction of “observation” is called mono-static case. The term “radar signature” instead, is usually referred to how the radar echo behaves over a meaningful bandwidth. The intensity of the echo is described by the Radar Cross Section of the object, hence the acronym RCS. Early papers on the subject called it the “echo area” or the “effective area”, terms still found occasionally in contemporary literature (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004). An intuitive definition of RCS may be derived referring to Figure 13.

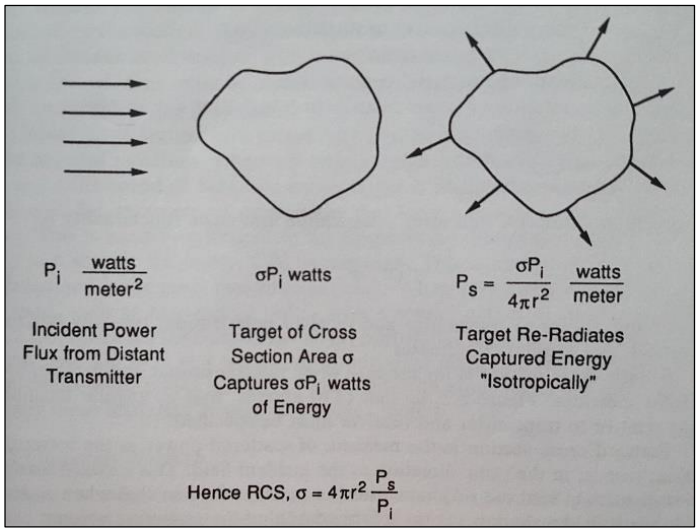



Figure 13: Intuitive definition of RCS based on power densities (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

Let the incident power density at scattering target from a distant radar be $P_i$ Watts/m$^2$. The amount of power intercepted by the target is then related to its cross section $\sigma$, with units of area, so that the intercepted power is $\sigma P_i$ Watts. The intercepted power is then either re-radiated as scattered power or absorbed and dissipated as heat per Joule Effect for instance. Assume for now that it is all uniformly reradiated as scattered power in all $4\pi$ sr of space so that the scattered power density $P_s$ in Watts/m$^2$ is given by

$$P_s = \frac{\sigma P_i}{4\pi r^2}$$

Equation 1

and solving for $\sigma$ at a distance $r$ such that the Far Field conditions are verified, we get:

$$\sigma = 4\pi r^2 \frac{P_i}{P_s}$$

Equation 2

which shows how RCS is fundamentally a ratio of scattered power density to incident power density. Since the power density of an EM wave is proportional to the square of electric or magnetic field, an alternative definition can be derived.

In fact, the IEEE [6] formal definition of RCS is (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004):

$$\sigma = \lim_{r \to +\infty} 4\pi r^2 \frac{\left|\underline{E_s}\right|^2}{\left|\underline{E_i}\right|^2}$$

Equation 3

where, referencing Figure 14 which portraits a generic body $M$ centered on the origin of the axis $O$, $\underline{E_i}$ is the electric field of the incident wave impinging on the target, $\underline{E_s}$ is the electric field of the scattered wave, and $\underline{r}$ is the vector which determines the point where the scattered field is observed and measured. It is important to note that Equation 3 contains the vector total electric fields, so that both the vertical and horizontal components are considered when calculating the square module. An equally valid definition results when the electric field magnitudes in Equation 3 are replaced with the incident and scattered magnetic field magnitudes.

6 IEEE: Institute of Electrical and Electronics Engineers.

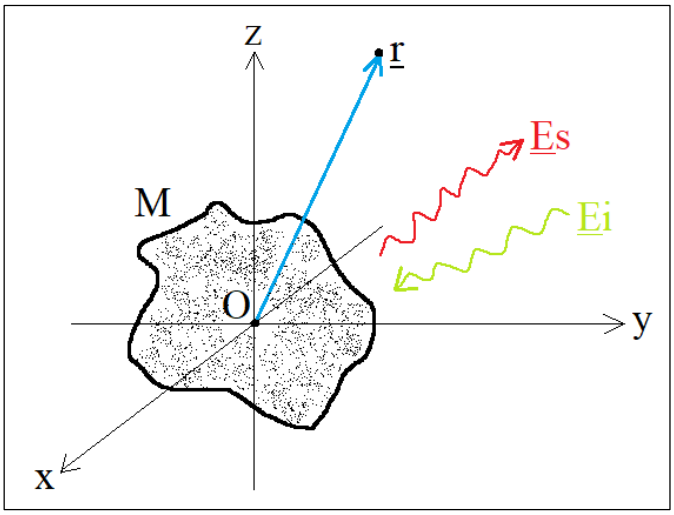



Figure 14: Scattering scenario

RCS is normalized to the magnitude of the incident wave at the target so that it does not depend on the distance of the target from the illumination source. RCS is also normalized so that inverse square fall-off of scattered intensity due to spherical spreading is not a factor. Therefore, the position of the receiver can be unknown. However, the limiting process in Equation 3 is more an academic refinement. In both measurement and analysis, the radar receiver and transmitter are usually located in the Far Field of the target, being at distances from tens to thousands of meters from it. Figure 15 portraits an indoor setup inside an anechoic chamber which has usually a small volume, except for the aeronautical framework where it can have sides measuring several tens of meters. Such distances, in combination with specific reflectors that create plane waves, are large enough to be acceptably considered at infinity at the frequencies of interest. Additionally, the scattered field square amplitude $\left|\underline{E_s}\right|^2$ decays inversely with $r^2$, resulting in an implicit $r^2$ term in the denominator that cancels the explicit $r^2$ in the numerator of Equation 3. Therefore, the dependence of the RCS on $r$ and the need to form the limit usually disappears.

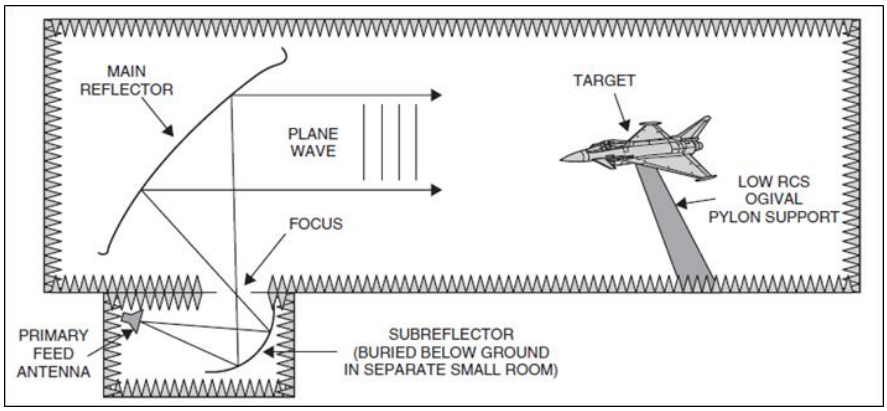



Figure 15: Schematization of indoor RCS advanced measurement set up for real targets

Figure 16 shows a part of one of the remarkable configurations for outdoor RCS measurements used in the early stages of the F-117 program, when scaled models were used for preliminary studies. Other more advanced configurations, both indoor and outdoor, exist; however, they are highly classified.

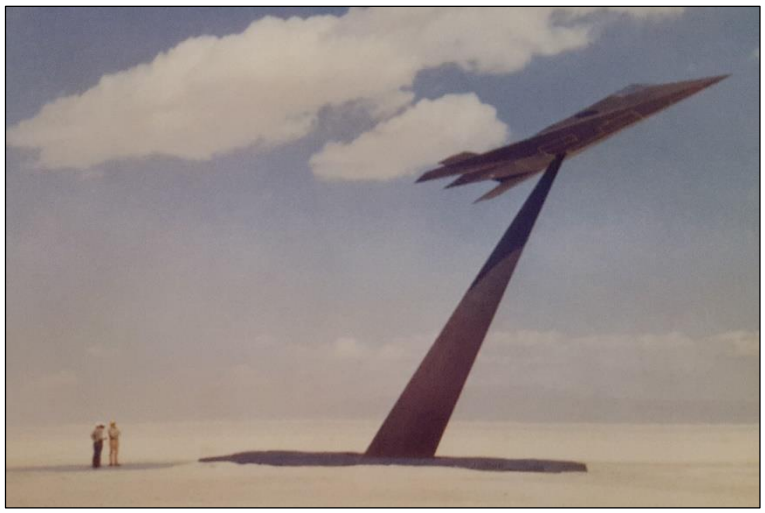


Figure 16: Outdoor RCS advanced measurement set up for scaled and real targets (Sweetman, 2001)

Considering a general case where the transmitter (Tx) is not necessarily co-located with the receiver (Rx), namely a "bistatic configuration", the RCS will be a function of (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004):

- position of Tx relative to target (expressed in terms of angles);
- position of Rx relative to target (expressed in terms of angles);
- target geometry (shape);
- target material composition;
- angular orientation of target relative to Tx (Tx aspect angle);
- angular orientation of target relative to Rx (Rx aspect angle);
- frequency;
- Tx polarization;
- Rx polarization.

The unit for RCS is area, usually in square meters, or may be nondimensionalized by dividing by $\lambda^2$ or expressed in $dB$ with respect to $1m^2$ with the equivalent symbols $dB_{sm}$ or $dB_{m^2}$.

### 1.4.1. Bistatic situation

It is often necessary to measure or calculate the power scattered in another direction than back to the transmitter: this is called a "bistatic" situation. A bistatic RCS may be defined for this case as well as for backscattering, provided it is understood that the distance R is measured from the target to the receiver. A "bistatic angle" is also defined as shown in Figure 17.

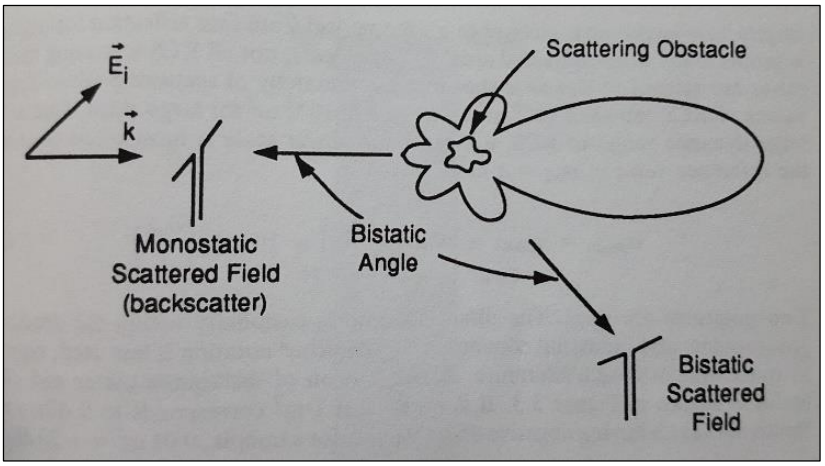



Figure 17: Bistatic situation
(Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

From an operational standpoint, bistatic, or, more in general, multi-static configurations, are of great interest when it comes to oppose a counter-measure to stealth technology. Since an airborne platform can be stealthy optimized just in certain aspect angles (usually ±50° AoN), due to necessary trade-offs related to its aerodynamics, illuminating it from the abeam or possibly from behind, will likely result in much greater radar return than from in front of it, where it is most stealthy. Modern deployable surveillance radar can be set up in networks in which every single radar constitutes a node covering a certain portion of the air space. Combining several radars in strategic key position may result in an effective multi-static configuration able to detect and track stealth platforms.

### 1.4.2. Accounting for polarization

The definition given in Equation 3 does not account for the polarization of the receiver, thus, a more precise expression would be (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004):

$$\sigma = \lim_{r \to +\infty} 4\pi r^2 \frac{\left|\hat{e}_s \cdot \underline{E_s}\right|^2}{\left|\hat{e}_i \cdot \underline{E_i}\right|^2}$$

Equation 4

where:
$\hat{e}_s$ is a unit vector aligned along the electric polarization of the receiver;
$\hat{e}_i$ is a unit vector aligned along the electric polarization of the transmitter;
$\underline{E_s}$ is the vector electric scattered field;
$\underline{E_i}$ is the vector electric impinging field.
The polarizations of both the receiver and the source of illumination may be arbitrary, but all the possible combinations can be characterized by means of a "polarization scattering matrix" $\underline{\underline{\sigma}}$.

The general notation to indicate the mentioned functionality from polarization and angle is

$$\sigma^{tr} = f(\theta^t, \varphi^t, \theta^r, \varphi^r)$$

Equation 5

where $t$ and $r$ refer to Tx and Rx polarization, typically vertical and horizontal and $\theta$ and $\varphi$ are the angular coordinates.

Figure 18 depicts the effects of the polarization as seen on basic geometries: (a) depicts the backscatter from a triangular cylinder illuminated with a parallel polarization plane waves, (b) with a perpendicular polarization, (c) with a circular polarization, and lastly (d) shows the same phenomenon for a diamond-shaped cylinder and parallel polarization.

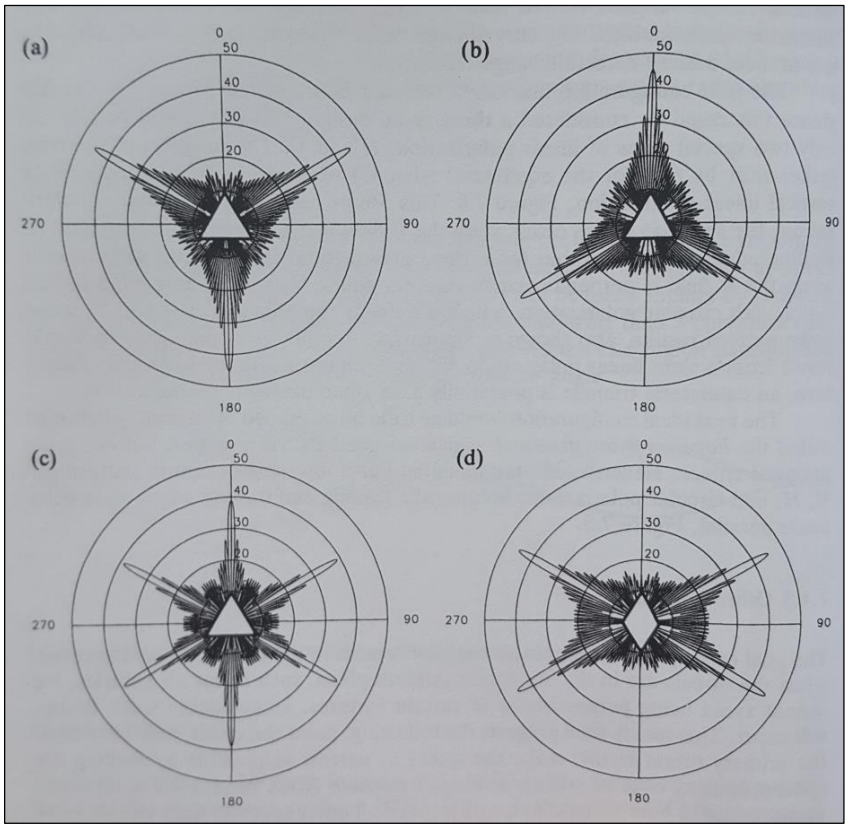



Figure 18: Backscatter from basic geometry reflectors
(Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

### 1.4.3. Accounting for phase

Because the radar cross section is a real number formed by squaring the amplitude of a complex number representing the scatted electric field, all phase information is ignored. To preserve phase information, it is often convenient to deal with the complex scattered fields themselves.

The phase relation can be addressed by assuming that even if $\sigma$ may be a pure real number, we may extract its root $\sqrt{\sigma}$ just in a formal notation as a complex number in which the phase information is

implicitly retained. When that is the case, the following expression is used (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004):

$$\sqrt{\sigma} = \lim_{r \to +\infty} 2\sqrt{\pi} r \frac{\hat{e}_s \cdot \underline{E_s}}{\hat{e}_i \cdot \underline{E_i}}$$

Equation 6

## 1.5. RCS reduction techniques

To better understand the role $\sigma$ plays in radar detection performances, let us recall the simplest form of the radar range equation which ignores number of effects that can be critical in detailed calculation but it is a valuable tool to roughly assess expected changes in radar performance for a given RCS change. The detection range is given by Equation 7 (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004) which relies on two main assumption, which are:

1) The transmitting antenna gain $G_{TX}$ is equal to the receiving antenna gain $G_{RX}$;
2) The distance between the transmitter and the target is equal to the distance between the target and the receiver

so that:

$$R_{max} = \sqrt[4]{\frac{P_{TX} G^2 \lambda^2 \sigma}{(4\pi)^3 P_{min}}}$$

Equation 7

where:

$P_{TX}$ is the transmitting power output of the radar in Watts;
$G$ is the peak gain of the radar antenna (equal for Tx and Rx);
$\lambda$ is the wavelength;
$\sigma$ is the RCS of the target;

$P_{min}$ is the required minimum input power to the receiver to achieve detection (based on the required probability of detection $P_d$).

Equation 7 shows that the maximum detection range varies with the fourth root of the RCS. Therefore, to halve the maximum detection range, a factor of 16 reduction in RCS will be required. To decimate (reduce of 10 times) the maximum detection range, a factor of 10000 reduction (-40 dB) will be required.

If no signal were competing with the target return, additional amplification could be added in the radar receiver to provide a detectable output, no matter how small the input. However, in a real-world scenario, we would find: cosmic and atmospheric noise (which may be neglected at L band and above), terrain backscatter (land and sea clutter), atmospheric clutter (backscatter from dust, refractive changes, and products of condensation or deposition of atmospheric water vapor), unintentional radio frequency interference from other emitters, and electronic countermeasure (jammers, chaffs, active decoys, etc…). Thus, the aim of RCS reduction is to make the RCS of the intended platform small enough to be "buried" within the noise level, so that a radar would not be able to discern the return from the noise floor, making the platform actually "stealth in the noise" (Skolnik, 2008).

In the following, the only four existing RCS reduction techniques will be briefly addressed and listed in order of decreasing practicability. They are: shaping, radar absorbing materials (RAM), passive cancellation, and active cancellation. Each method set trade-offs between advantaged and disadvantages. Typically, in current RCS designs, shaping is first employed to create a planform [7] design with inherently low RCS in the primary threat sectors (the abovementioned

7 As mentioned, in aeronautical design, "planform" or "plan view" is a vertical orthographic projection of an object on a horizontal plane, like a map. Similarly, in aviation, a planform is the shape and layout of an airplane's wing.

±~50° AoN) and then RAM are used to treat those areas whose shape could not be optimized, or to reduce the effects of creeping waves or travelling waves on the signature (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004).

### 1.5.1. Shaping

The objective of shaping is to orient the surfaces and the edges of the platform so that they deflect the scattered energy in directions away from the original direction, which corresponds to the radar one. Such optimized orientation cannot be achieved for all viewing angles within the entire sphere of $4\pi$ sr solid angle because, by necessary constraints imposed by the nature of the aircraft or the vessel, there will be aspect angles for which the surfaces are seen with normal incidence, resulting in strong radar returns along the directions characterized by those aspect angles. The applicability, and thus the success of the shaping technique, relies on the existence of angular sectors for which low RCS is less important than for others, so that aerodynamical trade-offs can be achieved.

Note that in design for reduced RCS, the emphasis has been focused almost exclusively on monostatic radars, implying an assumption that has proved to be a legitimate one so far. However, as introduced before, the future of military air surveillance seems to be aiming for multistatic configuration (possibly quickly deployable) employing several radar stations communicating via dedicated data-links.

Typically, as shown in Figure 19, a forward cone of about ±45° azimuth times about ±20° elevation, is of primary interest for RCS reduction, hence, large return affecting the RCS are shifted out of the forward sector and toward broadside.

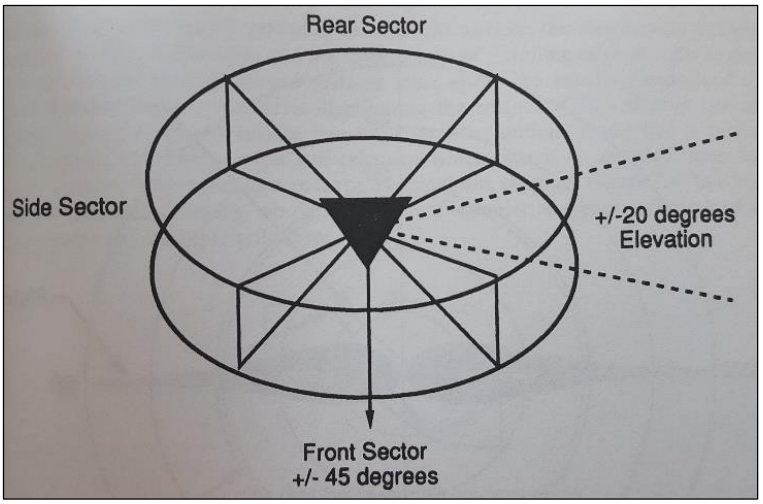



Figure 19: Front sector is the primary sector interested
by RCS minimization (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

For aircrafts, shifting strong returns to the abeam (i.e. about ±90° AoN) is accomplished by sweeping airfoils (i.e. wings) back at sharp angles. Figure 20 shows the comparison between two very similar basic shapes, observed under the aspect angle indicated by the arrows, when backswept edges are introduced in the middle section of the object instead of perpendicular [8] edges. The RCS polar plots let the reader understand how the right shape is much more detectable due to that strong backscatter, resulting in that strong spike oriented along the vertical direction on the right polar plot.

8 "Perpendicular" and "backswept" are referred to the aspect angle as seen from an illuminating radar located just in front of the target, namely along its longitudinal axis, as indicated by arrows in Figure 20Figure 20.

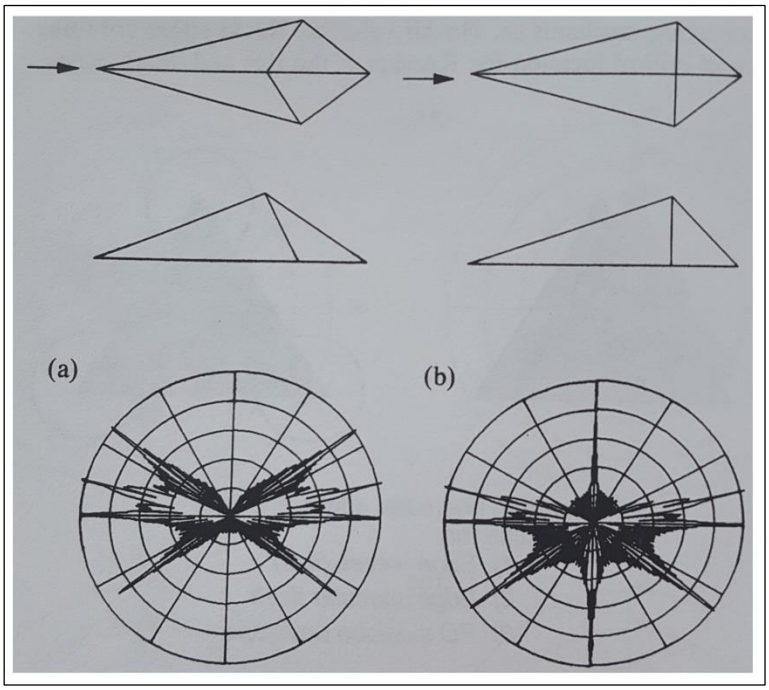



Figure 20: Effect of backswept edges on the RCS
of a basic geometry (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

Early designs obtained astounding stealth characteristics at the cost of degraded aerodynamics, and stability and controls performances, due to the faceted geometries employed, results of the limited processing capabilities of early CAD software, which were not able to properly manage significantly large doubly curved surfaces. Figure 21 shows two simulations, at 1 and 10 GHz respectively, run using a model of the stealthy B-2 strategic bomber. The fundamental shaping technique (i.e. backswept wings) was successfully employed to obtain very low RCS value in a ±30 degree azimuthal sector. These polar plots also show how the shaping technique is robust against significantly large frequency changes.

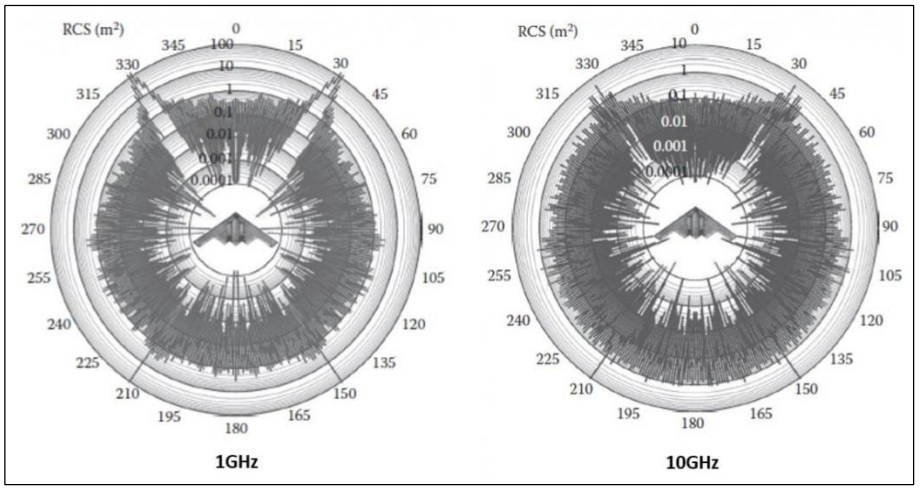



Figure 21: B-2 RCS polar plots at 1 and 10 GHz

Figure 22 compares two radically different designs: the first one based on doubly curved surfaces and the second one characterized by facets properly oriented for RCS reduction purposes. Due to data classification reasons, the value of the RCS is not available on the polar plots, however, it is evident how the first design RCS is characterized by spikes much larger than the baseline value, whereas the faceted cases results in a roughly constant RCS value without significant variants. In both cases, anyway, the RCS minima are oriented towards the positive direction of the longitudinal axis, in the abovementioned front sector of interest.

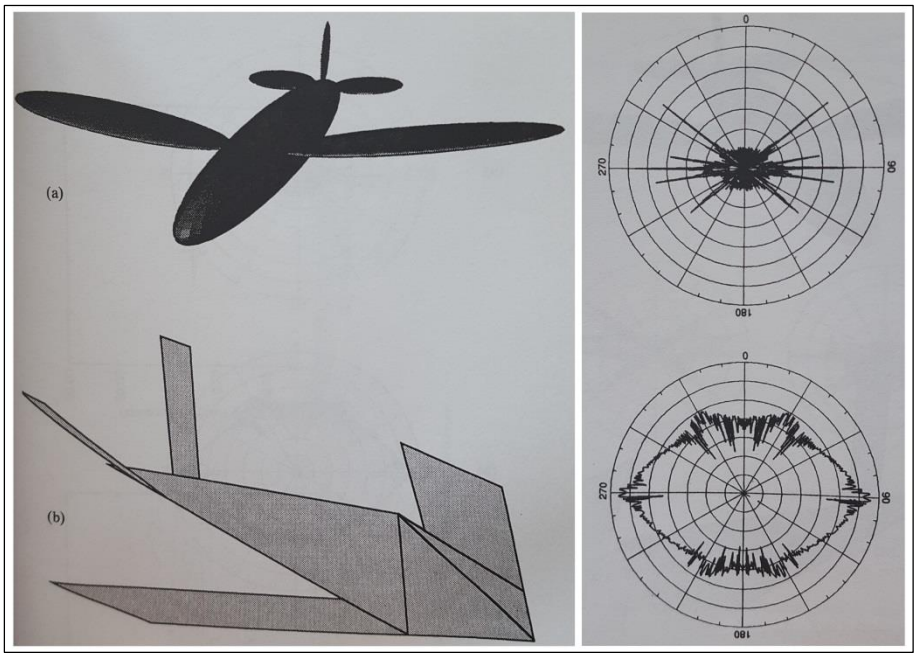



Figure 22: Comparison between a designed characterized by doubly curved surfaces design Vs a faceted design (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004)

Focusing just on the elevation plane, if the target is hardly seen from above (e.g. high-altitude deep strike bombers), echo sources such as engine intakes can be placed on the top side of the target where they can be hidden by the forward portion of the body when viewed from below, exactly as shown in Figure 23.

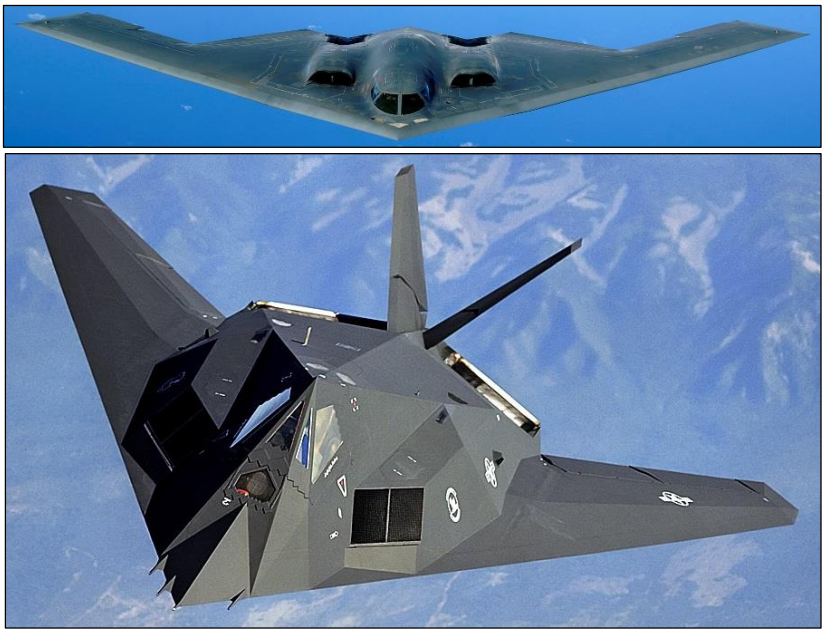


Figure 23: B-2 (top) and F-117 (bottom) feature
Distinct engine intakes on the top side

Similarly, for a low flyer (e.g. low-level deep strike bombers) whose major treats might be look-down radars, engine inlets will be placed on the underside of the fuselage (Figure 24).

Figure 24: B-1 features engine intakes on the bottom side (not visible in the picture)

Ships and ground vehicles are inherently more "boxy" and characterized by several dihedral and trihedral corners and "top hats" (circular cylinders with the axis perpendicular to a flat plate on top) as schematized in Figure 25. Returns generated by these scatterers are the major contributors to the overall RCS, and those can be avoided by bringing intersecting surfaces together at acute or obtuse angles.

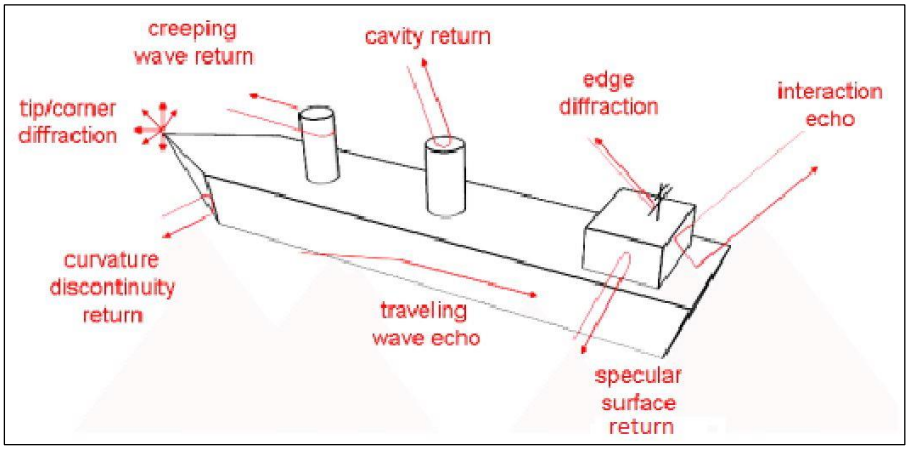



Figure 25: Sources of EM scattering in a generic vessel

Because of the presence of the sea surface, vertical bulkheads, masts and ships form efficient dihedral corners. The associated scattering effect can be reduced by tilting the bulkhead away from the vertical. Two examples of operational modern military vessels applying the same evident shaping techniques, despite a significant difference in size, are depicted in Figure 26.


Figure 26: Class Visby corvette (top) and
a Class Zumwalt destroyer employ shaping (bottom)

Since shaping is a major geometry feature, applying it to existing vessel is unpracticable, therefore it has to be accounted for since initial design. Although, even in the case of a new project, the amount of bulkhead tilt is a trade-off between RCS reduction, internal useful volumes and costs. Figure 27 shows an extreme example of shaping application: the experimental stealth ship “Sea Shadow” built by Lockheed for the United States Navy to determine how a low radar profile might be achieved and also to test high stability hull configurations.

Figure 27: Sea Shadow experimental stealth ship

## 1.5.2. Radar Absorbing Materials

If low RCS values are successfully obtained by means of shaping, the material treatment goals are then to reduce non-specular return from tips, edges and to reduce surface traveling, edge, and creeping waves. Specular RAM is reserved for those regions where shaping alone is not sufficient, such as cavities and edges viewed in a specular direction (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004).

As the name implies, RAM employ the absorption mechanism to reduce the backscattered energy. Several types of loss are exploited, which involve the dielectric and/or magnetic properties of the material. Physically, the loss is an actual conversion of energy into heat; however, most absorbers do not dissipate enough energy to become detectable by IR seeking devices. Two main categories of RAM are defined: the first one working on impedance matching principle (e.g. pyramidal and tapered absorber as the ones used in anechoic chambers) and the second one exploiting resonant absorbers. For obvious

mechanical and aerodynamical considerations, only the second type of RAM is used on military platforms.

Substances answering to specific requirements can be artificially fabricated whose indices of refraction are complex numbers. The index of refraction is responsible for magnetic and electric effects and its imaginary part accounts for losses. At microwave frequencies, the loss is also due to the finite conductivity of the material, but typically the effects of all loss mechanisms are recapped into the electrical permittivity ($\varepsilon$) and magnetic permeability ($\mu$) of the material because a cumulative effect is of more practical use (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004).

Carbon was the basic material used in the fabrication of early absorbers because of its non-optimal conductivity. Many commercial carbon-based materials now being marketed have designs that have not changed substantially for more than 20 years, but these materials are not easily applied to operational military platforms because of their bulky and fragile structure not adapt to operational environments. Magnetic absorbers are widely used for operational systems instead. Their loss mechanism is primarily due to a magnetic dipole moment whose basic component are iron compounds, such as carbonyl iron and oxides of iron (ferrites). Magnetic materials are typically a fraction of the thickness of dielectric absorbers, however, they are heavy (e.g. MX-410 experimental aircraft in 1945 could not fly due to coating weight) because of their iron content and are inherently more narrowband than their dielectric counterparts. The basic lossy material is usually embedded in a matrix or binder such that the composite structure has the EM characteristics appropriate to a given range of frequencies. Additionally, external layers of material transparent to EM radiation are applied for structural purposes.

Same examples of these materials are: aluminum oxide fibers, aramid fibers, boron, carbon-carbon composite, carbon fibers, ceramics, Fibaloy, Jaumann absorber, fiberglass epoxy resin

composites, Kevlar 49, Silag, Silicon-Carbide fibers, Spectra-100, Super-plastics, and thermoset composites. Examples of stealth paints are "Iron Ball", based on ferrite, and applied to the SR-71 (Figure 28) and the Retinyl Schiff Base Salt, a non-ferrous type of coating. Base salts are polymers that contain double-bonded carbon-nitrogen structures linking divalent groups in the linear backbone of the molecule's polyene chain (Jones & Thurber, 1989).

Figure 28: SR-71 "Black Bird" characterized by its famous black color, due to the ferrite based stealthy paint

### 1.5.3. Passive cancellation

Passive cancellation, intended as impedance loading, received a great deal of attention in the 1960s, but the method demonstrated to be severely limited. The basic concept is to introduce an echo source whose amplitude and phase can be adjusted to cancel another echo source by means of destructive interference. This can actually be accomplished for relatively simple objects, provided that a loading point can be identified on the body. An aperture can be machined in the

body, and the size and shape of the interior cavity can be designed to generate the required optimum impedance at the aperture.

Unfortunately, even for simple bodies, it is extremely difficult to generate the required frequency dependence for this built-in type of impedance, so that the RCS reduction obtained for one frequency rapidly disappears as the frequency changes (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004).

Furthermore, typical military platforms such as aircrafts and vessels are hundreds or even thousands of wavelengths in size and have tens of scattering hot spots, which makes unpractical to devise a passive cancellation treatment for each of these echo sources. In addition, the cancellation effect (i.e. destructive interference) can turn into a reinforcing effect (i.e. constructive interference) with a small change in frequency or viewing angle. For these reasons, passive cancellation has been discarded as a useful RCS reduction technique.

Within the passive cancellation techniques, another method is worth being mentioned: plasma. Sometimes addressed as active (since it involves the "production" of a plasma layer/cloud wrapping the object), in this chapter it is addressed as a passive method, since no EM field is produced to achieve RCS reduction.

A collisional unmagnetized plasma is a "quasi-neutral" (i.e. total electrical charge close to zero) mix of ions (i.e. atoms which have been ionized, and therefore possess a net positive charge), electrons, and neutral particles (i.e. un-ionized atoms or molecules) and has a complex dielectric constant, which makes it suitable to absorb EM energy over a wide range of frequencies.

For a plasma to act as an efficient absorber over a wide range of frequencies, without significant reflection of the incident signal, three conditions must be satisfied by both the electron density level and its spatial variation: firstly the electron density should be sufficiently high near the target whose RCS is sought to be reduced; secondly, the density should falloff with increasing distance from the target, and thirdly, the

electron-ion and electron-neutral collision rates should be sufficiently high to result in significant wave absorption.

Plasma employed as stealth technique was firstly developed by Russia by placing a plasma torch on the nose of an aircraft. The torch created a ionized cloud around the aircraft, which absorbed the incident waves. In fact, when an EM wave enters a weakly-ionized plasma, it is subjected to absorption as well as scattering. Absorption arises from loss of energy of the wave due to energy transfer to charged particles, and subsequently to neutral particles (atoms and molecules) by elastic and inelastic collisions. Therefore, the generated plasma "shield" partially consumes radar energy and also caused it to bend around the aircraft, thereby reducing the RCS. Wave scattering is due to spatial variation of the refractive index, such as during the transition from free space to a plasma, as well as density variation within the plasma (Hema, Simy, & Mohan, 2016) (Chaudhury & Chaturvedi, 2005).

### 1.5.4. Active cancellation

Also known as "active loading", active cancellation is even more ambitious than passive loading. In essence, the target must emit radiation in time coincidence with the incoming waves so to exploit the destructive interference between incident and reflected radiation obtaining an overall null EM field.

This implies that the target must be "smart" enough to sense the angle of arrival, intensity, frequency, polarization and waveform of the incident wave. It must also know its own echo characteristics for that particular wavelength and angle of arrival rapidly enough to generate the proper waveform at the required frequency. Such a system must also be versatile enough to adjust and radiate a pulse of the proper amplitude and phase with smaller aspect changes and where scattering patterns become more complex.

Due to the strict requirements, active cancellation appears to be unsuitable for practical application in the radar bands of interest. It may be employed for low-frequency RCS reduction application, where the use of absorbers and shaping become very difficult and scattering patterns exhibit broader lobes. Research on the technique is likely to continue because other practical means of RCS reduction are also difficult to apply for low frequencies.

Ultimately, consideration should be given to the fact that incorrect application of an active technique like this would turn into a beacon like type of source, making the platform even more detectable to enemy's surveillance radars or even passive ESM systems (Knott, Shaffler, & Tuley, Radar Cross Section, 2nd ed, 2004).

### 1.5.5. Penalties of RCS reduction

Typically, the requirements for reduced RCS echo conflict with conventional or traditional requirements for structures, in terms of aerodynamics of the overall geometry and specific geometry related to the engines, namely inlets and exhaust ducts and nozzles. As a result, the final system design is a compromise that inevitably increases the cost of the overall system, from initial engineering through production.

However, the high cost is only one penalty of RCS reduction, others are: reduced payload (i.e. the stores must be carried internally as much as possible), reduced range (i.e. if internal volumes are used for stores, they cannot be used for fuel), added weight (i.e. dense stealth coatings), increased maintenance (i.e. fragility of the stealth treatments). Not surprisingly, RCS reduction cannot always be justified, at least in terms of improved detection ranges.

Ultimately, despite how important it is, the radar signature is just one of the many specifications defining the desired performances of a platform. Trade-offs will always have to be made with respect to a large number of operational characteristics.

## 1.6. Peculiar sources of scattering and peculiar solutions for aeronautical military applications

In addition to the classical scattering sources and the respective design expedients described in Paragraph 1.5, other challenging aspects have to be taken into account.

### 1.6.1. Ownship radar and Frequency Selective Surfaces (FSS)

The majority of military aircraft uses airborne radars, typically installed in the nose cone, to detect and track enemy's platforms, taking advantage of the scattering features described so far. The radome cone must permit the EM waves to radiate from within and, necessarily, it will permit radiation in the opposite direction, namely from the outside of the cone, coming inside towards the radar antenna.

However, the radar antenna dish itself is a highly-optimized reflecting surface, paradoxically making the radar a major scattering hot spot besides an incredibly useful tool for detection and tracking. Additionally, old generation mechanically scanned radars featured moving dishes with many metallic moving parts, which will increase even further the unwanted scattering properties of the radar assembly.

For these reasons, modern low-observability fighter aircrafts usually employ PESA/AESA [9] radars which scan the intended volume using steerable beams rather than mechanically scanned dishes. Having a fixed dish, a backward inclined configuration can be used to facilitate the deflection of impinging waves away from the illuminating radar, as

9 PESA/AESA: Passive/Active Electronically Scanned Array. These types of radar do not require mechanical moving dishes since they rely on electronically configurable beam.

shown in Figure 29, which depicts a real-world operational AESA radar.

Figure 29: Raven ES-05 radar as installed in the nose of the Saab Gripen NG fighter aircraft

The described solution can be furthermore improved. In fact, the ideal configuration would feature a radome letting the radiation coming out but preventing at the same time unwanted radiation coming in from the outside world. The discriminating criteria to select which frequency can or cannot pass through the radome can be the frequency, angle of incidence and polarization of the impinging EM waves.

Therefore, the choice of the radome material not only satisfies the need of structure and intensity, but also realizes a selection primarily based on frequency: for this reason, the materials used for this type of application constitute the so-called Frequency Selective Surfaces (FSS). The FSS belong to the surfaces that are band-pass or band-stop at a given frequency and can efficiently control the transmission and reflection of the incident electromagnetic wave. Unlike traditional microwave filters, the FSS frequency responses are not only functions of frequency, but also functions of the incident angle and the

corresponding polarizations. Airborne radome which applies the FSS design technology can significantly reduce the interference of the spurious electromagnetic clutter with the electronic equipment and microwave radio-frequency antennas, improving stealth characteristics of the aircrafts (Sun, Xie, & Zhang, 2016).

Figure 30 shows a possible FSS implementation based on circular elements, whereas in literature and in practical application many diverse geometries can be found, each one relying on basic elements specifically shaped to obtained desired resonating-dissipating characteristics.

F-22 (Figure 5 and Figure 7) and F-35 (Figure 31) strike-fighter aircrafts successfully adopted FSS. Research on FSS structures is therefore regarded as a key technique to control the stealth characteristics and reduce the RCS of air vehicles (Sun, Xie, & Zhang, 2016).

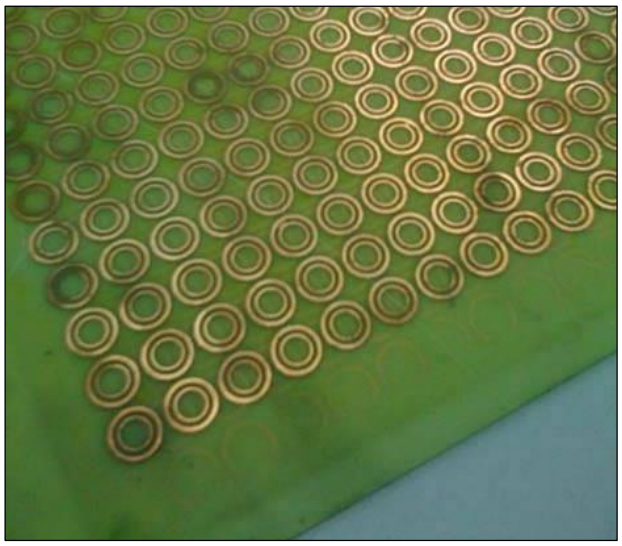


Figure 30: Circular elements geometry FSS

Ultimately, despite not being a resource relying on shaping or material considerations, there is another technique specifically designed

to reduce the detectability due to the use of the own-ship radar: the Low Probability of Intercept Radar (LPIR) technology. A LPIR is designed to be difficult to detect by passive radar detection equipment (such as a radar warning receiver – RWR) while it is searching for a target or engaged in target tracking. This characteristic is desirable in a radar because it allows finding and tracking an opponent without alerting them to the radar's presence. Ways of reducing the profile of a radar include using wider bandwidth (wideband), frequency hopping, using a frequency-modulated continuous-wave signal, and using only the minimum power required for the task. Using pulse compression also reduces the probability of detection, since the peak transmitted power is lower while the range and resolution is the same. Constructing a radar so as to emit minimal side and back lobes may also reduce the probability of interception when it is not pointing at the radar warning receiver. However, when the radar is sweeping a large volume of space for targets, it is likely that the main lobe will repeatedly be pointing at the RWR. Modern phased-array radars not only control their side lobes, they also use very thin, fast-moving beams of energy in complicated search patterns. This technique may be enough to confuse the RWR so it does not recognize the radar as a threat, even if the signal itself is detected.

### 1.6.2. Air-Data System

The Air Data System (ADS) is one of the essential systems in every airborne platform which computes several fundamental parameters (e.g. pressure altitude, calibrated airspeed, true airspeed, Mach number, static air temperature, and others) then used in the management of other main systems as, first of all, the engines and the Flight Control System (FCS).

Besides the internal circuitry dedicated to the processing of the gathered parameters, the ADS requires a series of probes that are placed

externally, along the "skin" of the aircraft. These probes are: Pitot-static tubes (Figure 31), static pressure ports (usually flush with the aircraft skin but still made of unpainted PEC metal), Total Air Temperature (TAT) probes (Figure 32 left), Angle of Attack (AoA) indicators (Figure 32 right). Moreover, for symmetry and redundancy considerations, the aircrafts usually feature more than one probes of the same category, positioned on both sides of the fuselage, typically in the nose area, so that they can sense an unperturbed airflow.

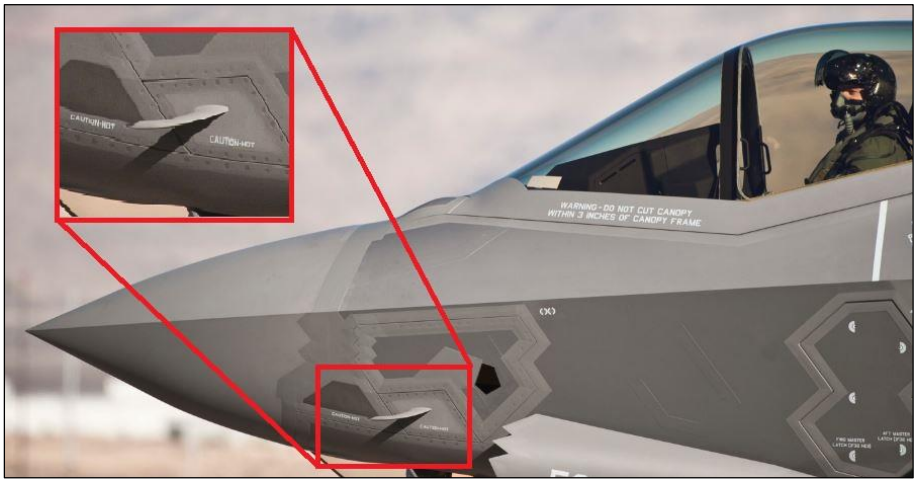



Figure 31: F-35 Pitot-static probe

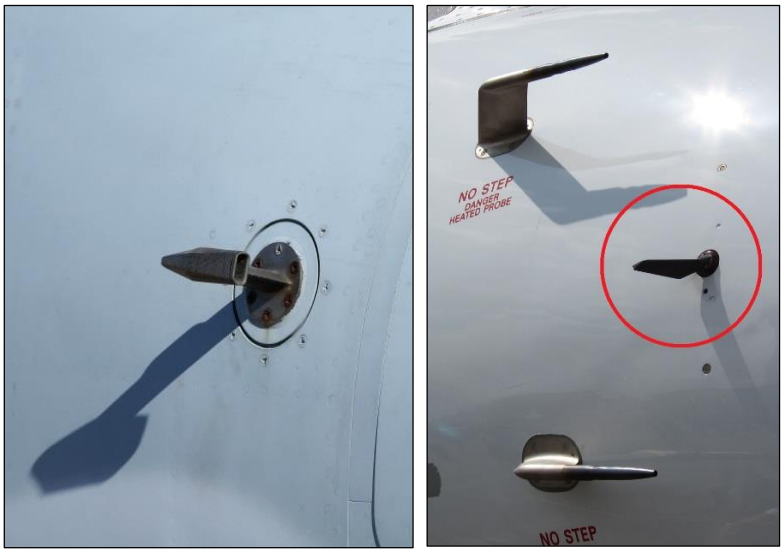



Figure 32: Total Air Temperature (left)
and AoA indicator between Pitot two tubes (right)

As one can clearly deduce from the pictures, these probes affect the smooth profile of a stealthy platform and so its RCS, due to the tip and edges diffraction type of scattering. However, some of these sensors can be designed in a low-observability fashion so that their impact is minimized: the probes are designed to be flush with the fuselages skin so that, despite the different material, the overall geometry is not a scattering geometry. A great example of a LO ADS (low-observable Air Data System) was implemented by means of several flush ports on the nEUROn UCAV, object of a test campaign accounted for during the second year of the PhD (Marchetto, Mercurio, Migliozzi, Piccinotti, & Risoldi, 2015). Unfortunately, for classification reasons, no pictures are available, as long as an assessment of the actual performance as substitute of the “classical” ADS.

Despite the efforts put in RCS-friendly designs for ADS, the Pitot-static tubes still maintain a RCS-disturbing geometry, as shown in Figure 31 which depicts the very low-observable F-35 aircraft with very evident Pitot-tubes.

### 1.6.3. External stores

Another common trend for strike-fighter modern low-observable aircraft is to carry the stores (e.g. launch or delivery weapons) within bays "buried" in airframe. The nEUROn UCAV technology demonstrator mentioned in the previous since involved in 2nd year activity during this PhD encompassed this type of technology. F-35 "JSF" (Figure 33), the well-known top-notch strike-fighter which is starting being employed by several military aviation around the globe, including the Italian Air Force, employs this technology as well. The drawback of such solution is that the room available in these internal bays is limited by aerodynamical and structural trade-offs. This imply the necessity, in specific type of mission profile (e.g. long-endurance deep strike with many stores), of resorting to conventional stores carriage with the weapons hung under the wings. The conventional solution significantly increases the overall RCS of the platform, since new tips, edges and interactions type of scattering are generated.

Figure 33: A F-35 JST showing internal stores carried within the bays Vs external stores conventionally carried under the wings

### 1.6.4. Stealth treatments deterioration and LO maintenance

As previously described, a key component in the RCS reduction deal is the use of several advanced materials, purposely designed to be transparent to, absorb, dissipate, and conduce EM fields and the associated currents. Stealth coatings include: RAM pastes, polycarbonate transparencies, resins, foams, edge sealing compounds, adhesives, primers, conductive films, multi-part adhesives, sealants, fillers, fairing materials, and organic topcoats. The majority of these elements are fragile or somehow prone to deterioration because of structural decay and corrosion, especially when exposed to adverse environmental conditions.

A contemporary real-world example can be found in the F-22s of the United States Air Force which are currently operating in the Middle East: these aircraft are the state of the art of stealth technology, protected by the highest level of classification. However, official sources recently made public that the radar-absorbing coating of the aircrafts warped and started to peel off. According to the US Air Force, climatic conditions in the area of usage of the warplanes are one of the reasons of this inconvenient. According to head of the F-22 program, the coating wrinkled and peeled off due to the fact that it lost its hardness and turned into its original liquid state. He also noted that this process is accelerated by external factors, including rain and sand dust. Additionally, during normal operations, RAMs were found to be severely deteriorated by contact with fuel and lubricating oil.

However, Developers of the F-22 Raptor claim that they have created a new, more stable formulation and are going to apply it to all the aircraft during maintenance operations. According to preliminary estimates, this process will take at least three years.

A stealthy platform needs peculiar maintenance process, overall indicated as “LO maintenance”. Such process may include the general

steps chronologically listed in the following and partially showed in Figure 34 thru Figure 37:

1) Inspect coatings, structures, and components to determine operational status;
2) Identify and remove corrosion using mechanical and chemical procedures (Figure 34);
3) Assess damage impacts to aircraft RCS signature;
4) Interpret inspection findings and determine corrective actions;
5) Remove RAMs by sanding, scraping or pulling using manual special tools (Figure 35);
6) Fabricate RAM repair parts such as strips and panels, and adhere them to aircraft surfaces and fasteners using vacuum bags, fixtures, and other pressure inducing processes, to cover skin slots, gaps, voids and gouges (Figure 36 and Figure 37);
7) Use ambient and accelerated cure processes to cure adhesives, sealants, fillers, fairing materials, and organic topcoats;
8) Apply organic low observable topcoats and rain erosion materials using spray equipment, brushes and rollers;
9) Inspect LO repairs to ensure compliance with technical data specifications also using portable maintenance aids (PMA) and automated maintenance systems;
10) Analyze and validate data processed with automated systems.

A fast RCS prediction tool could be used during the initial steps (i.e. step 3) of the mentioned process to obtain a quick assessment of the RCS deterioration due to unintentional damage and operational deterioration of the stealth treatments. As well, the tool could be used as the final check-out tool to verify the desirable RCS value have been reached after maintenance (i.e. steps 9 and 10).

Figure 34: Step 2


Figure 35: Step 5

Figure 36: Step 6


Figure 37: Step 6

# Chapter 2

# Implementation of a hybrid GO-PO parallelized algorithm for RCS prediction on GPU

The objective of the second Chapter is to describe the development and optimization of an efficient and accurate algorithm able to predict the RCS of an electrically large, arbitrarily shaped target, exploiting the concept of massive parallel computing applied to inherently parallelized processors, such as Graphical Processing Units (GPUs), commercially available on graphic cards.

Thanks to the introduction of the RCS concept provided in Chapter 1, it is now clear how RCS prediction is a deal of great interest in the military framework. However, for electrically large objects, as aircrafts and vessels in the radar band (S, C and X bands, typically), RCS predictions become drastically demanding from a computational stand-point. Therefore, it becomes necessary resorting to hardware expedients (i.e. parallel processors) and software expedients (i.e. powerful and efficient algorithms relying on advanced mathematical and algorithmical tools) in order to achieve an adequate trade-off between computational time and necessary processing power, for a required accuracy of the solution.

## 2.1. Introduction to the existing methods used to deal with scattering problems

Solving Maxwell's Equations has to be practically realized using a numerical approach, and several methods were developed along this direction. They are broadly distinct into two main branches: "exact" or "full-wave" methods (Figure 38, left tree) and "approximated" or "asymptotic" methods (Figure 38, right tree). However, depending on the nature of the given problem and scenario, the method must be chosen wisely: a wrong technique may yield to inconsistent results or require prohibitive computational time or memory usage.

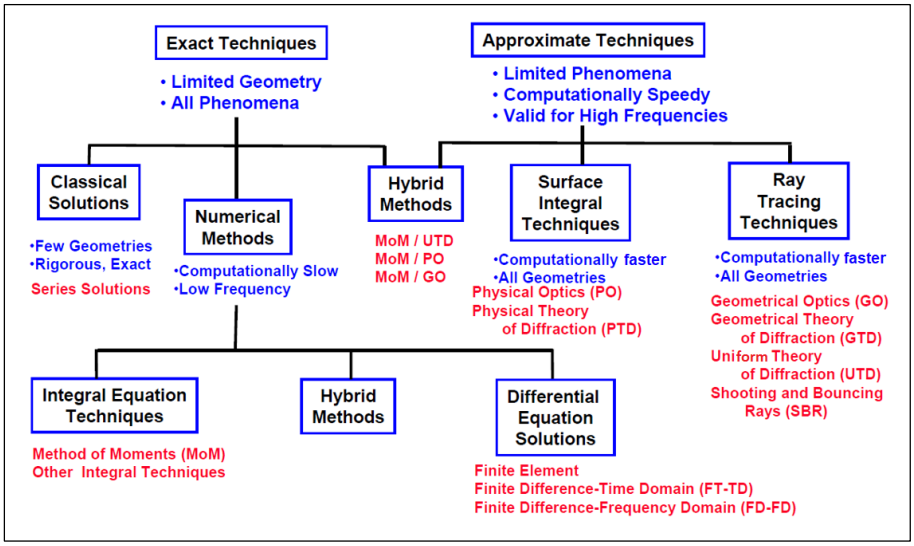



Figure 38: Overview of the methods developed to solve Maxwell's Equations

Using an exact, full-wave approach to solve Maxwell's equations with no a-priori approximations to predict the propagation of electromagnetic fields, both for indoor and outdoor scenarios, is difficult when dealing with microwaves frequencies: the computational burden and memory requirements increase extremely rapidly with the

electrical dimensions of the problems. Fortunately, electrically large problems can be tackled using a different category of techniques: the asymptotic methods which rely on initial approximations that alleviate the computation for the solution to Maxwell's Equations since the beginning.

The Geometrical Optics (GO) is an asymptotic method used at high frequencies that allows studying the EM field propagation by means of an optical ray-tracing model. It is generally a valid approximation when the index of refraction changes very little over a distance that is large compared with the wavelength. Geometrical Optics theory can be derived from Maxwell's equations as an asymptotic solution obtained in the limit as the frequency approaches infinity (Collin, 1985). In GO, the power density associated to the electromagnetic field is assumed to propagate along slender tubes called "rays" according to (Kline & Kay, 1965):

$$\begin{cases} \underline{E}(\underline{r}) = \underline{E}(\underline{r})e^{-jk_0L(\underline{r})} \\ \underline{H}(\underline{r}) = \underline{H}_0(\underline{r})e^{-jk_0L(\underline{r})} \end{cases}$$

Equation 8

where $k_0 = \omega\sqrt{\mu_0\varepsilon_0}$ is the propagation vector in free space and $L$ is called "Eikonal function" and describes the phase of the front waves. Note that both the amplitude information (i.e. $\underline{E}_0$ and $\underline{H}_0$) and the phase information (i.e. $L$) are functions of the position $\underline{r}$.

It can be applied to different scenarios and is suitable to deal with both Perfectly Electric Conductor (PEC) bodies and dielectric bodies. Additionally, it can model multiple interactions between different bodies as consequence of the reflection and refraction phenomena. However, GO has some major limitations: it cannot determine the EM field in those regions that are not reached by any ray, due to the shadowing of certain portions of the region of interest; additionally, the obtained solution is not valid near the edges of the scatterer; and finally,

the method suffers when dealing with almost-flat surfaces, resulting in huge unrealistic values for the RCS (Knott, A Progression of High-Frequency RCS Prediction Techniques, 1985).

A step forward from the GO is the Physical Optics (PO) which helps to overcome some of the GO shortcomings: PO uses a geometrical optics approximation of the currents induced on a PEC surface and integrates the induced currents to obtain the scattered field. The Far Field scattered from a PEC body is given by:

$$\underline{E_s} = -z_0 \hat{r} \times \underline{H_s}$$

Equation 9

$$\underline{H_s} = -\frac{jk}{4\pi}\left(\frac{e^{-jkr}}{r}\right)\underline{U}$$

Equation 10

where:

- $z_0$ is the free space impedance (free space was assumed);
- $r$ is the distance from the field point to the origin;
- $\hat{r}$ is the unit vector in the direction of observation;
- $\underline{U}$ is the vector Far Field amplitude and is given by:

$$\underline{U} = \hat{r} \times \int_S \underline{J}(\underline{\rho}) e^{jk\underline{\rho}\cdot\hat{r}} dS$$

Equation 11

in which $\underline{J}(\underline{\rho}) = \hat{n} \times \underline{H}(\underline{\rho})$ is the surface current, $\hat{n}$ is the normal unit vector to the surface, $\underline{\rho}$ is a position vector from the origin to a point on the surface $S$ of the scatterer, $\underline{H}(\underline{\rho})$ is the total magnetic field (Kouyoumjian, 1965) which could be approximately calculated from the incident field.

The advantage of using the PO over GO is that PO is often acceptably accurate for flat and singly curved surfaces, provided that the scattering direction remains within a few side-lobes of the specular direction. Moreover, the surface integral for a flat plate can be expressed in terms of a simple contour integral around the perimeter of the plate, thereby simplifying computations (Knott, A Progression of High-Frequency RCS Prediction Techniques, 1985).

Seeking to improve the accuracy of the EM field calculation in those regions where the GO and the PO fails to be adequately accurate, the Geometrical Theory of Diffraction (GTD) was developed. GTD introduced a new kind of special rays, called “diffracted rays”, which are produced by objects discontinuities, such as edges, wedges and abrupt variations in the curvature of the surfaces. Diffracted rays allowed to partially model diffraction effects. Eventually, GO extended and improved by PO and GTD, let us deal with configurations characterized by very large electrical dimensions and complex shape, showing a significant numerical efficiency and achieving adequate accuracy with respect to reality (Schmitz, Rick, Karolski, Kuhlen, & Kobbelt, 2011) (Buddendick & Eibert, 2010).

## 2.2. Approaching the RCS scattering problem

The research activity consisted in developing and validating a hybrid GO/PO/GTD tool suitable to deal with electrically large complex scenes. In particular, the implemented algorithms employed parallel codes running on GPU, aiming at achieving the best performance in terms of computational times, for a given required accuracy (Breglia, Capozzoli, Curcio, Liseno, & Piccinotti, GPU implementation of hybrid GO/PO BVH-based algorithm for RCS predictions, 2015).

The algorithm was able to determine the RCS of a PEC object in a set of designated directions, for a given set of sources and for a given

scene, represented in terms of the geometrical shape of the object under evaluation, which was located in the Far Field region with respect to the sources.

The obtained results were promising, allowing in the future studying how the codes would perform when applied to optimizations problems, where recursive use of the algorithm is required to find the best solution for a complex problem, such as antenna placing, for instance. Additionally, as future development, consideration should be given to the effects caused by introducing different materials characterized by diverse electrical permittivity and magnetic permeability.

The electromagnetic scattering problem, as considered during this research, can be modeled as described in the following. Accordingly, the GO algorithm can be divided in two main consequent steps (Tao, Lin, & Bao, 2010) (Gao, Tao, & Lin, 2013): firstly, the ray tracing phase which focuses on the paths followed by the rays, and secondly, the EM field transportation phase.

As input, the algorithm was expecting a 3D geometric scene, characterized by one or several arbitrarily shaped objects, represented by means of discretizing mesh-grids composed by geometric elements called “primitives”, which, in general, may be rectangles (Figure 39) or, as for the specific tool described in these research, triangles (Figure 40). The mesh-grid density, and thus the total number of primitives, must allow reaching an accuracy in the scene description so that the approximations imposed by the GO theory can be applied (Gao, Tao, & Lin, 2013).

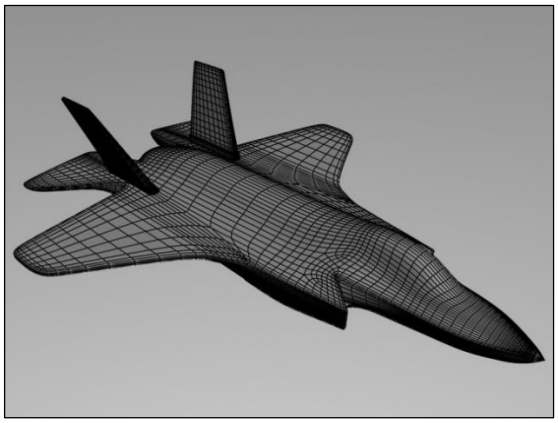


Figure 39: Rectangular element meshgrid

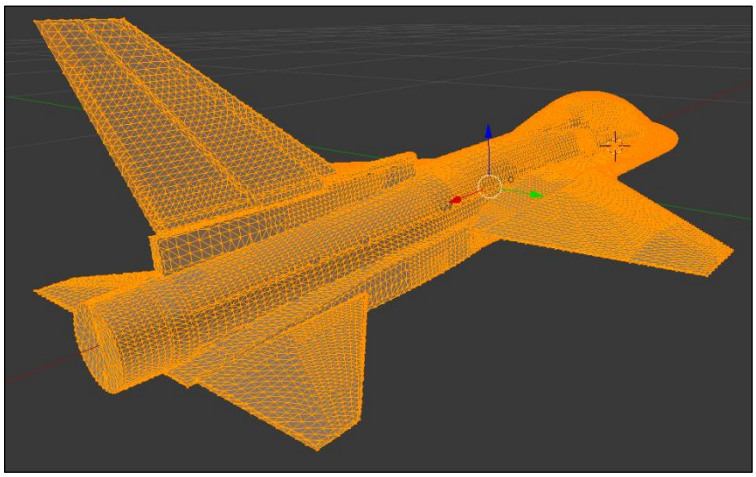


Figure 40: Triangular element meshgrid

It can be shown that such density should be directly proportional to the radii of curvature of the modeled surfaces: a flat surface will require few large primitives, whereas a surface characterized by abrupt changes in shape with tight radii of curvature will require many more triangles of a smaller size. Additionally, the GO theory requires the radii of curvature of the whole surface to be much greater than the wavelength of the waves associated to the impinging EM field. Based on these starting criteria, when modeling complex scenes, the result of the

discretization is a large set of triangles with a highly variable density, which results in a significantly non-uniform mesh-grid.

To analyze the EM propagation phase, a homogenous medium such as air or free space was assumed: this resulted in the rays following rectilinear trajectories. The description of the sources generating the incident EM field was then required as input for the algorithm. A suitable technique for defining the sources consisted in providing, for each source, the radiated field on a wave front: in the RCS case, since the interest was only on the Far Field range, the wave front was always assumed as a plane wave front. Then, GO rays were launched from this wave front to model the propagation phase.

The algorithm output consisted in the hybrid GO/PO field calculated along the directions of interest by means of the two steps previously mentioned: ray tracing and field transport. It can be shown that, from a numerical standpoint, the ray tracing, if not optimized, would result very demanding. The choice of which method to use for the ray tracing was based on the following premises:

- geometry complexity;
- computational hardware architecture (i.e. GPU-based dedicated workstation with a high degree of parallelization vs standard CPU-based serial workstation);
- requirements in terms of computational times.

## 2.3. Algorithm phase one: ray tracing

The ray tracing phase consisted in finding all the relevant paths followed by the GO rays. The rays were launched with no dependency among them, which made this algorithm extremely suitable to be parallelized since each ray could be traced concurrently with the others. The number of launched rays for each iteration was generally extremely high, so that a maximum number of paths could be simultaneously

analyzed, saturating the available resources of the GPU (i.e. available memory, number of threads, streaming processors, etc).

When introducing objects in the scene (e.g. a target of interest) for highly advanced electromagnetic applications, a simple brute force approach that performed a ray-primitive test (Figure 41) for each ray and each primitive in the scene would have been unfeasible since the computational time to evaluate the intersections grows linearly with the number of objects per ray.

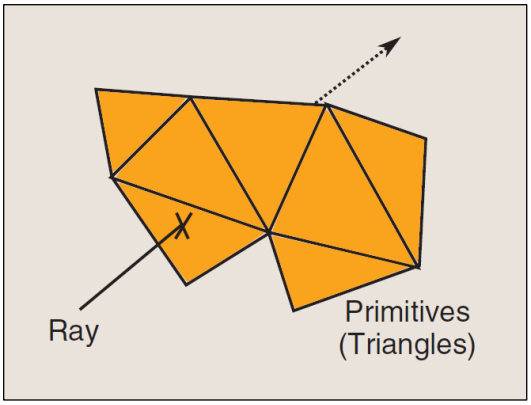



Figure 41: Ray-primitive test

This problem had already been extensively studied in computer graphics and several accelerating structures had been developed, some of them optimized for parallel architecture machines. By enclosing the objects of interest in bounding volumes, the intersection tests can be performed on the bounding volumes prior to testing the objects, drastically reducing the number of potential candidates to deal with intersections. Providing the bounding volumes with a tree-like organization hierarchy, namely wrapping bounding volumes in larger bounding volumes, further improves the computational efficiency. It should be noted that in many iterative applications involving non-dynamic scenarios (e.g. antenna placement), tree-like data structures can be constructed once for all and then reused during the whole

optimization process, so that construction costs are amortized (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

To date, in the field of computer graphics, several of such acceleration schemes have been developed, including: regular grids, octrees, KD-trees, Bounding Volume Hierarchies (BVHs), and others (to be noted, in literature and within this work as well, the term "Bounding Volume Hierarchy" and the corresponding acronym "BVH" are used to indicate either the general concept of hierarchy of volumes, and the specific method used to construct the hierarchy: this nomenclature abuse will be clarified in the following). Some of these schemes make the computational time per ray logarithmically dependent on the number of primitives. Currently, the most widely used and fast techniques in Computer Generated Imagery (CGI) are the KD-tree and the BVH with its variants.

Over the last decade, the performance of the ray-primitive test phase has been constantly improved by smart implementations of both these techniques, along with the employment of parallel computational capabilities of modern GPUs. In particular, an impressive peak-performance has been achieved by using GPUs of the current generation in conjunction with the BVH, indicating the feasibility of very high-performance ray tracing (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

By contrast, in the past, within the electromagnetics community, the attention was totally focused on the KD-tree scheme with either sequential or parallel approaches. Although for sequential algorithms the KD-tree showed very good performance, results indicated that parallel ray tracing on GPUs could benefit from the different properties of the BVH strategy coming from CGI world. Thus, a need for more efficient data structures arose within the electromagnetic community,

leading to an investigation of which technique was the most appropriate for a given scenario.

Within the purpose of this work, the most suitable acceleration data structure for electromagnetic ray-tracing on GPUs using the Nvidia CUDA language was identified, leading to an ultra-fast ray-tracing when compared to what is currently available in the literature. In particular, both KD-tree and BVH approaches were considered and adapted to the electromagnetic case, and their performances compared. The comparison was mainly focused on computational speed, but the differences in terms of ease of implementation, numerical robustness in geometrical calculations, and memory occupancy, which are the parameters identifying the convenience of a data structure against the others, were also pointed out (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

The first effort was aimed at developing two ray tracing algorithms, respectively employing the KD-tree and the BVH variant techniques, with the purpose of determining which one would obtain the best performance on a GPU-based parallel computing machine. The scenarios on which the algorithms were tested included objects with major electrical sizes which were discretized in a very large number of triangles. Typically, with the tools available in the past, handling such scenarios would have been very difficult. Some of the scenes, as the one depicted in Figure 42, were of actual electromagnetic interest, whereas others were inherited from the CGI world and were exploited as benchmark for computational testing purposes.

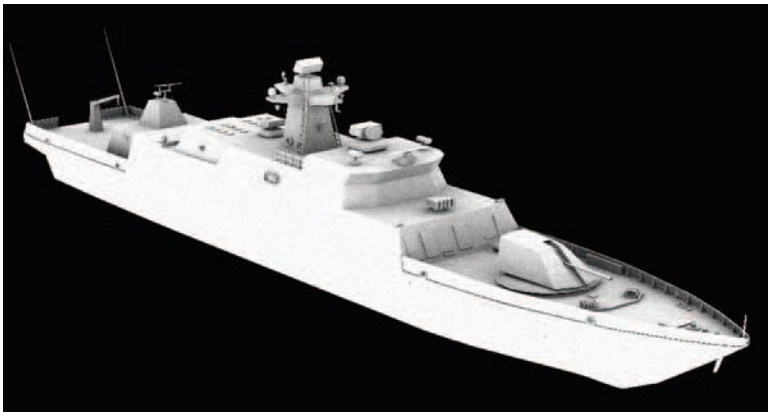


Figure 42: Scene of electromagnetic interest

For the sake of simplicity, we only considered objects as impenetrable from an electromagnetic point of view and with triangles meshing their surfaces. However, the approach could have also been applied to the case of penetrable objects to evaluate the congruencies of refracted rays. Additionally, we focused our attention exclusively on the efficiency of ray tracing and ray intersecting, which are of electromagnetic interest, avoiding discussing any electromagnetic aspect within the ray tracing itself, such as phase, intensity, and polarization transport, assuming it feasible.

Furthermore, even "ambient occlusion rays" have been considered, which are capable of providing a more consistent estimate of the speed reachable in electromagnetic ray tracing, which was in the order of magnitude of $\sim 10^6$ ray/second for the developed tool. Ambient occlusion rays are also suitable to deal with rough surfaces (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

In the scenarios considered during the development of the tool, the rays were propagating along linear trajectories because of the homogenous media employed, so that each path was characterized by a polygonal chain composed by a finite number of rectilinear segments. In particular, each ray launched from the source, experienced a certain

number of interactions (i.e. reflections, refractions and diffractions) until it exited the scene or it reached the maximum number of allowed interactions (set as an algorithm parameter).

In order to determine all the relevant paths, the number of possible paths to be evaluated and the number of rays to be launched and traced was significantly large (i.e. $10^8$ order of magnitude), so that a fast ray tracing algorithm was necessary. Each ray was modeled as a half-line described by an originating point and a direction. Tracing a ray was realized through the following steps, which had to be repeatedly executed:

1) locate the intersection point between the ray and the object surface which resulted being as close as possible to the originating point;
2) if the ray intersected any surface, then a new ray had to be launched, whose originating point is at the mentioned intersection and whose direction is determined by the laws of GO (i.e. reflection, refraction, diffraction).

Steps 1 and 2 were repeated until the ray experienced a pre-selected maximum number of reflections or exited the domain set for the calculation. From a computation standpoint, step 1 was the most onerous time-wise. A thorough evaluation testing for all the triangular primitives on the surface could not be adopted in complex scenes due to the extremely large number of intersection tests to be performed. Fortunately, though, for a given ray, the majority of the triangles constituting the scene was not of interest for that specific ray. This allowed employing a tree-like structure to hierarchically organize the primitives in the scene. Such expedient drastically reduced the required number of ray-primitive intersection tests and the overall computational burden of the algorithm.

It will be shown in the following how the optimized BVH approach can be faster than the current approaches available in the literature.

## 2.4. Ray tracing acceleration schemes

A common concept of all acceleration schemes is the bounding volume. A bounding volume is a simple geometrical entity that can enclose a group of geometrical primitives or even other bounding volumes. When a ray intersects a bounding volume, that ray will also most likely intersect some of the objects inside the bounding volume. The term "object" signifies a list of primitives that, in principle, can even belong to different physical scatterers. The volumes have a shape simple enough to be tested for the primitive-ray intersections by means of fast algorithms. In fact, the volumes are usually designed as parallelepipeds with the axes aligned to the reference system axes. In this case, it is reasonable to search for an intersection with those objects, since evaluating the intersection with a parallelepiped is easier than with a more complex object.

When testing the primitives and rays for intersection, the test is firstly performed on the bounding volumes: if the ray is not intersecting a volume, then all the primitives contained in that volume will not be intersected as well, and can obviously be neglected within that test.

However, to drastically improve the performance of ray tracing, a hierarchy must be employed. Incidentally, in the literature and within this work as well, the term "Bounding Volume Hierarchy" and the corresponding acronym "BVH" are used to indicate either the general concept of hierarchy of volumes, and a specific method used to construct the hierarchy.

Regardless of the specific criterion used, the hierarchical structure is practically obtained like so: depending on the chosen approach a specific partitioning criterion is given and then, starting from the bounding volume of the whole scene, the hierarchy of volumes is built by iteratively partitioning "parent volumes" into "child volumes" (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration

Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

The BVH is then managed like a tree-based data structure, with two types of nodes: inner nodes and leaf nodes (Figure 43). Both inner and leaf nodes hold bounding volumes, with the specification that the bounding volume corresponding to a leaf node contains only primitives and the volume wraps all of them. In practice, the primitives are then stored in memory and the object is a list referring to the stored primitives.

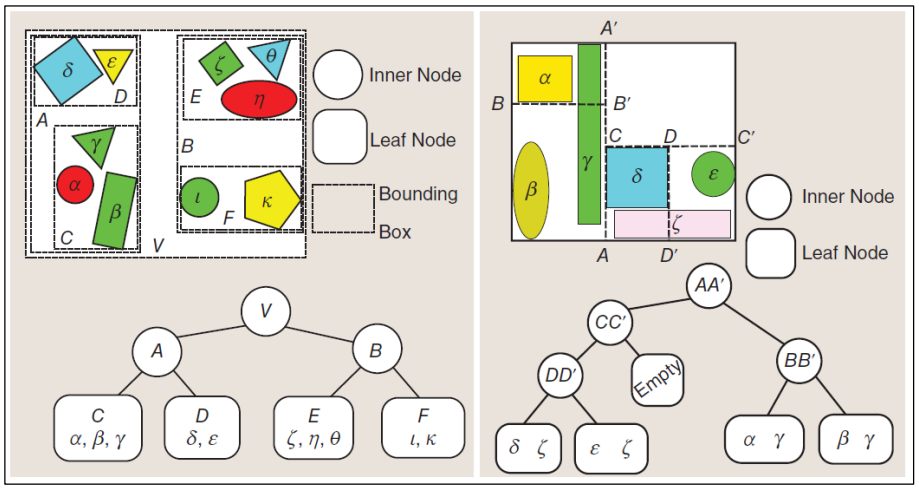



Figure 43: Object partitioning BVH (left) vs Spatial sub-division KD-tree (right)

With such a hierarchy, once a ray is launched, a tree-search algorithm can be used to find the hit object closest to the launching point. The intersection tests are then performed firstly with the bounding volume at the root of the hierarchy (i.e. the biggest external box), and secondly with only the child nodes that originates from parent nodes that were intersected. Whenever a leaf node is intersected, its primitives are considered for intersection as well.

The shape of the bounding volume should be optimized such that a low memory occupancy is necessary and fast ray-primitive intersection

tests can be executed. Reducing the memory allocation for the single bounding volume will result in an overall small required memory and the fast ray-primitive tests will allow a fast traversing of the hierarchy. In fact, many different trees can be built for a given scene, but they will have different efficiency in ray traversal. With regard to this question, an important characteristic of a BVH is the number of children per node. Practically, two children per node is by far the most common choice because a binary tree is the easiest to build and manage.

Furthermore, the most commonly used bounding volume is an Axis-Aligned Bounding Box (AABB), which is completely determined by the "minimum" and "maximum" corners and which allows fast intersection tests. An AABB is a rectangular six-sided box whose normals to its faces are parallel with the axes of the given coordinate system. As depicted in Figure 44, the two vertices $C_{min}$ and $C_{max}$ identify the region $R = \{(x, y, z) | C_{min}^{x} \leq x \leq C_{max}^{x}, C_{min}^{y} \leq y \leq C_{max}^{y}, C_{min}^{z} \leq z \leq C_{max}^{z}\}$ (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

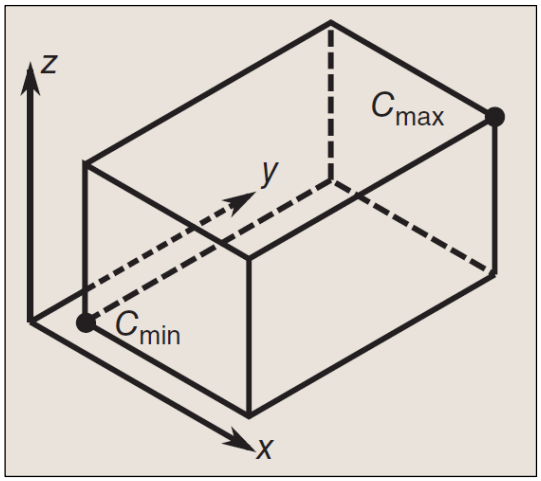



Figure 44: Axis-Aligned Bounding Box

Within the purpose of this work, only BVHs with AABBs were considered. In particular, we examined two types of structures: the KD-tree and the BVH ("BVH" now addresses the specific technique). Both are binary trees with AABBs, but they feature different constraints that will be discussed in detail in the following: the first approach operates a spatial sub-division and leads to the KD-tree, whereas the second approach works on the objects and leads to the BVH. In the first case the space is divided, in the second case the objects are grouped (Havran, 2001).

As consequence of the described working principles, both for the KD-tree and the BVH, the ray-primitive tests are executed following a tree-like flow where only the "promising" volumes of the tree are considered. This type of hierarches are not the only ones available and within all the possible ones, some of them are faster in terms of ray tracing. The differences in performance depends on:

- intrinsic capability of the hierarchy to effectively group a certain scene;
- distribution of the rays during the launch phase;
- adaptability of the structure to the parallel computing hardware configuration.

The mixing of the two presented methods generates some interesting hybrid structures such as the Split BVH (SBVH) or the Bounding Interval Hierarchy (BIH) (Wachter & Keller, 2006), which allow reaching the highest performance when employing GPUs. The KD-tree, normal BVH and SBVH will be presented in the following.

### 2.4.1. Spatial sub-division: the KD-tree

The KD-tree algorithm (Popov, Unther, Seidel, & Slusallek, 2007) analyses the geometry and then iteratively divides the volume of interest using planes. At each step, the tree is automatically built by a process that considers a specific volume which then cuts in two halves

spatially disjoint by means of a plane aligned with the axes. Thus, step by step, the primitives are sorted between the two halves.

A list of primitives along with the corresponding AABB that envelops all of them is given as in Figure 45. A binary spatial subdivision scheme splits the node AABB, which is the root node indicated by the blue continuous line, into two non-overlapping sub-AABBs.

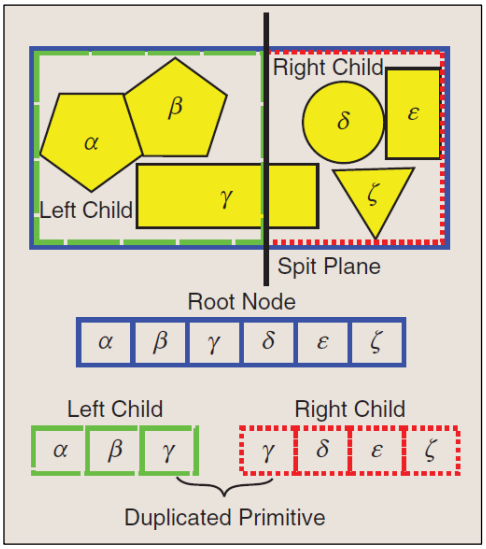



Figure 45: KD-tree working principle

The process used to divide the volume is based on a "heuristic technique" [10]. The specific heuristics adopted to perform volume and object partitioning will be addressed in the following. Therefore, based on such heuristic, the spatial subdivision algorithm selects as best as it can an axis aligned splitting plane, which is a plane with the normal unit

10 A heuristic technique, often called simply a "heuristic", is generally defined as an approach to problem solving, learning, or discovery, that employs a practical method not guaranteed to be optimal, but sufficient for immediate goals. Where finding an optimal solution is impossible or impractical, heuristic methods can be used to speed up the process of finding a satisfactory solution.

vector aligned with one of the coordinate axes. This plane creates two nonoverlapping sub-AABBs and the corresponding sub-nodes are depicted in Figure 45 with the green dashed-line and the red dotted-line. The primitives in the list are divided accordingly, where straddling primitives are copied in both lists creating duplicates and allowing a partial overlap between the two new lists of primitives.

Recursively subdividing each sub-volume and sub-list with an axis aligned plane will generate a KD-tree (Figure 46). It is important to underline that the parent nodes do not hold the primitives: only child nodes contain primitives and are stored. As per this mechanism, following the construction of the tree until the end, only the leaf nodes will store the primitives (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

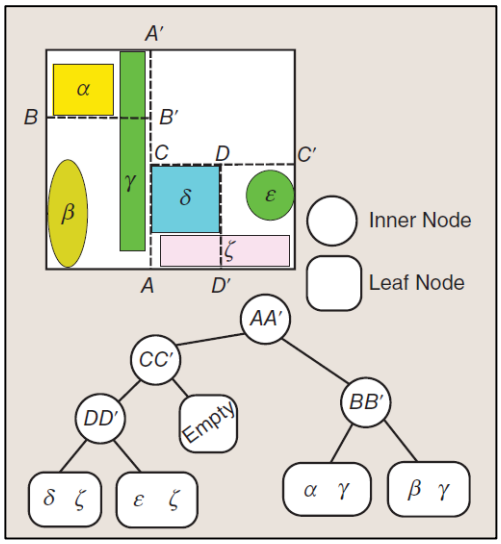



Figure 46: KD-tree structure

Initially the algorithm is applied to the entire scene which is going to be divided in many non-overlapping volumes: the recursive process is terminated when the space left to be divided contains a number of

primitives under a fixed threshold. It is important to locate the dividing plane in a position such that the number of intersections with the primitives is minimized: this will optimize the ray tracing.

As consequence of the described process, the only information needed to store primitives in a KD-tree node is:

1) the axis (*x*, *y* or *z*) normal to which the split occurs, which requires only 2 bits of information;
2) the references to the child nodes, which is an information required for traversing the tree;
3) the references to the specific primitive, only if the node is a leaf node.

KD-tree structures have an important property that follows directly from spatial subdivision: if a ray intersects a primitive in the volume nearest to the ray origin *O*, then all other potential intersections in a farther volume will be even farther from *O*. Accordingly, in the case of impenetrable objects, there is no need to perform other ray-primitive intersection tests for the farther volumes to check for closer ray-primitive intersections, as illustrated by Ray 1 in Figure 47.

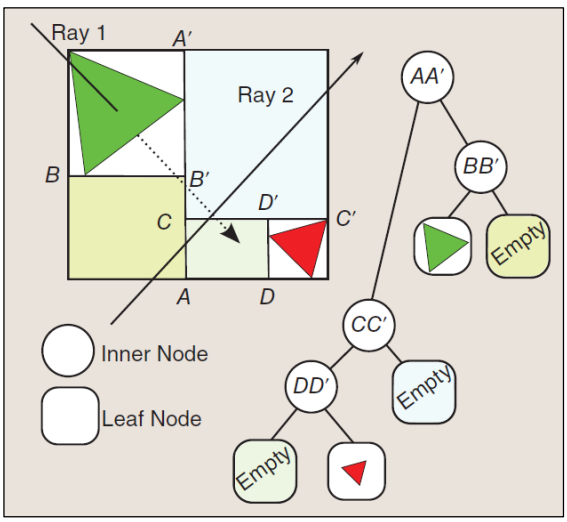



Figure 47: Primitive impenetrability

Note that a KD-tree may produce empty leaves to cut off empty spaces. As shown in Figure 47 by Ray 2, the spatial subdivision algorithm may split the AABB containing the two triangles (red and green), along the line *AA*', creating two sub-AABBs. The leftmost AABB is split again along *BB'* producing an empty leaf node and a leaf node with the green triangle. On the other side, the algorithm first subdivides the rightmost AABB along *CC*', thus creating an empty leaf node. Afterward, it subdivides the remaining space along *DD*', forming another empty leaf and the leaf with the red triangle. Ray 2 will only visit empty leaves and no ray-primitive tests will be performed.

Despite being possibly the fastest sequential data structure known nowadays, KD-tree structures suffer from high memory consumption due to high primitive duplicates. Also, the spatial subdivision requires care with round-off errors to avoid splitting off some primitives, especially axis-aligned primitives commonly encountered in the meshes of the scenes of interest. These drawbacks will be discussed in the following (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

#### 2.4.1.1. Finding the Splitting Planes for the KD-Tree

For the KD-tree algorithm, the best axis-aligned plane which minimizes the cost must be chosen. For a given list of primitive references and for a given AABB to split, said cost is calculated according to the Surface Area Heuristic (SAH) algorithm, which will be discussed in the following.

Firstly, the split axis is chosen and secondly, for the selected axis, the position of the plane is continuously varied allowing a variable optimization. However, since it can be shown that the local minima of the cost function occur only at the starting or ending position of the

primitives (including the clipped ones), only these planes are considered as good split candidates. It is important to clarify that, when a primitive is clipped, the starting (or ending) position where the minima are located, is the one of the clipped primitive and not the one of the original entire primitive. Accordingly, for the rest of the section, for the sake of simplicity, the primitives will be addressed as “clipped” regardless as whether they are actually clipped or not.

In general, for each node partitioning, three subsequent optimizations along the three coordinate axes are performed. Furthermore, for each coordinate axis (i.e. x, y, z), the algorithm considers only the splitting plane candidates parallel to the coordinate planes (e.g. the y-z plane for the case of optimization along the x-axis). Therefore, the splitting plane that results in the minimum cost along the three axes is used to perform the spatial subdivision.

Figure 48 shows a split executed along the x-axis, where five triangles are depicted, including one (black) parallel to the y-z plane. Dashed lines represent the potential split positions along the x-axis.

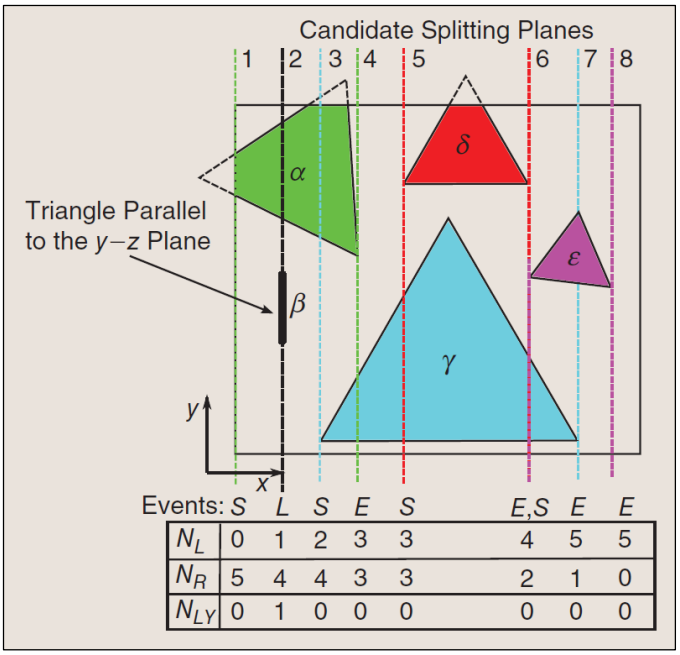



Figure 48: Split along the x-axis

As previously explained, the potential positions of the splitting plane only occur at the starting or ending positions of the clipped primitives. For each possible position, the algorithm calculates the cost for the AABB split, which means solving Equation 12, which will be further commented in the following section dedicated to the SAH.

$$C(B_0) \approx K_T + K_1 \, , \left(P_{[B_1|B_0]} N_1 + P_{[B_2|B_0]} N_2\right)$$

Equation 12

Such an equation requires calculating $N_1$ and $N_2$ which are the number of primitives falling in the two child nodes after the splitting. A naive approach would reclassify each time the primitives into the child nodes for each candidate splitting plane: this would have a considerable cost since there are at most six planes to consider for each primitive. This leads to a computational complexity of $O(N^2)$, where $N$ is the

number of primitives. Fortunately, a more efficient way to perform the reclassification exists, which scales as $O(N(\log N)^2)$ (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

For a fixed splitting axis, the algorithm described in the following creates a sorted list of candidate splitting planes from the primitives belonging to the node to be split. The list is ordered because the triangles belonging to a node are not ordered, and so, a listing criterium must be adopted. When scanning the list in an ordered fashion and moving from one candidate to the next, the mentioned criterium eases the calculation of the number of primitives left behind and still to come across. The list of candidate splitting planes is managed as a list of events identified with the pair $(p, type)$ that contains the position $p$ of the splitting plane, and the type $type$ of the event, coded by an integer value as:

- E: ending = 0;
- L: lying = 1;
- S: starting = 2;

where "ending" means that the splitting plane touches the clipped primitive at its rightmost point, while "starting" at its leftmost point, and "lying" that the clipped primitive lies on a candidate splitting plane. Numerically, the type of the event is an integer (i.e. 0, 1, or 2) to enable an ordering process of the event, according to the following comparison operator:

$$a < b \coloneqq a_p < b_p \; OR \; (a_p = b_p \; AND \; a_{type} < b_{type})$$

Equation 13

According to Equation 13, the event $a$ comes before event $b$ (i.e. $a < b$ is true) if and only if the candidate splitting plane $a_p$ is on the left of $b_p$ (i.e., $a_p < b_p$) or if the two candidate splitting planes

coincide (namely, $a_p = b_p$) and the type of the event $a$ is less than that of $b$.

Again, Figure 48 presents a concrete sample case. The first examined plane is plane 1. Since plane 1 is the first and since it is of "starting" type, then the number $N_L$ of primitives on its left is initialized to 0. The number $N_{LY}$ of lying primitives, such as axis-aligned triangles, is initialized to 0. The number $N_R$ of primitives on the right is initialized to the overall number of primitives in the AABB, namely 5. When moving to plane 2, the number of $N_R$ primitives is reduced by 1. Since plane 2 is "lying," both $N_{LY}$ and $N_L$ are increased by one. Recapping, for a generic position in the event list:

- $N_L$ is the sum of the events of the starting and lying categories prior to that position;
- $N_R$ is the sum of the events of the ending and lying categories after that position;
- $N_{LY}$ is the sum of events of the lying category on that position.

Note that events may "overlap" in the sense that the same plane may correspond to two different event types (see the event associated with plane 6 in Figure 48, which simultaneously corresponds to a "starting" and "ending" event). In this case, a single but multiple-type event must be taken into account. Additionally, straddling primitives appear as members of both starting and ending categories.

## 2.4.2. Object partitioning: the BVH

In this grouping scheme, the list of primitives is recursively partitioned creating at each step a couple of disjoint non-empty lists [11] and the bounding volume for each list is then calculated (Aila & Laine,

11 When located in the 3D space of the geometry the primitives can be grouped in 3D "sets" (i.e. volumes); instead, when treated as items in a sorting algorithm, they are obviously better organized and processed in "lists".

Understanding the efficiency of ray trasversal on GPUs, 2009) and accounted as a node (Aila, Laine, & Karras, Understanding the efficiency of ray trasversal on GPUs - Kepler and Fermi addendum, 2012). The choice of the items to be placed in one of the two lists is based on the heuristic process called Surface Area Heuristic (SAH) process which optimizes the items locations. The iteration stops when a minimum number of primitives to be sorted is reached or when sorting the primitives even further is not convenient anymore. When it comes to perform automatic sorting, the BVH results being fast and efficient, since is basically a "simple" sorting algorithm.

Figure 49 shows a list of N primitives and the external minimum AABB enclosing them, which is the algorithm starting point. A binary object partitioning scheme directly subdivides the list of primitives into two separated sub-lists according to a specific heuristic, showing a single object partitioning step. Said sub-lists are non-empty and disjoint, even if the related bounding boxes are overlapping. Note that there are $2^N - 2$ binary partitions of the list. The root node contains a list of the six references to the primitives and its AABB encloses all the primitives. The object partitioning algorithm first subdivides the list of the root node into two disjoint sub-lists and for each sub-list, the algorithm computes the minimum AABB that contains all of the primitives. Two child nodes are then created with their respective lists of references and AABBs. The result of this subdivision is a pair of non-empty and non-overlapping sub-lists.

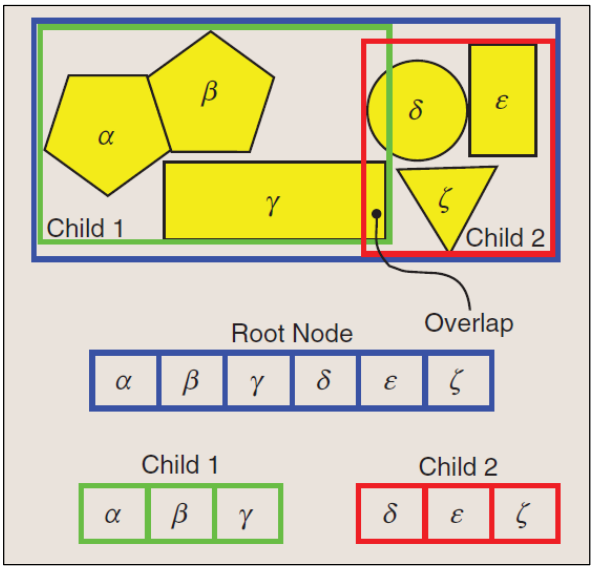



Figure 49: BVH working principle

In general, and for the particular example of Figure 49, the AABBs belonging to the two sub-lists may partially overlap. As a consequence, an intersection in one of the AABBs does not exclude a closer intersection in the other. Also, it is worth remarking that the object list partitioning does not create primitive duplicates and is the only possibility if one wants to minimize memory occupancy.

As shown in Figure 50, recursively partitioning each sub-list will create a BVH of AABBs. Differently from the previous case, a BVH node needs more information to be recorded than a KD-tree. In fact, all the nodes need to store:

1) the coordinates of the bounding boxes which require six floating points: three for $C_{min}$ and three for $C_{max}$;
2) the references to the child nodes, which is an information required for traversing the tree;
3) the references to the specific primitive, only if the node is a leaf node.

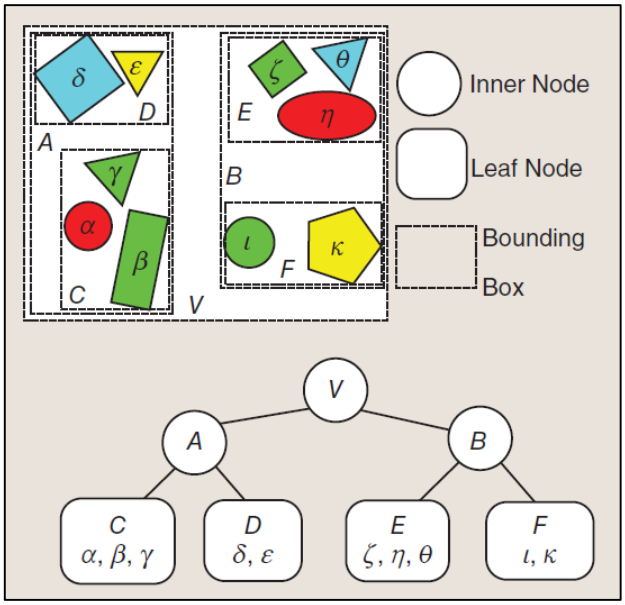



Figure 50: BVH basic working principle

Despite requiring a larger storage memory than the KD-tree, in the BVH, the number of references to the primitives is limited by the total number of primitives themselves, since no primitive clipping and duplication occurs, resulting in only comparison operations required. Accordingly, the memory occupancy of the BVH is much smaller compared to the KD-tree one. There are also other pros in the BVH choice: from a coding standpoint, building a good BVH is easier than building a KD-tree; from a GPU-based computational standpoint, BVH is faster since it is more suitable for parallelization; finally, from a numerical standpoint, it is less prone to numerical errors (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

#### 2.4.2.1. Finding the Bounding Boxes for BVH

The object partitioning is significantly simpler than finding the splitting planes for the spatial subdivision algorithm. A list of $N$ primitives is sorted according to a certain criterion and the best partition for said list is found, creating two non-empty disjoint lists. Therefore, while the KD-tree often produces empty leaves to cut off empty space, this does not happen with the BVH. Indeed, BVH nodes enclose only non-empty space thanks to the flexibility of the AABBs.

In most implementations, the BVHs are constructed in the following way:

1) in each partition step, the list of primitives is sorted along each axis with respect to the primitive centroids;
2) the ordered list is then split into two sub-lists so that the SAH cost according to Equation 22 is minimized;
3) for each sub-list, a bounding box is created and assigned to the corresponding child node.

The process is recursively executed. Note that a BVH construction algorithm does not require clipping the primitives, because they are always entirely bounded.

The described approach is widely used but produces sub-optimal BVHs. This is particularly accentuated, for instance, when dealing with architectural scenes, which are those scenes that include floors and walls, and for which the rays are launched from within the scene. The "Conference" scene (Figure 51 left) is an example of architectural scene, while the "Bunny" scene (Figure 51 right) or the "Ship" scene (Figure 42) are not. Therefore, architectural scenes are characterized by indoor or urban propagation scenarios: in these cases, large primitives crossing all over the scene, such as primitives belonging to the walls or to the floor, will force the algorithm to generate a child node that includes another one (Figure 52 a) (Breglia, Capozzoli, Curcio, &

Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

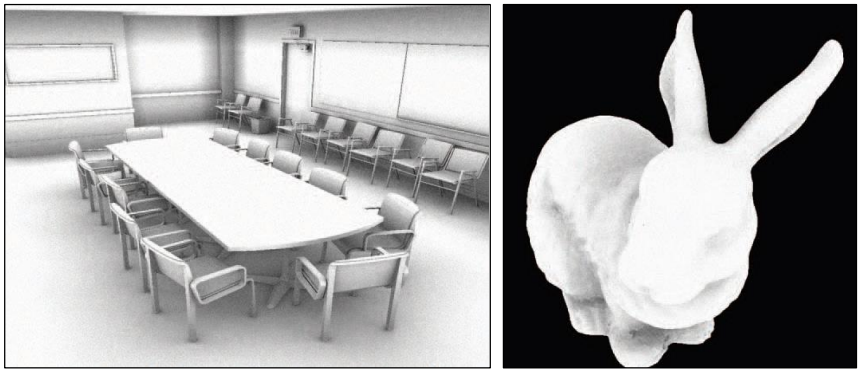


Figure 51: Architectural scene (left) vs non-architectural scene (right)

An example of this unwanted phenomenon is depicted in Figure 52 (a) and (b), where a representative case in which the object partitioning scheme fails to create a good BVH is shown. The large triangle $\alpha$ induces the algorithm to create a leaf node $A$ that contains the other one $B$. This configuration does not reduce the number of nodes to examine: in fact, the ray depicted in Figure 52 (a) intersects both leaf nodes. Because of the particular configuration of the leaves, it is forced to visit node $A$ firstly and node $B$ secondly, for which the nearest intersection occurs. The dotted line around the node indicates the path of the ray in the BVH (Figure 52 (b)).

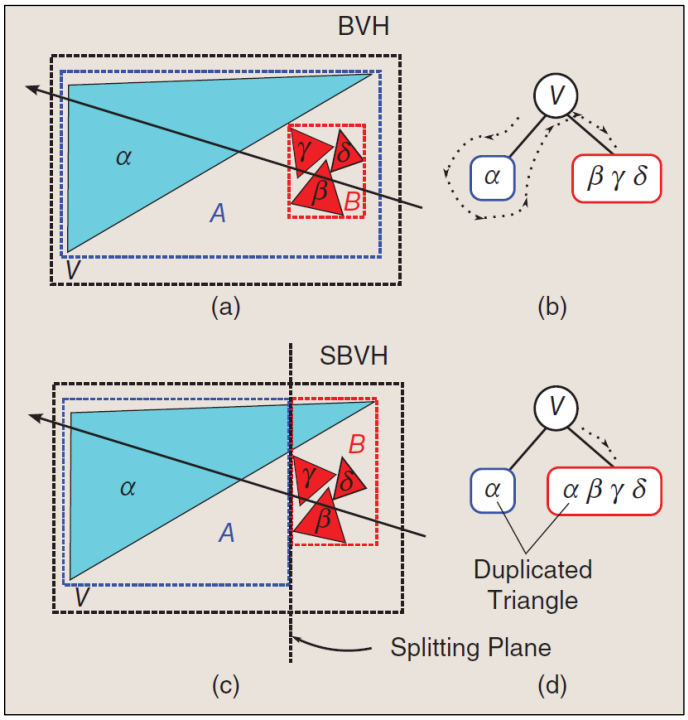



Figure 52: BVH (top) vs SBVH (bottom)

### 2.4.3. Coding the tree-building routine

In terms of numerical implementation, the main difference between the KD-tree (spatial sub-division) and the BVH (object list partitioning) concerns the "FindBestSplit" function. This subroutine accepts as input the structure $P$ in the current step and returns $(P1, B1)$ and $(P2, B2)$ containing the list of references to primitives and the AABBs of the two created partitions. A cost $C$ is introduced according to a criterion as shown in the next paragraph "Surface Area Heuristics". The recursion terminates in three cases:

- the tree has reached the maximum preassigned depth;
- the number of primitives of the parent object is less than the preassigned minimum value;

- if a further subdivision does not significantly improve the tree in terms of costs (Equation 23, which will be addressed in the next paragraph).

The cleverness and so the efficiency of the algorithm resides in the "FindBestSplit" function. The choice of this function directly influences the number of visited nodes and the number of primitive intersection tests. In the next section, simple heuristics that make FindBestSplit effective for both the KD-tree and the BVH are discussed (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

#### 2.4.3.1. Surface Area Heuristic (SAH)

The most studied and most commonly accepted heuristics in ray tracing is called "Surface Area Heuristics" (SAH). It defines a cost function which is a measure of the computational cost of traversing the tree.

SAH defines two constants that govern the construction of the tree, which will be described in the "Stack-Based Traversal" section. The first constant $K_T$ measures the cost to advance one level deeper in the tree while tracing a ray; the second constant $K_I$ measures the cost of the ray-primitive intersection test.

Let us denote a bounding box with $B_0$ and with $B_i$ a sub-box of $B_0$. For spatially uniformly distributed rays and AABBs, it can be shown that the conditional probability that a ray hits $B_i$ once it has hit $B_0$ is:

$$P_{[B_i|B_0]} = \frac{SA(B_i)}{SA(B_0)}$$

Equation 14

where $SA$ is the surface area of the bounding box. Let us suppose that $B_0$ is subdivided in two bounding boxes, namely, $B_1$ and $B_2$. The expected cost $C(B_0)$ to traverse the branch of the tree, starting from $B_0$ and subdivided it into $B_1$ and $B_2$, is the cost of advancing one level deeper in the tree, namely $K_T$, plus the expected cost of intersecting the two children, namely:

$$C(B_0) = K_T + P_{[B_1|B_0]}C(B_1) + P_{[B_2|B_0]}C(B_2)$$

Equation 15

where $C(B_i)$ stands for the cost of the entire child tree enclosed by $B_i$ with $i = 1, 2$. If $B_1$ and $B_2$ are further subdivided into $B_{1,1}$, $B_{1,2}$, $B_{2,1}$, and $B_{2,2}$, then, according to Equation 15, it results:

$$C(B_1) = K_T + P_{[B_{1,1}|B_1]}C(B_{1,1}) + P_{[B_{1,2}|B_1]}C(B_{1,2})$$

Equation 16

$$C(B_2) = K_T + P_{[B_{2,1}|B_2]}C(B_{2,1}) + P_{[B_{2,2}|B_2]}C(B_{2,2})$$

Equation 17

so that Equation 15 becomes:

$$\begin{aligned} C(B_0) = K_T + P_{[B_1|B_0]}K_T + P_{[B_2|B_0]}K_T + \\ + P_{[B_1|B_0]}P_{[B_{1,1}|B_1]}C(B_{1,1}) + \\ + P_{[B_1|B_0]}P_{[B_{1,2}|B_1]}C(B_{1,2}) + \\ + P_{[B_2|B_0]}P_{[B_{2,1}|B_2]}C(B_{2,1}) + \\ + P_{[B_2|B_0]}P_{[B_{2,2}|B_2]}C(B_{2,2}) + \end{aligned}$$

Equation 18

Taking into account that according to Equation 14,

$$P_{[B_X|B]}P_{[B_Y|B_X]} = P_{[B_Y|B]}$$

Equation 19

then Equation 18 simplifies as:

$$\begin{aligned} C(B_0) = {} & K_T + P_{[B_1|B_0]}K_T + P_{[B_2|B_0]}K_T + \\ & + P_{[B_{1,1}|B_0]}C(B_{1,1}) + \\ & + P_{[B_{1,2}|B_0]}C(B_{1,2}) + \\ & + P_{[B_{2,1}|B_0]}C(B_{2,1}) + \\ & + P_{[B_{2,2}|B_0]}C(B_{2,2}) + \end{aligned}$$

Equation 20

When $B_0$ is a leaf, then the cost will be that of intersecting a leaf which is roughly $C_{leaf} = N_l K_l$, where $N_l$ is the number of primitives in the leaf. Accordingly, after having applied Equation 20 to each tree level, the cost of a complete tree can be expressed as:

$$C = \sum_{n \in inner\ nodes} P_{[B_n|B_S]}K_T + \sum_{l \in leaves} P_{[B_l|B_S]}K_l N_l$$

Equation 21

where $B_n$ and $B_l$ are the bounding boxes of node $n$ and of leaf $l$, respectively, and $BS$ is the AABB of the complete scene $S$. The best KD-tree or BVH for a given scene $S$ is the one for which the cost in Equation 21 is minimal.

A global optimization of $C$ against all the possible spatial partitioning of the KD-tree or BVH is currently unfeasible, so, a local greedy approximation is used. In such approximation, the cost of subdividing $B_0$ is computed as if both resulting children were leaves. That means that $C(B_1)$ and $C(B_2)$ in Equation 15 are given values as

they were leaves. So, the expected cost in the current step is approximated such that

$$C(B_0) \approx K_T + K_I\left(P_{[B_1|B_0]}N_1 + P_{[B_2|B_0]}N_2\right)$$

Equation 22

where $N_1$ and $N_2$ are the number of primitives in the two child nodes, respectively. The SAH also gives a criterion to terminate the recursive process: the recursion is terminated when an additional subdivision does not decrease the cost. In other words, the process stops when the cost of dealing with $B_1$ and $B_2$ as leaves is less than or equal to the cost of a further optimal subdivision:

$$K_I(N_1 + N_2) \leq C(B_0)$$

Equation 23

where the left-hand side corresponds to dealing with $B_0$ as a leaf and where $C(B_0)$ is provided in Equation 22.

### 2.4.4. Combined technique: Spatial BVH

Spatial sub-division and object partitioning can be combined to increase the performance of the ray tracing phase, especially when running the algorithm on GPU (Aila & Laine, Understanding the efficiency of ray trasversal on GPUs, 2009) (Aila, Laine, & Karras, Understanding the efficiency of ray trasversal on GPUs - Kepler and Fermi addendum, 2012). The result of this combination is referred as "Spatial BVH" (SBVH) and the core idea is to select the best partitioning scheme based on the SAH algorithm. It employs spatial-subdivision instead of object partitioning in some stages while keeping memory consumption low.

The idea is to select which is the best partition scheme (i.e., object partitioning or spatial partitioning) based on the SAH cost. The cheapest scheme of the two is chosen. Of course, the more spatial subdivisions are performed, the more the SBVH looks like a KD-tree. Thus, spatial splitting is performed only when the expected improvements are significant with respect to an object splitting solution.

In particular, spatial sub-division is used when object partitioning would produce many overlapping bounding volumes, which is non-efficient, as explained previously. When the SBVH is employed, the structure of the tree is the same as the pure BVH, the only difference is the necessity to add the clipping algorithm for the primitives lying over the dividing planes, which is typical of the KD-tree. This does not affect the robustness of the whole algorithm since the spatial sub-division is used just for the first iterations when primitives clipping is an accurate process.

The existing literature (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015) suggests employing the amount of overlap in the child nodes produced by the best object split as a criterion to decide which scheme to prefer. In particular, we compute:

$$k = \frac{SA(B_{child1} \cap B_{child2})}{SA(B_{parent})}$$

Equation 24

where $B_{child1}$ e $B_{child2}$ are the bounding boxes of the two child nodes. Therefore, $k$ is 0 when the two child nodes do not overlap, while it is 1 when the two child nodes completely overlap. When $k$ is greater than a threshold $\alpha$, spatial subdivision is employed. It can be shown that, for the most scenes, a good choice for $\alpha$ is $10^{-5}$. Additionally, when using the SBVH, there is no need for storing the information on whether the

node has been created by space partitioning or object partitioning, since the employed ray traversal algorithm is the same as for the BVH.

Nevertheless, and at variance with BVH, clipping of primitives may be required. Figure 52 (c) and (d) shows the reason why a space subdivision scheme is more effective than object partitioning when there is a large amount of overlapping between nodes. The bounding boxes of the two leaf nodes of the SBVH, created according to the splitting plane in Figure 52 (c), do not nonoverlap and so the ray will only visit node $B$ since node $A$ will provide a farther intersection.

The space subdivision scheme employed in the framework of the SBVH saves even more computational time with respect to the BVH case when node $A$ is the parent of a deep hierarchy. Indeed, in this case, the whole subhierarchy starting from node $A$ is not required to be explored.

## 2.5. Algorithm phase two: determination of the electromagnetic fields

In order to obtain the EM field scattered from the target of interest, and thus the target RCS, the currents induced on the surface of the target were required. To obtain these currents, the incident EM field was necessary. The incident EM field might have been extremely onerous to calculate because of the multiple reflections and contributions caused by the complex geometry of the target of interest (e.g. a jet aircraft). In fact, the GO portion of the algorithm was used to easily obtain a good approximation of the incident EM field on the surface of the target. As the EM field impinged on the surface, currents were induced. In this phase the PO came into play: employing the tangent plane

approximation [12] and assuming that the mentioned impinging field was a plane wave, the induced current could easily be found by means of the vector product between the normal unit vector to the surface in that point and the incident magnetic field associated to the plane wave. Doing so, the induced currents were calculated. At this point, since the interest was only on the Far Field range because of the nature of the RCS concept, the scattered EM field could be calculated by means of an asymptotic approximation of the surface integral of the mentioned currents.

## 2.6. Application of the algorithm to CGI benchmark scenes and real-world scenarios

### 2.6.1. Choice of the acceleration structure based on CGI benchmark scenes

The choice of the data structure (BVH/SBVH vs KD-tree) to accelerate the algorithm running on GPU was based on: computational time, numeric robustness, speed during the construction of the hierarchy, and the suitability to deal with the electromagnetic problem. Known computer graphics benchmark scenes were initially adopted to check that the developed algorithm attained performances comparable to the existing state of the art codes available for GPU.

The purpose was to trace rays representing an electromagnetic wave front radiated from an antenna and reflected or diffracted by scatterers: this kind of rays are somewhat different from those usually dealt with in CGI. In the scenarios considered throughout this research work, the rays were generally “uncoherent”, meaning they could

12 The tangent plane approximation consists in assuming that the local portion of the surface where the EM field is impinging on, can be considered as flat.

traverse the scene following completely different paths. However, since they were often launched in a cylindrical or radial ray congruence or diffracted by corners or edges, they could be generally assimilated to the ambient occlusion rays used in CGI to provide more photo realism.

In fact, as schematized in Figure 53, during this preliminary benchmark test two types of rays were traced: primary rays and ambient occlusion rays, as seen in previous works according to the current literature. Ambient Occlusion is a three-dimensional computer graphics technique that produces realistic images by calculating the amount of diffuse light blocked by the environment for each point of a surface. This technique tries to approximate what happens on cloudy days for diffusive surfaces. The more a point is occluded, the darker it appears.

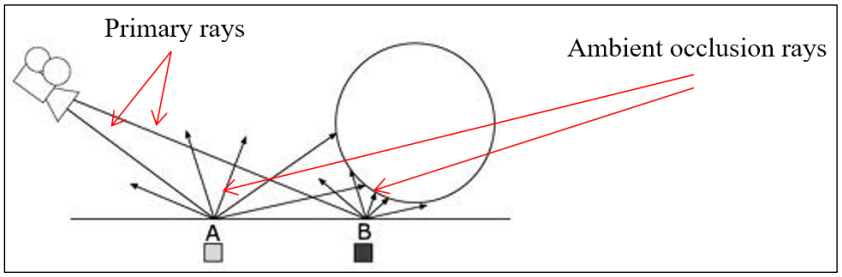



Figure 53: Primary vs Ambient occlusion rays

However, the focus of these research work was not on computer graphics but rather on how ambient occlusion rays were calculated, so that they could be employed for electromagnetic purposes. The most straightforward way to approximate the ambient occlusion of a point was to cast rays from that point in any direction over a hemisphere and test for intersections. Rays that did not hit anything (i.e. un-occluded rays) increased the illumination of the point. The randomness of ray occlusion allowed reliable speed tests of the implemented ray tracers on GPU because it stressed the single-instruction-multiple-threads architecture.

Simple scenes were rendered by tracing primary rays launched from a camera and shading with information of ambient occlusion. In order to assess the quality of both the KD-tree and the BVH approaches for the mentioned scenarios, some commonly used statistics were employed, which were defined by:

$$E_T = \sum_{n \in inner\ nodes} \frac{SA(B_n)}{SA(B_S)}$$

Equation 25

$$E_L = \sum_{n \in leaf\ nodes} \frac{SA(B_n)}{SA(B_S)}$$

Equation 26

$$E_I = \sum_{n \in leaf\ nodes} N_n \frac{SA(B_n)}{SA(B_S)}$$

Equation 27

where $E_T$ was the expected number of visited inner nodes, $E_L$ was the expected number of visited leaf nodes, $E_I$ was the expected number of ray-primitive intersections per ray, $B_n$ was the bounding box of the $n$-th node, and $B_S$ was the bounding box of the whole scene (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

To measure the tool performance against other tools used in computer graphics, the comparison between the KD-tree and the SBVH was accomplished by using both classical CGI benchmark scenes, like “Conference” and “Bunny,” and the “Ship” scenario, which was of mere electromagnetic interest. All the scenes are shown in Figure 54.

Figure 54: Benchmark scenes: Ship, Conference, and Bunny

Table 1 reports several comparative statistics for the two approaches when applied to the known scenarios. The costs *C* (Equation 22) regarding the KD-tree reported in Table 1 agreed with the results available throughout the literature, and sometimes appeared also slightly lower than other existing available results. Furthermore, from Table 1, the costs associated with the SBVH were significantly smaller than those of the KD-tree, a result reported here for the first time in the literature. Also, the KD-tree leaded to a significantly larger number of primitive duplications with respect to the SBVH, which instead leaded to a significantly larger memory occupancy. Additionally, the statistics in Equation 25 and Equation 27 were more favorable for the SBVH rather than the KD-tree.

**OPTIMIZED VALUES OF $K_T = 15$ AND $K_I = 20$**

| | Ship | | Conference | | Bunny | |
|---|---|---|---|---|---|---|
| | KD | SBVH | KD | SBVH | KD | SBVH |
| tris | 447k | 447k | 283k | 283k | 69k | 69k |
| $N_N$ | 254k | 318k | 116k | 270k | 380k | 64k |
| $N_{ref}$ | 1.1M | 540k | 138k | 358k | 326k | 69k |
| dups | 142% | 20% | 389% | 27% | 370% | 0.5% |
| $N_L$ | 127k | 159k | 581k | 135k | 190k | 32k |
| $N_{NE}$ | 77k | 159k | 396k | 135k | 108k | 32k |
| $E_T$ | 42.6 | 24.2 | 35.5 | 24.5 | 55.4 | 26.1 |
| $E_I$ | 16.1 | 9.94 | 10.3 | 11.3 | 14.7 | 5.23 |
| $E_L$ | 11.0 | 2.7 | 9.5 | 9.26 | 7.8 | 2.28 |
| $C$ | 961 | 562 | 765 | 620 | 988 | 495 |

Tris: overall number of triangles. $N_N$: number of nodes (inner nodes plus leaves). $N_{ref}$: number of triangle references. dups: number of duplicated triangles. $N_L$: number of leaf nodes. $N_{NE}$: number of non-empty leaf nodes. $E_T$: expected number of inner-node traversal steps. $E_L$: expected Leaf visits.

Table 1: KD-tree and SBVH statistics (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015)

Table 2 reports the processing speed, in millions of rays per second, for the two approaches. The tests were performed on a workstation equipped with an Intel Xeon E5-2650 2.00 GHz eight-core processors CPU and Nvidia Kepler K20C GPU video-cards. The convenience of the SBVH approach can be easily appreciated from the table (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015).

**THE PROCESSING SPEED OF THE TWO COMPARED APPROACHES**

| | Ship | | Conference | | Bunny | |
|---|---|---|---|---|---|---|
| | KD | SBVH | KD | SBVH | KD | SBVH |
| $P$ | 95 | 360 | 101 | 257 | 77 | 300 |
| $AO_{ave}$ | 52 | 310 | 33 | 95 | 33 | 190 |
| $AO_{min}$ | 18 | 94 | 17 | 76 | 15 | 138 |
| $AO_{max}$ | 243 | 929 | 70 | 131 | 179 | 460 |

$P$: primary rays. $AO_{ave}$: ambient occlusion rays average speed. $AO_{max}$: ambient occlusion rays highest speed. $AO_{min}$: ambient occlusion rays lowest speed. Speeds are in Mrays/s.

Table 2: KD-tree and SBVH processing speed (Breglia, Capozzoli, Curcio, & Liseno, Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs, 2015)

Thanks to these tests, the BVH resulted as the fastest algorithm and was therefore chosen as ray tracing algorithm. Besides, the BVH was faster and more robust especially in the initial construction phase when compared to the KD-tree. In fact, the BVH involved only comparison operations, where the KD-tree required clipping operations for the primitives which could introduce further accuracy problems especially in case of triangles aligned with the reference axes, which were actually frequent when dealing with engineering problems rather than computer graphics. Also, from the tests, it was evident that the KD-tree was better suited for scene with a significant variance in the dimension of the mesh-grid triangles, whereas the BVH gave undesirable results. However, if the spatial variant of the BVH was employed (i.e. SBVH), the results were anyway better than the KD-tree.

### 2.6.2. RCS prediction code in practice

The GO algorithm was adapted to calculate the monostatic RCS of several objects, with an increasing complexity in the geometry of the scenarios, so that the EM field accumulation phase of the tool could be tested. The following steps were required to determine the RCS:

1) the target was illuminated with a plane wave coming from pre-determined directions, which were meaningful directions for those objects which had a real-world relevance, such as the ship;
2) the currents induced by the impinging plane wave on the surface of the PEC target were calculated, considering not only the first reflection contributions of the rays, but also all the contributions given by subsequent reflections of the same ray, until a pre-fixed maximum number of reflections was reached;
3) the scattered EM Far Field was calculated along certain directions of interest by means of a surface integral of those induced currents;
4) the ratio between the scattered Far Field and the incident field gave the Radar Cross Section.

Specifically, the Far Field along the direction of interested was given by:

$$\underline{E}_f = \frac{jk_0}{2\pi r}\, e^{-j\,k_0 r}\, \underline{\hat{E}}_t(0,0)$$

Equation 28

where:

- $k_0$ is the propagation constant;
- $r$ is the distance between the origin of the reference system and the point where the Far Field is calculated;
- $\underline{\hat{E}}_t(0,0)$ is the tangent component of the GO electric field.

## 2.6.3. Electromagnetic results

To evaluate the RCS prediction tool several canonic objects were used, whose results were known in closed form. Additionally, complex scenes were simulated using both the developed tool and some full-

wave algorithms available within the commercial electromagnetic CAD FEKO, allowing a comparison between the results.

#### 2.6.3.1. Perfect Electric Conductor sphere

This test analyzed the scattering of a X-band plane wave linearly polarized along the $x$-axis and propagating along the positive direction of the $z$-axis, impinging on a 20 $\lambda$ radius Perfect Electric Conductor sphere located at the origin of the reference system. Figure 55 depicts the amplitude of the EM field obtained with the GO compared with the true reference described by an expansion of Mie series on a 40 $\lambda$ radius circle.

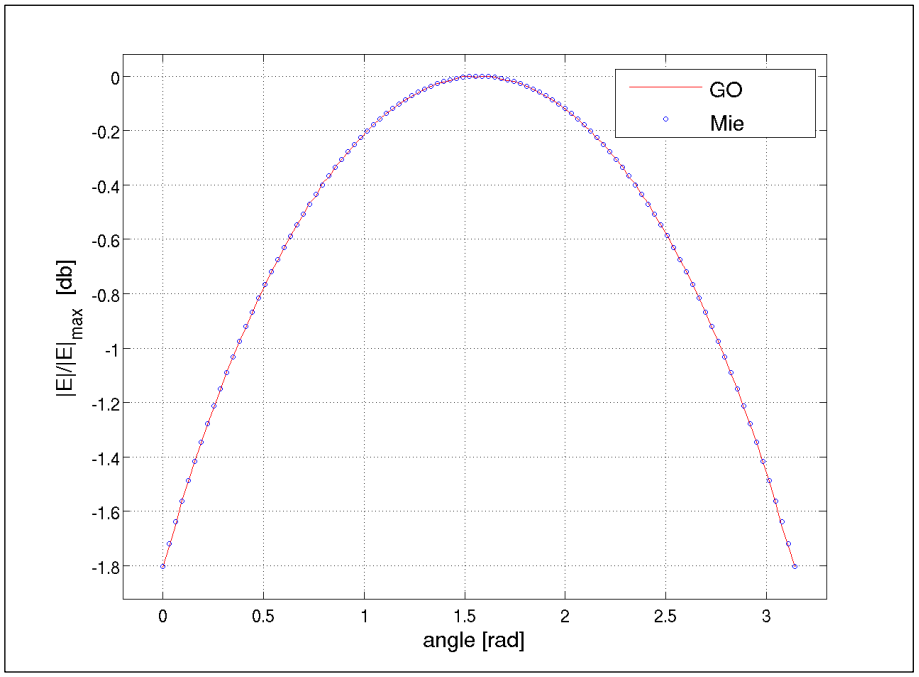



Figure 55: Amplitude of the reflected field from a 20 $\lambda$ radius PEC sphere

### 2.6.3.2. Perfect Electric Conductor cylinder

This test involved an indefinitely long 25 cm diameter PEC cylinder centered at the origin of the reference system and aligned with the $y$-axis, arranged as shown in Figure 56. The cylinder was illuminated by a plane wave at 8 GHz, linearly polarized along the $y$-axis and propagating along the positive direction of the $z$-axis. The Separation Algorithm was used for the accumulation of the EM field. Figure 57 and Figure 58 show respectively the amplitude and the phase of the $y$ component of the reflected EM field collected on a 2-meter long cut located at 2 meters from the cylinder and oriented with a 30-degree angle with respect to the $z$-axis (Figure 56).

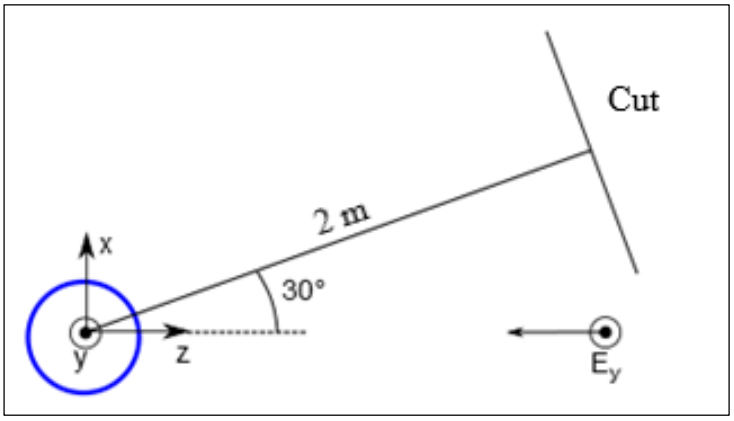



Figure 56: Setup for the cylinder test

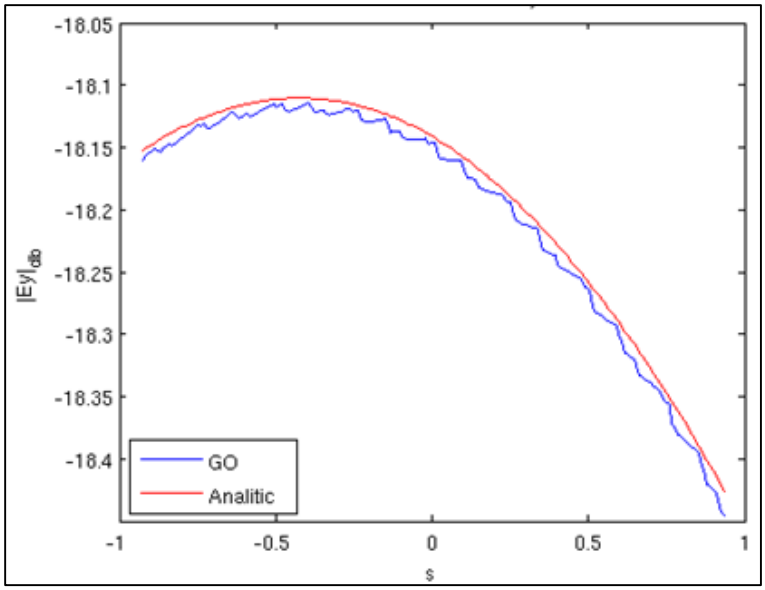



Figure 57: Absolute value of the y component of the electric field

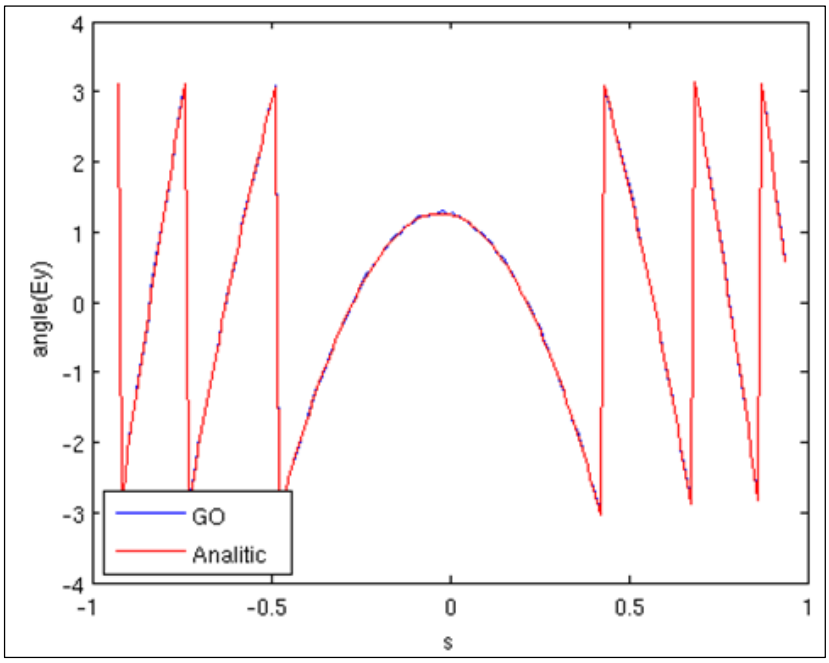



Figure 58: Phase of the y component of the electric field

Another PEC cylinder simulation was conducted at 10 GHz involving a cylinder with $5\lambda$ radius and $10.47\lambda$ height illuminated by a plane wave sampling the whole 360-degree for the variable $\theta$, as depicted in Figure 59. The simulation lasted 4.7 s (13 ms per angle). Normalized Radar Cross Section is shown in Figure 60.

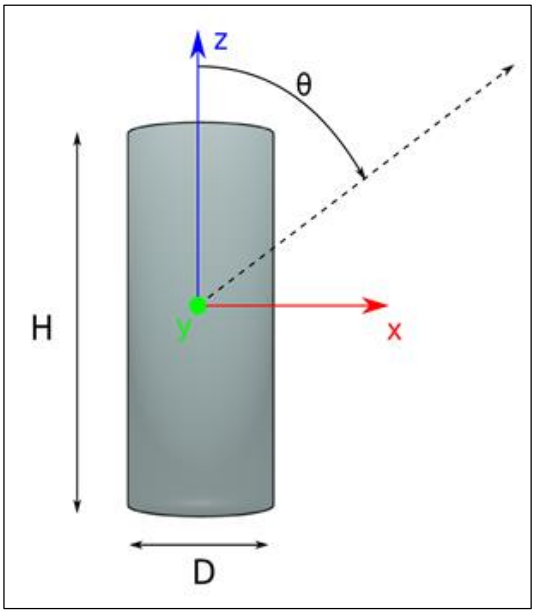



Figure 59: Geometry of the second PEC cylinder simulation

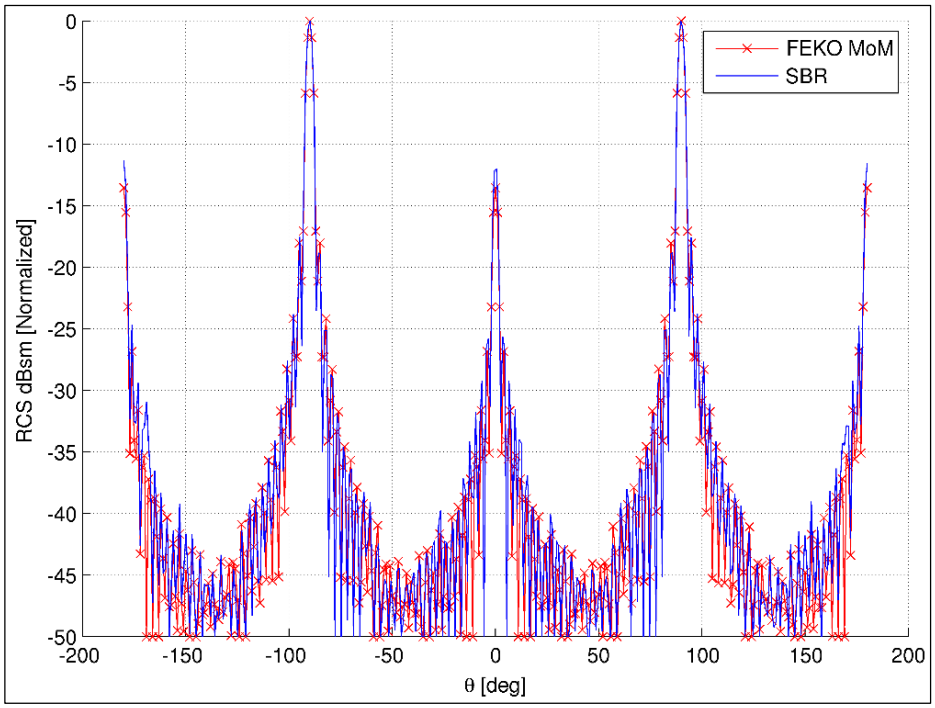



Figure 60: Normalized RCS amplitude for the second PEC cylinder simulation

### 2.6.3.3. Perfect Electric Conductor corner reflector

A standard $5.5\lambda$ side corner reflector was also simulated at 10 GHz focusing on a ±60-degree angle centered on the boresight, for a fixed $\varphi = 54.73°$, as shown in Figure 61. The simulation lasted 4.1 s (34 ms per angle). Normalized Radar Cross Section is shown in Figure 62.

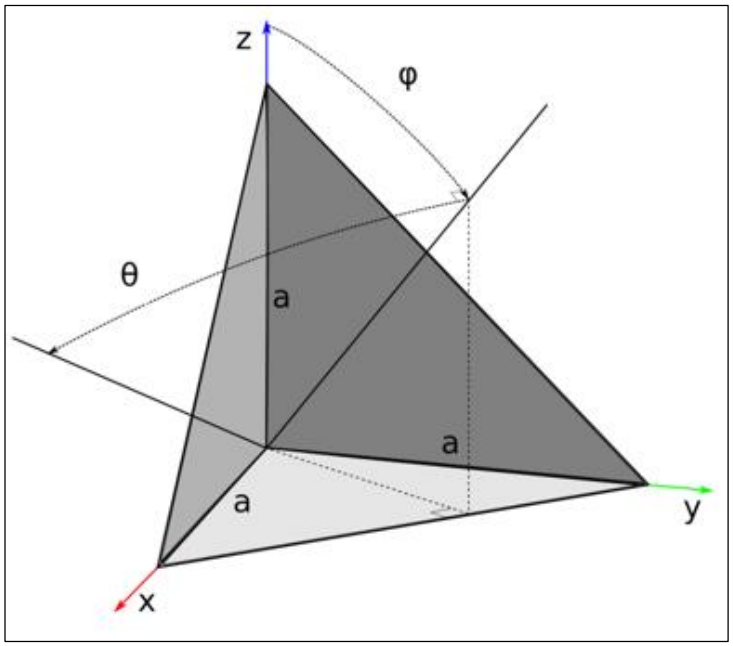



Figure 61: Corner reflector setup

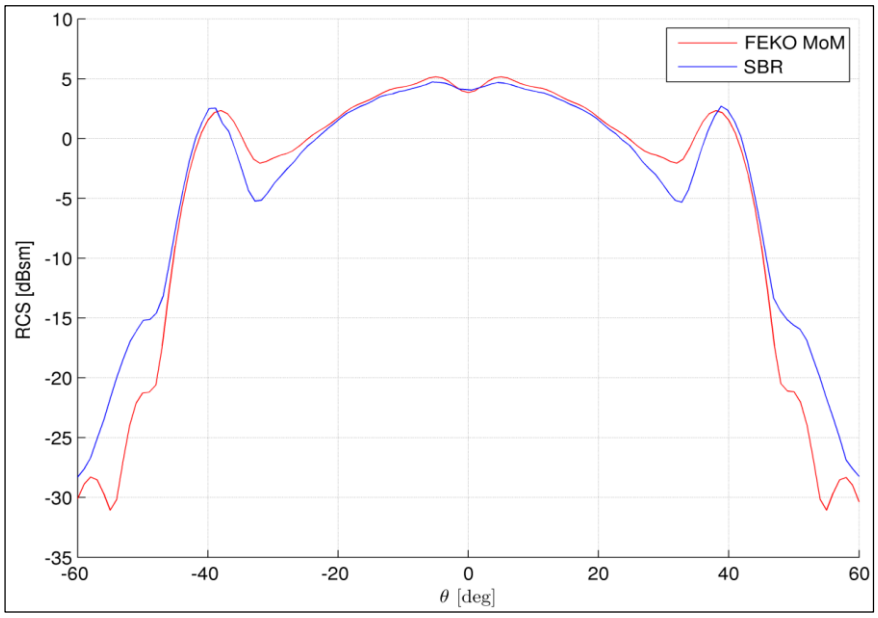



Figure 62: Corner reflector RCS amplitude

#### 2.6.3.4. Perfect Electric Conductor real-world representative target: Ship

As last case of study, a real-word representative target was considered: a 50×8×13 meter ship, discretized in 4477188 PEC triangular primitives, was simulated in several conditions.

Firstly, a 1 GHz Hertzian dipole was placed on top of the ship, as shown in Figure 63, and the absolute value of the scattered electric field was measured on the $x$-$z$ plane, as shown in Figure 64.

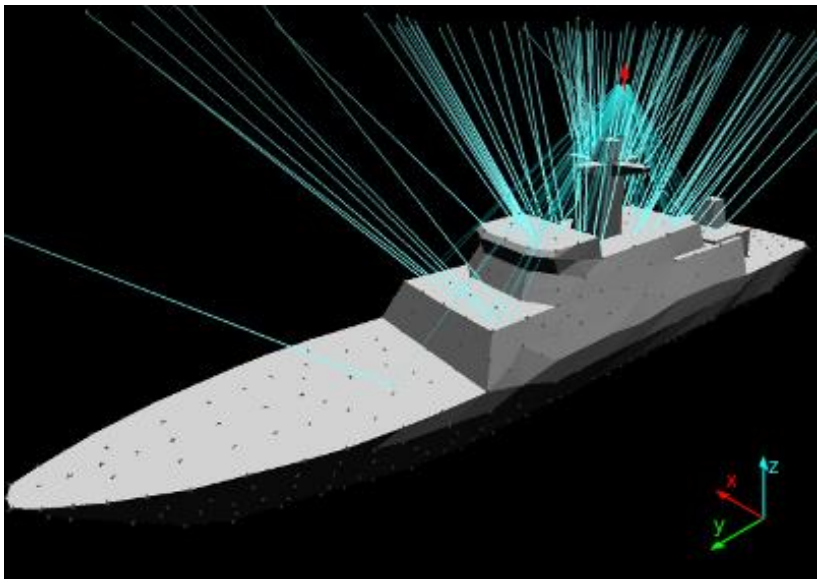



Figure 63: 1 GHz Hertzian dipole on top of the 52 meter long ship

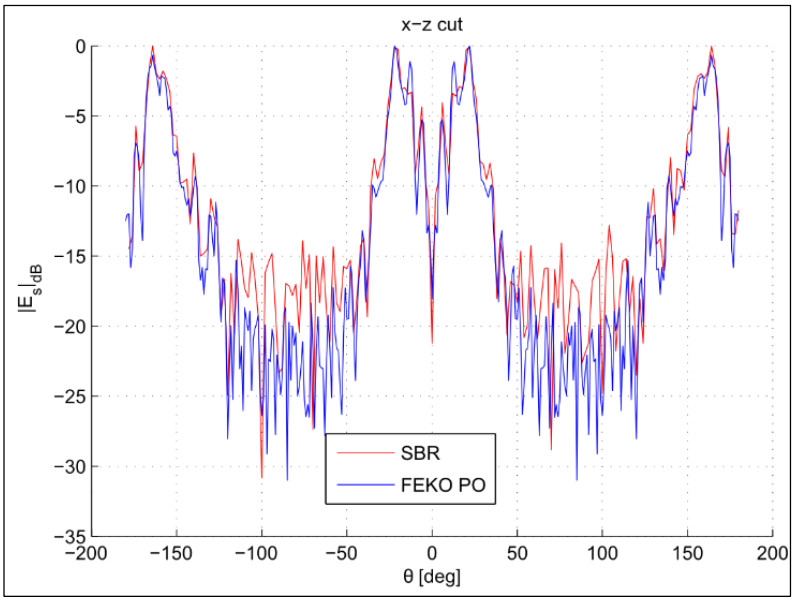



Figure 64: Absolute value of the amplitude of the scattered electric field on $x$-$z$ cut

Secondly, the ship was illuminated with plane waves at 300 MHz (resulting electrical dimensions of the ship: 50×8×13 λ) and 10 GHz (electrical dimensions: 1666×266×433 λ) and the back-scattering RCS calculated on the $x$-$z$ plane and the $y$-$z$ plane with reference to what shown in Figure 65. The results are presented in Figure 66 through Figure 69.

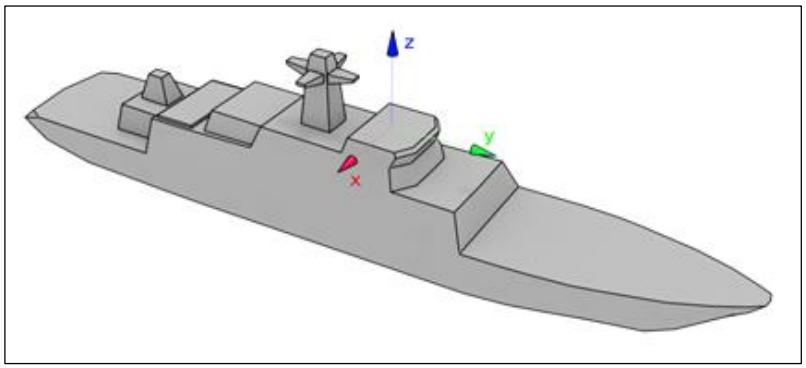



Figure 65: Ship geometry and dimensions (50x8x13 meters)

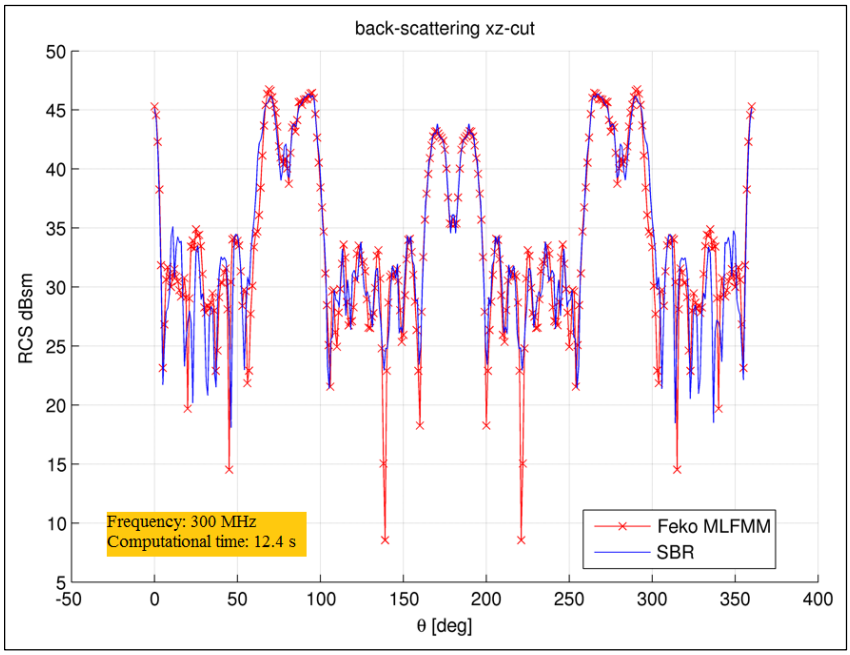



Figure 66: Results at 300 MHz for x-z cut

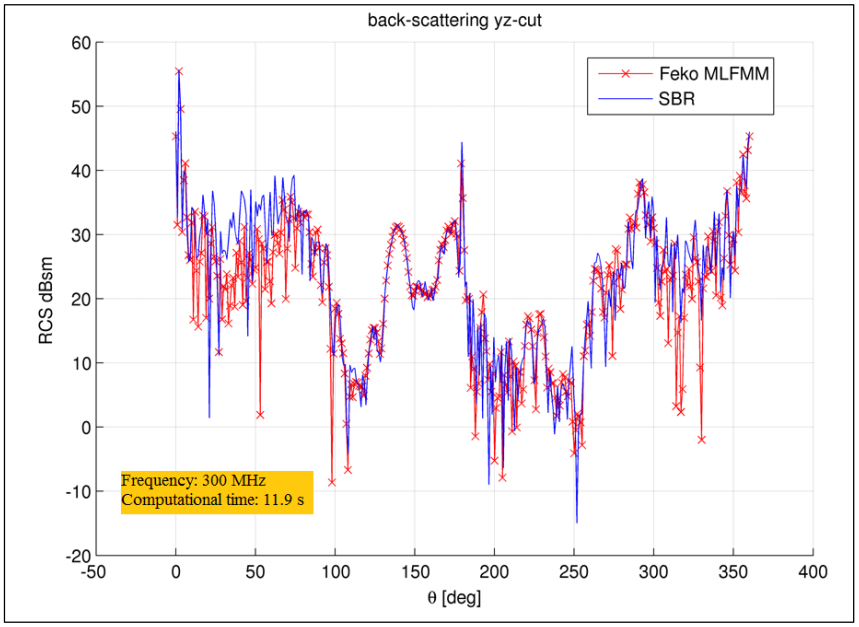



Figure 67: Results at 300 MHz for y-z cut

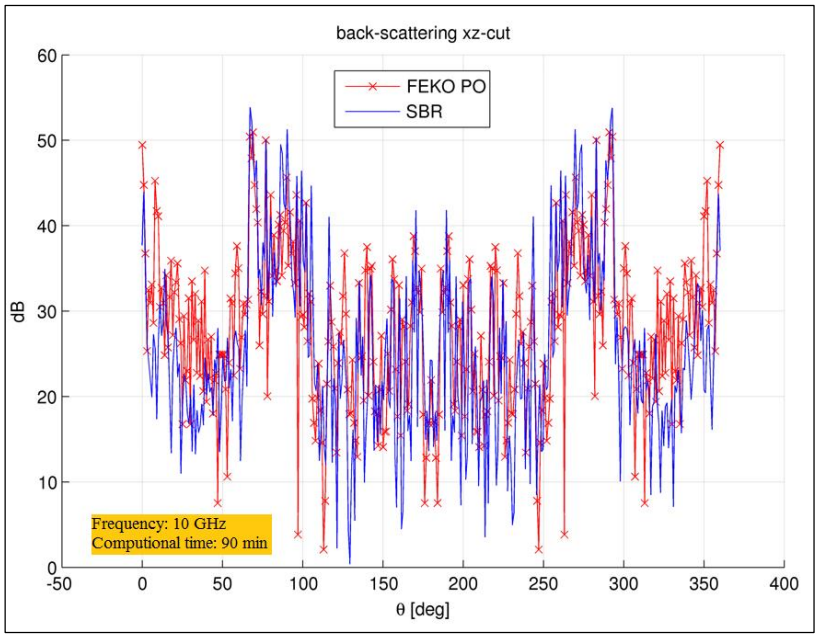



Figure 68: Results at 10 GHz for x-z cut

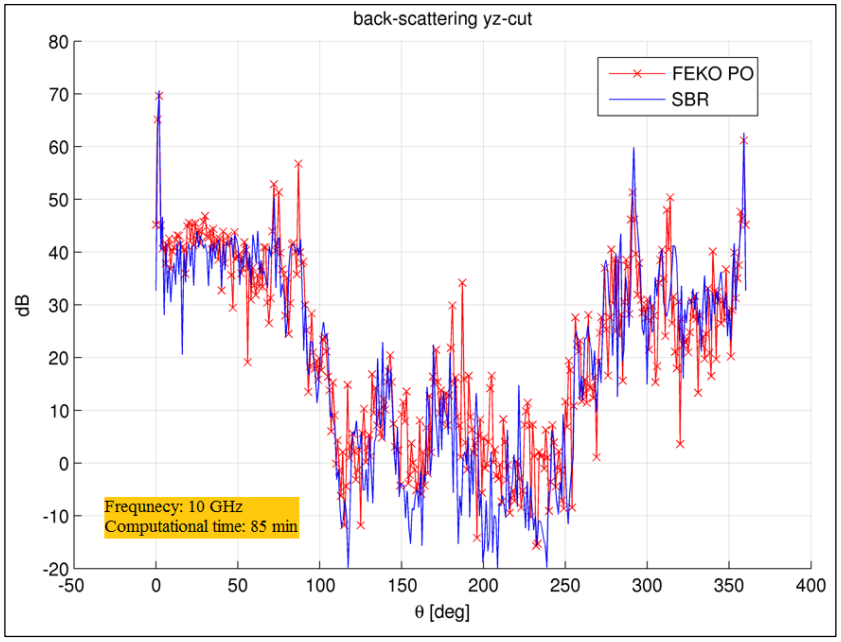



Figure 69: Results at 10 GHz for y-z cut

## 2.7. Measurements

A series of measurements was conducted at the head quarter of the Ingegneria Dei Sistemi (IDS) private company, inside the main semi-anechoic chamber within the premises of the main facility, located in Pisa, Italy. Ingegneria Dei Sistemi is a leader company in the framework of radar applications: in particular, since the ’80s, one of the specific field of business and research has always been both simulation and measurement of the Radar Cross Section of those military targets which are of particular interest within the aeronautical and maritime scenarios.

During the period of time spent at IDS’ facilities, a total of about 40 hours of work, focus was dedicated to the understanding of the simulation software for RCS prediction and IR evaluation of aeronautical targets; after that, a detailed study of the measurement set-ups was performed.

Eventually, to practice with these articulated measurement tool, pre-existing benchmark measurements were repeated, also to be used as reference for the simulation accomplished with the GO/PO tool. Such benchmark measurements involved the PEC sphere, PEC cylinder, and PEC corner reflector.

### 2.7.1. Measurement set-up

The main measuring equipment for static RCS evaluation developed and produced by IDS at the time of this work is a system relying on a large planar scanner (about 10 meters wide and 5 meters tall) which is normally set-up in a very large semi-anechoic chamber (about 30x20x12 meters) (Figure 70) which can also be deployed in a controlled open-field environment (Figure 71).

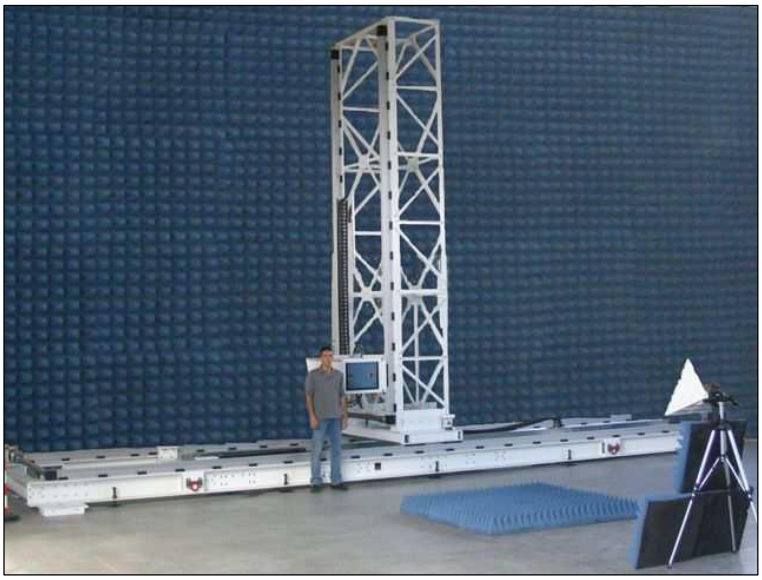


Figure 70: Planar scanner - indoor configuratioin

Figure 71: Planaer scanner – outdoor configuration

The set-up is constituted by a near-field measurement system which employs a Near-to-Far Field conversion and Synthetic Aperture Radar (SAR) imaging to obtain the RCS of a target over a large span of

frequencies (i.e. around 1÷40 GHz) with a peak transmission power around 10 Watts. The system main components are: the RF section (Vector Network Analyzer plus a custom radio frequency pulser) with the test antennas (two identical horn antennas placed side by side), the test antennas positioner (Figure 72), the Device Under Test (DUT) positioner (Figure 73), the data acquisition and control software, and the post-processing software.

Figure 72: RF section: VNA + Pulser (left) and Antennas positioner (right)

Figure 73: Different types of DUT positioners

In general, the samples required for the generation of a typical 2D SAR image are collected in few minutes whereas it takes around 1.5h

for a typical 3D SAR image. Besides 2D and 3D SAR images, high-resolution range profile (i.e. 1D-like) and ISAR imaging are also available among the acquisition modes. Figure 74 shows the steps required to create a RCS plot starting from the very raw samples collected with the planar scanner. However, the system also provides a real-time visualization of the SAR image of the target, allowing for a quick detection of the DUT scattering hot-spots, realized with the same processing tool which elaborates the final image.

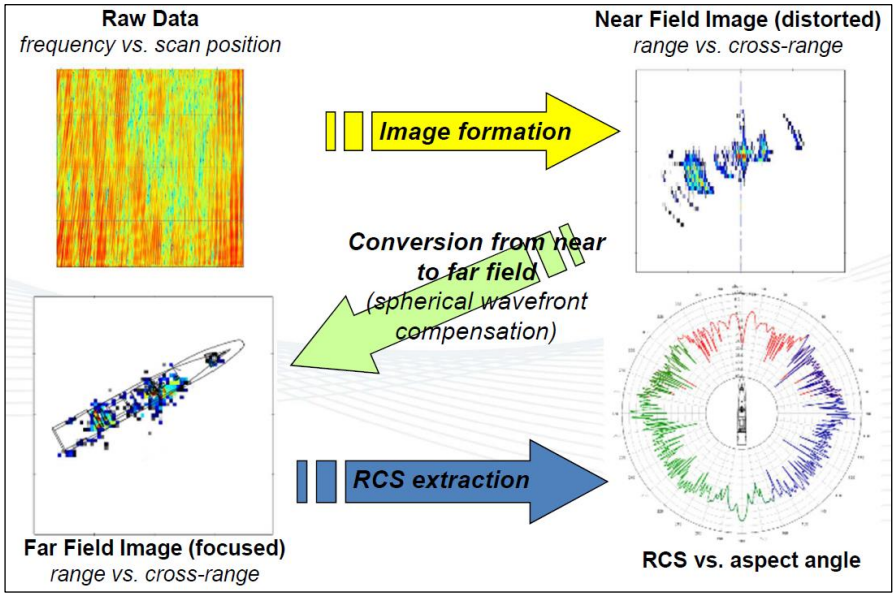



Figure 74: SAR image formation process

IDS is also capable of providing dynamic RCS measurements of high-speed flying targets thanks to a unique system which is basically an extremely sensitve radar: this equipment is called FARAD (Flying Aircraft Radar signature Acquisition and Determination) and is portrayed in Figure 75, where it was set-up in an open-field range. Figure 76 focuses on the RF generation and radiation segment of the system. It is easily transportable and quickly deployable, and the measurements can start within 1 hour from the completion of the set-

up. Additionaly, it is self-sufficient in terms of power supply, which makes it ideal for applications in the aeronautical framework since it does not require excessive logistics contraints.

The FARAD operates in the X band, with a radiated peak power around 125 W and a remarkable sensitivity of -30 dBsm @ 10 km, and it is able to track and measure target moving at a tangential speed up to 250 m/s (roughly 500 knots, a considerable speed even for jet fighter aircraft).

Figure 75: FARAD equipment

Figure 76: FARAD antenna assembly

Using the FARAD against a real target, possibly belonging to the fleet of the Italian Air Force, was a possibility seriously considered throughout the PhD: this kind of activity is extremely complex both on the technical and on the logistic aspect and will probably be accomplished in a dedicated measurement campaign after the completion of the present PhD.

However, the planar scanner was actually used with some canonic targets described in the following to validate the prediction of the monostatic RCS accomplished using the hybrid Geometrical Optics / Physical Optics tool developed during this research.

### 2.7.2. Canonical targets

Some canonic targets whose RCS is known in closed form are normally used by IDS as a reference to validate other measurements. In particular, during the time spent at IDS, four targets were employed as a reference to be compared against the simulations: square plane plate, cylinder, sphere, and corner reflector.

Figure 77 recaps an overall comparison performed using a 16.2 cm side corner reflector as target where the FEKO Physical Optics algorithm, the IDS Physical Theory of Diffraction algorithm, the developed hybrid GO/PO tool are compared against the IDS measurement, accomplished at 10 GHz for the vertical polarization, making the side of the corner reflector 5.4 $\lambda$. As depicted in the figure, within ±30° from the observation direction along the azimuthal plane, the simulations and the measurement show good adherence.

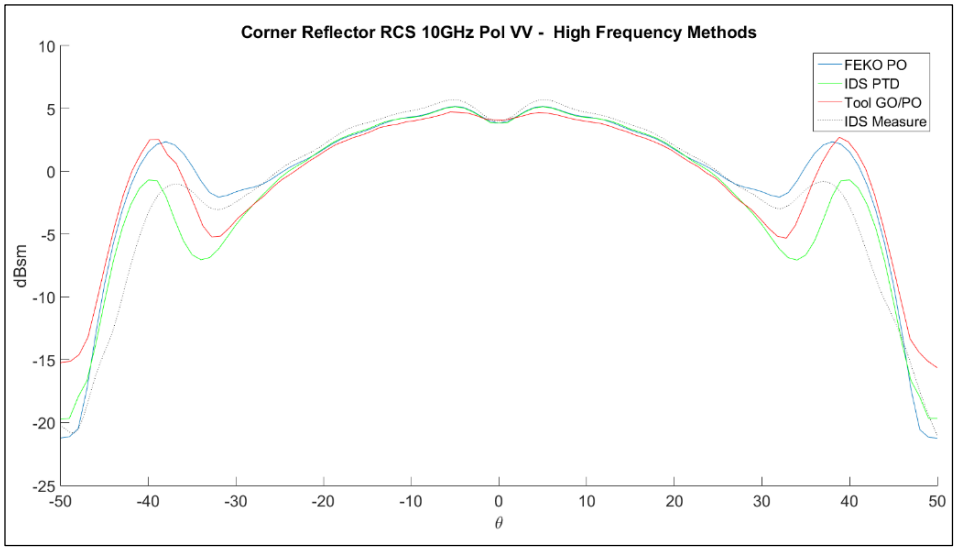



Figure 77: High Frequency methods comparison against IDS measurement

### 2.7.3. Special targets

In addition to the canonic targets, a complex-shape real word representative target was considered: the AT2000 is an all metal fighter jet mock-up which realistically encompasses all the significant characteristics that constitute a target of interest in the aeronautical military framework. Figure 78 shows the configuration used for the measurement inside the semi-anechoic chamber where the planar scanner was employed. Figure 79 gives an overview of the AT2000 RCS measurement process also pointing out the main scattering hot spots of this kind of geometry.

The aim is to focus the final part of the research work simulating the AT2000 RCS under different conditions to stress the capabilities of the developed tool in order to have a good feedback in terms of performance. This is possible since IDS provided the AT2000 3D model and mesh-grid together with previously collected measurement data.

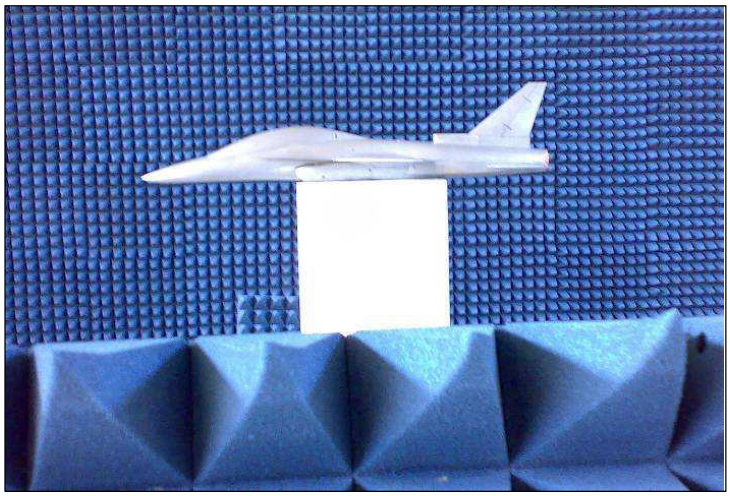


Figure 78: AT2000 mock-up

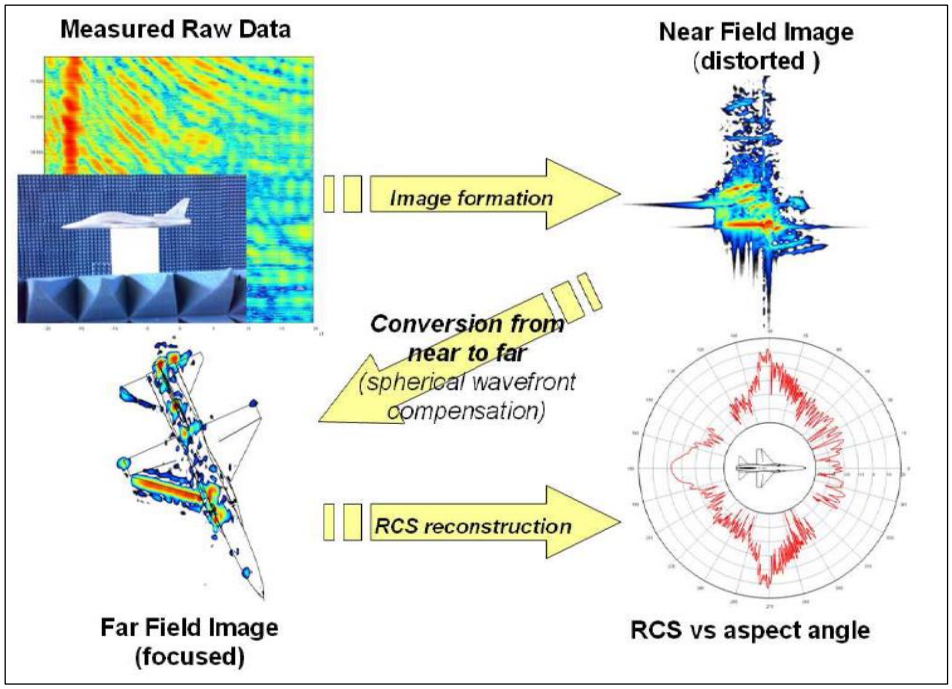



Figure 79: AT2000 RCS measurement

# Chapter 3

# Efficient computing of the Far Field radiation phase by means of the pruned Non-Uniform FFT and the Domain Decomposition technique

Radar Cross Section problems involve incident electromagnetic radiation generated by external sources, creating currents on the scatterer that re-radiate a scattered field. The third Chapter describes the process used to implement an efficient algorithm for the computation of the Far Field scattered by a volumetric scatterer (i.e. a tridimensional object of arbitrary shape).

## 3.1. Scattering scenario

Using the preferred lexicon of the theory of scattering, the specific problem which had to be solved was a "multi-view multi-static" scenario. This situation is of particular interest in the military aeronautical framework: a target aircraft is illuminated from ±50 degrees around the nose on the azimuthal plane and, for each direction of illumination, the scattering was observed within an angular sector of ±30 degrees centered around said direction (Figure 80). The mentioned angles are typical values of interest when dealing with military fighter-bomber jets, but they vary based on particular requirements for a specific geometry.

Similarly, moving along the elevation plane would be necessary as well for a complete analysis: directions of illumination displaced in the interval ±30 degrees above and below the nose should in fact be considered. For sake of simplicity, Figure 80 schematizes only the azimuthal plane geometry.

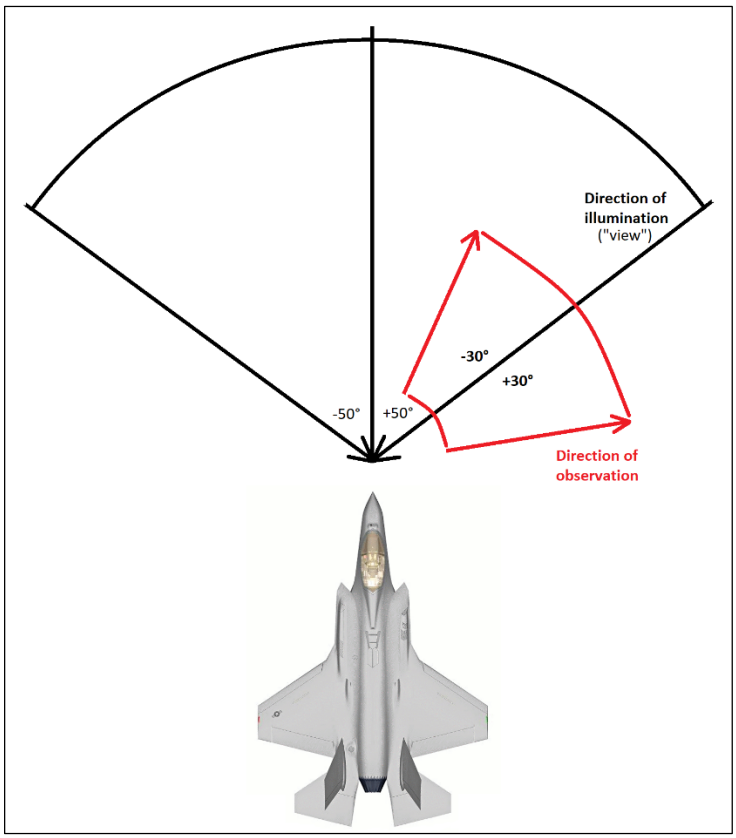



Figure 80: Multi-view multi-static scenario

From an electromagnetic standpoint, this situation is a tridimensional scattering problem where the scatterer is a complex-shape electrically large fast-moving metallic and composite material made target which is illuminated by an impinging plane wave of arbitrary polarization of a frequency belonging to a large interval, such as from 500MHz to 12GHz, since it may be generated by airborne or

ground-based radars. This wave induces currents on the surface of the target which then re-radiate a scattered field. In the process of determining this scattered field, the asymptotic calculation of the radiation integral is necessary. To reduce the computational complexity, the use of algorithms based on the Fast Fourier Transform (FFT) would be particularly convenient. Indeed, Fourier Transform is numerically implemented by the FFT algorithm which has been optimized in many ways through the years.

When dealing with FFTs it is usually assumed that the input and output vectors have the same size and are made of samples which were collected in a uniform fashion (e.g. uniformly-spaced 2D meshgrid). This is just a simple particular case which may be ideal for specific applications, such as the time-frequency transform, but could be definitely not ideal when applied to scenarios such as a 3D scattering problem. In such a complex electromagnetic scenario, the target may have an intricate shape with rapidly changing radii of curvature which demand a non-uniform meshgrid thus a non-uniform sampling of the domain. Moreover, in the transformed domain, the interest may not be uniformly distributed among all the or directions of observation, and, most importantly, the visible domain of the FFT, which corresponds to the real physical scattered Far Field, is calculated only on a small part of the whole transformed domain.

Because of these several reasons, it appeared necessary to implement a Non-Uniform FFT which needed to be non-uniformly sampled both in the non-transformed domain and in the transformed domain.

## 3.2. Introducing the Non-Uniform Fast Fourier Transform

The FFT comes into play when evaluating the radiation integral via the Far Zone approximation: depending on the characteristics of the

object (i.e. PEC or dielectric), the integral can be a surface or a volumetric integral, but in both cases, it will be under the form a Fourier Transform.

Now, when predicting the RCS in the monostatic case the direction of observation is one and only one, namely the same direction from which the illuminating plane wave is coming from. This particular condition simplifies the computation since the Far Field pattern (Equation 29) becomes a sum: in fact, the direction of observation is described only by three scalars which are the components of the unit vector representing that direction in a 3D space. However, since in the monostatic case only the back-scattering direction is observed, if more than one direction of observation needs to be evaluated, every time the direction is changed, the induced currents on the body surface change as well and the radiation integral needs to be re-calculated.

In the multi-static case instead, since there is a finite set of directions of observation which may also be displaced in a non-uniform manner, there will be three vectors containing the triplets of scalars identifying all the directions of interest. Additionally, these directions may be displaced in a non-uniform fashion (i.e. the visible domain is embedded in the set containing the points where the Far Field is calculated via the standard FFT), so that the integral in Equation 29 would result in tri-dimensional Non-Uniform Fast Fourier Transform (NUFFT 3D) when asymptotically evaluated in the Far Field zone. The NUFFT can be numerically approached using a "divide et impera" technique which divides a big set of simple calculations in many sub-sets which are executed in parallel if an appropriate processor is available.

As previously mentioned, the non-uniform sampling of the transformed domain was not only related to the fact that one may be interested only in a certain set of directions of observation, but it was also required because the visible domain of the FFT has a dimension less with respect to the domain on which it is calculated. In fact, in 2D, the visible domain of the FFT is a circumference (i.e. $u^2 + v^2 = 1$

where $u$ and $v$ are the directional cosines of the direction of observation for the scattered Far Field) whereas the FFT output is calculated over a full circle, and in 3D, the visible domain is a spherical surface whereas the FFT output is calculated for all the spherical volume. This must be taken into account in order to alleviate the computational complexity of the algorithm.

There are three types of NUFFTs: Type 1 considers a non-uniformly sampled domain transformed onto a uniformly sampled co-domain, Type 2 considers a uniformly sampled domain transformed a non-uniformly sampled co-domain, and finally the Type 3, where both domains are non-uniformly sampled. In this research, the NUFFT was implemented using the Lee-Greengard version of the algorithm (Lee & Greengard, 2005), in the Type 3 case, where both the non-transformed and the transformed domain are non-uniformly sampled. Since the NUFFT algorithm is inherently complex, a build-up approach was adopted: the code was firstly developed using MatLab, a high-level intuitive programming environment, to later facilitate writing the code also with C++ and CUDA. In fact, CUDA, the proprietary NVidia programming language used to develop GPU routines, was the final objective of this work, whereas MatLab and C++ were intermediate steps.

First off, a Non-Uniform Discrete Fourier Transform (NUDFT) was implemented in MatLab as exact reference for the NUFFT codes about to be realized. The first NUFFT was indeed the one-dimensional version which was immediately compared against the NUDFT until a positive match of the results was reached. Then, a bi-dimensional case of the NUFFT was quickly realized starting from the 1D case. At this point, the 1D and 2D NUFFTs codes were converted in C++, a well-known programming language that was used as intermediate step to avoid going directly from MatLab to CUDA, which is an arduous language to use, if not particularly familiar with it.

At this point, the MatLab and C++ 1D/2D NUFFTs were all tested using short input vectors (i.e. up to 128 or 256 elements) to check for congruency in the results. Done that, a 3D version of the NUFFT was developed in MatLab. This step took a considerable time because of the increased complexity of the iterative routines (i.e. "for" loops) contained inside the code. Once the 3D MatLab version was working correctly, the C++ and CUDA implementations followed. Once that all the 3D codes gave the same exact results, the NUFFT algorithm was considered complete.

Running some preliminary tests with short vectors as mentioned before, despite the good implementation of the algorithm, the 3D version was extremely demanding in terms of required memory to store the matrices involved in the actual calculation of the FFT. This resulted in the code being unserviceable for practical use, considering the expected dimensions of the vectors generated in a scattering problem as the ones described in Chapter 2, involving objects hundreds or even thousands of wavelengths long.

These considerations made obvious the necessity to contain the memory usage: this was implemented exploiting the concept of "pruned FFT", namely an FFT where only a certain portion of the transformed domain is considered, so that the matrices involved in the calculation of the FFT get significantly reduced in size, resulting in a much lower memory occupancy. Analogously, to reduce the size of the non-transformed domain to work with, the concept of "domain decomposition" can be used: employing a geometrical partitioning of the starting domain, only the portion of interest is selected and processed, reducing the required memory storage.

## 3.3. Far Field evaluation by Fourier matrices

The computation of the Far Field radiated/scattered by free-space sources/objects takes place in many areas of applied electromagnetics (Pike & Sabatier, 2002) (Capozzoli A. , Curcio, D'Elia, & Liseno,

2010). Solving this problem by a brute-force approach requires managing a complexity that may be unacceptable for large problems, especially for the 3D case (Boag & Letrou, 2003). The observation that the Far Field of planar (in 3D) or linear (in 2D) radiators/scatterers can be computed by a standard Fast Fourier Transform (FFT) has pushed towards the development of algorithms aiming at reducing the computational complexity. For example, the approach in (Boag & Letrou, 2003) seeks to achieving the same complexity of the FFT, however without the explicit use of the FFT algorithm. As a result, this trend is obtained only when sacrificing accuracy. The problem of computing the Far Field radiated/scattered by 2D volumetric objects can be recapped in three steps (Lee & Greengard, 2005) (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013) (Capozzoli A. , Curcio, Liseno, & Piccinotti, 2017):

1) non-uniformly sampled Fourier exponentials are represented by a finite number of uniformly sampled ones, with an accuracy controlled in terms of number of involved uniform exponentials used;
2) execution of a possibly pruned FFT computation (Sorensen & Burrus, 1993) (Knudsen & Bruton, 1993);
3) interpolation, again with controlled accuracy, of the uniformly sampled exponentials onto the non-uniformly sampled ones which correspond to the relevant directions of observation.

Since the output of the FFT step is only needed at a few output points, the possibility of employing a pruned FFT scheme was considered (Sorensen & Burrus, 1993) (Knudsen & Bruton, 1993) and the consequent mitigation of the number of computations is discussed.

Considering now a 2D radiator/scatterer embedded in free-space as in Figure 81, the source $J$ can be due to primary radiators or be a contrast source in the case of scattering (Abubakar, Hu, Van den Berg, & Habashy, 2008) and its support is assumed to be $S$. Without loss of generality, a $z$-directed current $= J\hat{\imath}_z$ is considered. Then, the only $z$

component of the Far Field Pattern (FFP) $P(\phi)$ is, apart from unessential factors:

$$P(\phi) = \int_S J(\underline{r}')e^{j\beta \underline{r}'\cdot \hat{\imath}_r} dS$$

Equation 29

where $\hat{\imath}_r = (\cos\phi\,, \sin\phi)$ and $\underline{r}' = (x', y')$. Following the application of a quadrature rule (Richmond, 1966), the FFP as evaluated at the discrete angles $\phi_k, k = 1, \ldots, K$ can be written as:

$$P(\phi_k) = F(s,t)\big|_{(s,t)=-(\beta\cos\phi_k, \beta\sin\phi_k)} = \sum_{i=0}^{N-1} w_i J_i e^{-j(x_i' s + y_i' t)}\Big|_{(s,t)=-(\beta\cos\phi_k, \beta\sin\phi_k)}$$

Equation 30

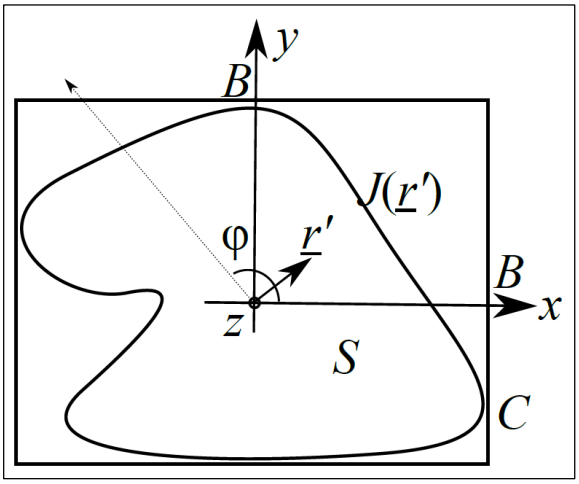



Figure 81: Geometry of the problem

where $w_i$ are the weights of the quadrature.

Therefore there are two grids to deal with: one in the spatial $(x, y)$ plane as defined by the $(x_i', y_i')$'s, and one in the spectral $(s, t)$ plane as

defined by the $(s_k, t_k) = -(\beta \cos\phi_k\,, \beta \sin\phi_k)$'s. Equation 30 can be recast as a matrix-vector multiplication $\underline{P} = \underline{\underline{\mathrm{A}}}\,\underline{f}$, where $\underline{P}$ is the vector of the $P(\phi_k)$s, $\underline{f}$ is the vector of the $w_i J_i$s and the elements of the $K \times N$ matrix $\underline{\underline{\mathrm{A}}}$ are the $exp[-j(x_i' s_k + y_i' t_k)]$'s. Matrix $\underline{\underline{\mathrm{A}}}$ resembles, but is not in the form of, a Fourier matrix, namely a matrix $\underline{\underline{\mathrm{F}}}_M$ whose generic $(ijpq)$-th element is $\omega_M^{ijpq}$, $\omega_M$ being equal to $exp[-j2\pi/M]$, namely, a root of unity.

The efficient computation of the FFP amounts thus at the efficient calculation of a matrix-vector multiplication, whose complexity strongly depends on the structure of the matrix. Fortunately, when $\underline{\underline{\mathrm{A}}}$ has peculiar characteristics (e.g. Vandermonde matrix or Fourier matrix), the complexity can be significantly improved. Morgenstern's theorem (Morgenstern, 1973) in the $c$-restricted computational model has been a cornerstone result in algebraic complexity theory (Burgisser, Clausen, & Shokrollahi, 1997), stating that the complexity associated to Fourier matrices arising from 1D problems of size $L$ be not less than $(L/2)\log_c L$. Accordingly, in the case of a Fourier matrix $\underline{\underline{\mathrm{F}}}_M$ the asymptotic complexity drops to $M^2 \log M$. Therefore, recasting the calculation in terms of a matrix-vector multiplication involving a Fourier matrix becomes convenient.

The problem with Equation 30 is that $\underline{\underline{\mathrm{A}}}$ is not in the form of a Fourier matrix, so that reformulating the problem by interpolating non-uniformly sampled exponentials by uniformly sampled ones is in order. This can be achieved by the Poisson formula (Trigub & Belinsky, 2004):

$$e^{-j\xi x} = \sqrt{2\pi}\,\frac{\sum_{m\in\mathbb{Z}} \mathcal{F}\left[\Phi(\xi)e^{-j\xi x}; m\right] e^{jm\xi}}{\sum_{m\in\mathbb{Z}} \Phi(\xi + 2m\pi)\, e^{-j2m\pi x}}$$

Equation 31

where $\Phi$ is an appropriate interpolation window and $\mathcal{F}$ denotes the Fourier transformation. Accordingly, a computational scheme analogous to a Type 3 NUFFT procedure (Lee & Greengard, 2005) (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013) can be set up. Such procedure is illustrated in the following by assuming the window functions $\Phi$ to be Gaussian. This choice is motivated by the availability of bounds concerning the maximum errors pertaining the uniformly discretized operator mapping functions in the $(x, y)$ domain onto functions in the $(s, t)$ domain (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013).

### 3.3.1. Procedure step #1

The contributions from non-uniformly spaced input sampling points corresponding to $exp[-j(s_k x_i + t_k y_i)]$ are spread by Gaussian windows $exp\left[-j(x^2/(4\tau_x) - y^2/(4\tau_y))\right]$ with parameters $\tau_x$ and $\tau_y$, to a regular grid $(n\Delta x, m\Delta y)$. Step #1 thus produces (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013):

$$f_\tau^{-\sigma}(n\Delta x, m\Delta y) = \frac{e^{\left[\sigma_x (n\Delta x)^2 + \sigma_y (m\Delta y)^2\right]}}{\sqrt{4\sigma_x \sigma_y}} \sum_{i=0}^{N-1} f_i e^{-\left[\frac{(n\Delta x - x_i)^2}{4\tau_x} + \frac{(m\Delta y - y_i)^2}{4\tau_y}\right]}$$

Equation 32

with $f_i = w_i J_i$ and where the presence of the exponential function $exp\left[\sigma_x x^2 + \sigma_y y^2\right]$ is related to the pre-compensation of the Gaussian window which will be used in Step #3.

Due to the rapid decay of the exponential functions, $f_i$ significantly contributes to only few samples of $f_\tau^{-\sigma}(n\Delta x, m\Delta y)$. On defining $Int[\alpha]$ as the nearest integer to $\alpha$, by letting $\xi_i = Int[x_i/\Delta x]$ and $\eta_i = Int[y_i/\Delta y]$, $i = 0, \dots, N-1$, denoting the nearest regular grid points to $x_i/\Delta x$ and $y_i/\Delta y$, respectively, and assigning $n' = m - \eta_i$, the contributions of each $f_i$ to $f_\tau^{-\sigma}(n\Delta x, m\Delta y)$ can be ignored when $|n'| >$

$m_{sp}$ or $|m'| > m_{sp}$, where $m_{sp}$ is a parameter properly selected according to the required accuracy. In other words, the summation in Equation 32 can be truncated to $(2m_{sp}+1) \times (2m_{sp}+1)$ terms. This step is illustrated in Figure 82, where the available current sample locations are denoted by red crosses, the black empty circles represent the regular $(n\Delta x, m\Delta y)$ points and the blue filled circles represent the spreading due to Equation 32 of the $(x_i', y_i')$ onto the $(n\Delta x, m\Delta y)$ grid.

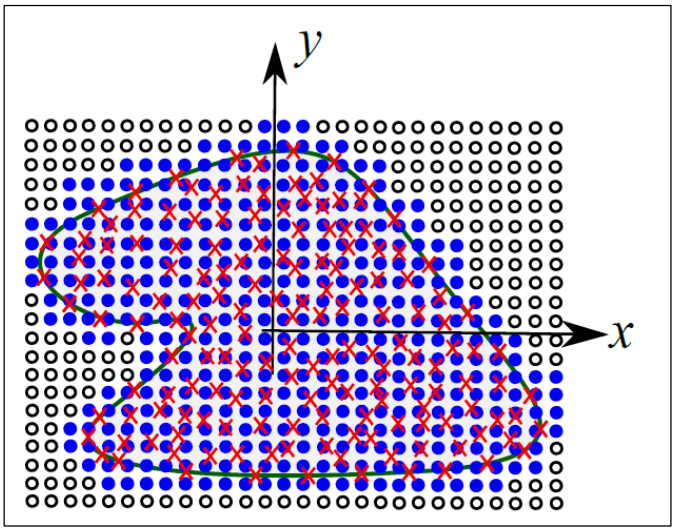



Figure 82: Illustrating Step #1

### 3.3.2. Procedure step #2

The spread contributions are transformed to the spatial frequency domain via a standard FFT thanks to the discretized version of the operator mapping the $(x, y)$ plane into the $(s, t)$ domain. In other words, the second step produces:

$$F_{\tau}^{-\sigma}(p\Delta s, q\Delta t) \simeq \frac{\Delta x \Delta y}{2\pi} \sum_{n=-M_{r_x}/2}^{M_{r_x}/2-1} \sum_{m=-M_{r_y}/2}^{M_{r_y}/2-1} f_{\tau}^{-\sigma}(n\Delta x, m\Delta y) \times e^{-jpn\Delta x\Delta s} e^{-jqm\Delta y\Delta t}$$

Equation 33

The FFT then allows evaluating $F(s,t)$ in Equation 30 at the sampling points $(p\Delta s, q\Delta t)$, i.e. the black circles in Figure 83.

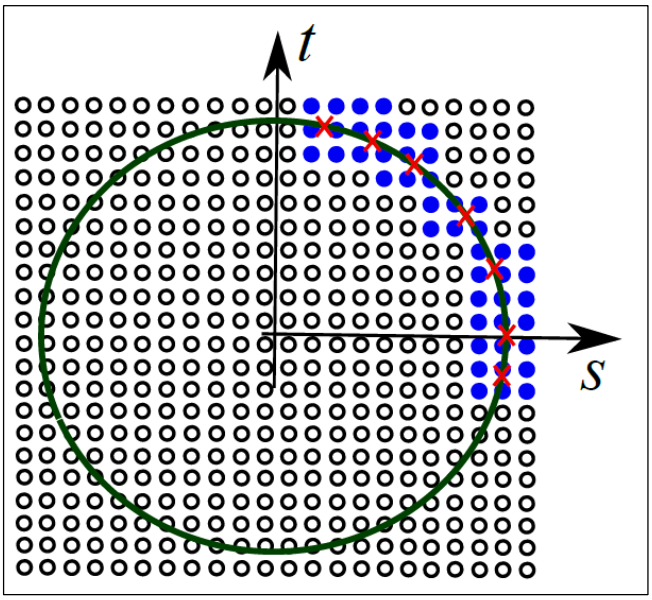



Figure 83: Illustrating Step #3

### 3.3.3. Procedure step #3

The transformed data are interpolated from the FFT output uniform grid to the non-uniform grid $\{(s_k, t_k)\}_{k=0}^{K-1}$, again by Gaussian windows, $exp\left[-s^2/(4\sigma_x) - t^2/(4\sigma_y)\right]$. The final output is thus:

$$P(\phi_k) = F(s_k, t_k) = \frac{\Delta s \Delta t}{4\pi\sqrt{\tau_x \tau_y}} e^{\tau_x s_k^2} e^{\tau_y t_k^2} \sum_{n=-M_{r_x}/2}^{M_{r_x}/2-1} \sum_{m=-M_{r_y}/2}^{M_{r_y}/2-1} F_\tau^{-\sigma}(n\Delta s, m\Delta t) e^{-\frac{(n\Delta s - s_k)^2}{4\sigma_x}} e^{-\frac{(m\Delta t - t_k)^2}{4\sigma_y}}$$

Equation 34

Similarly to Step #1, the presence of the Gaussian functions $exp\left[\tau_x s^2 + \tau_y t^2\right]$ is related to the post-compensation of the Gaussian windows used in Step #1. Again, due to the rapid decay of the involved exponential functions, $F_\tau^{-\sigma}(p\Delta s, q\Delta t)$ significantly contributes to only few samples of the $F_\tau(s_k, t_k)$. In particular, on letting $\widetilde{\xi_k} = Int[s_k/\Delta s], k = 0, \dots, N_s - 1$ and $\widetilde{\eta_k} = Int[t_k/\Delta t], k = 0, \dots, K - 1$ and $p' = q - \widetilde{\xi_k}$ and $q' = q - \widetilde{\eta_k}$, the contributions of $F_\tau^{-\sigma}(p\Delta s, q\Delta t)$ can be ignored when $|p'| > m_{sp}$ or $|q'| > m_{sp}$, where $m_{sp}$ is a parameter properly selected according to the required accuracy. In other words, the summation in Equation 34 can be truncated to $\left(2m_{sp} + 1\right) \times \left(2m_{sp} + 1\right)$ terms. This step is illustrated in Figure 83, where the red crosses represent the sampling points at which the FFP is required, while the blue filled circles represent those regular grid points contributing to the value of the FFP samples of interest.

### 3.3.4. Centering and choice of the relevant parameters

Before applying the abovementioned procedure, a centering of the input and output sampling points is required. Similarly, for the choices of $\Delta x$, $\Delta y$, $\tau_x$, $\tau_y$, $\sigma_x$, $\sigma_y$ and $m_{sp}$ see (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013) and Table 3. In Table 3, $R$ was chosen strictly larger than 2, and $X$, $Y$, $S$, and $T$ were chosen as follows:

$$X = \max_{\cdot}\{|x_i'|\}_{i=0}^{N-1}$$
$$Y = \max_{\cdot}\{|y_i'|\}_{i=0}^{N-1}$$
$$S = \max_{\cdot}\{|s_k|\}_{i=0}^{K-1}$$
$$T = \max_{\cdot}\{|t_k|\}_{i=0}^{K-1}$$

Following the "centering" step, $m_{sp} = 2\pi b$, $b$ is chosen according to successive approximations of the following equation:

$$b = \frac{1}{\gamma}\log\left(\frac{4\alpha}{e}b + \frac{9\alpha}{e}\right), \alpha = 2 + \frac{1}{\sqrt{2\pi}}, \gamma = \pi^2\left(1 - \frac{2}{R^2}\right)$$

Equation 35

where $e$ is the requested accuracy (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013).

| | |
|---|---|
| $\Delta x = \frac{\pi}{RS}$ | $\Delta y = \frac{\pi}{RT}$ |
| $\Delta s = \frac{2\pi}{\Delta x M_{r_x}}$ | $\Delta t = \frac{2\pi}{\Delta y M_{r_y}}$ |
| $M_{r_x} \geq 2\left(\frac{XSR^2}{\pi} + 2\pi Rb\right)$ | $M_{r_y} \geq 2\left(\frac{YTR^2}{\pi} + 2\pi Rb\right)$ |
| $\tau_x = b\Delta x^2$ | $\tau_y = b\Delta y^2$ |
| $\sigma_x = b\Delta s^2$ | $\sigma_y = b\Delta t^2$ |

Table 3: Summary of the parameters choice (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013)

### 3.3.5. "Optimality" of the approach

Concerning the evaluation of Equation 30 in terms of a matrix-vector multiplication, see the linear computational model described in (Burgisser, Clausen, & Shokrollahi, 1997). In this respect, Winograd's theorem provides an evaluation of the computational complexity which amounts to be $K(2N-1)$ for a "generic" rectangular $K \times N$ matrix.

According to Morgenstern's theorem (Morgenstern, 1973), it is expected that, as long as the computation is rearranged in terms of Fourier matrices, the complexity can be significantly reduced. An estimate of the complexity reduction is now in order.

The input sample locations (red crosses in Figure 82) are available after the sampling step employed for their calculations, typically non-uniform, which can be in the order of, say, $\lambda/10$, as a result of the numerical solution of Maxwell's equations in a scattering case. Opposite to that, the sampling steps $\Delta x$ and $\Delta y$ arising from the above scheme to get an accuracy up to machine precision (in double precision arithmetics) of the Far Field radiation operator are in the order of $\Delta x \sim \Delta y \sim \lambda/4$ according to the formulas in (Capozzoli A. , Curcio, Liseno, & Riccardi, 2013). Consequently, if we assume that the radiator/scatterer is contained within the minimum box sides $2B \times 2B$ (Figure 81[13]), then $K \simeq 20B/\lambda$ (Chew, Wang, Otto, Lesselier, & Bolomey, 1994) and $N \simeq (20B/\lambda)^2 \simeq K^2$, so that $K(2N-1) \sim O(K^3)$.

On the other side, the standard FFT step above costs $L^2 \log L$ with $L = 8B/\lambda \simeq K/2$, while Steps #1 and #3 cost $O(N) = O(K^2)$ and $O(K)$, respectively.

In conclusion, the matrix-vector multiplication costs $O(K^3)$, while the proposed approach costs $O(K^2 \log K)$ to get the FFP with machine precision. For the purposes of Steps #3, the FFT samples are required only at the blue filled circles of Figure 83. Accordingly, a pruned FFT is a further possibility to save computations (Sorensen & Burrus, 1993) (Knudsen & Bruton, 1993).

13 We suppose that the radiator/scatterer essentially fits a square box. In the opposite case, a domain decomposition into rectangular boxes can be fruitfully exploited for an efficient hierarchical computation.

## 3.4. Pruned FFT algorithm

The computational saving of pruned FFTs depends on the required output samples pattern. For 1D FFTs of length, say $W$, when only the first few $V$ output samples are required, pruning can reduce the computational complexity from $O(W \log_2 W)$ to $O(W \log_2 V)$ (Sorensen & Burrus, 1993).

In the 2D case (Byun, Park, Sun, & Ko, 2016), similar results can be obtained for some specific patterns. Unfortunately, for the output samples pattern in Figure 83, the asymptotic computational complexity is less favorable, keeping $O(W^2 \log_2 W)$. To roughly estimate it, a surface approach is used.

To this end, the output pattern in Figure 83 can be approximated by a circular annulus of radius $B$ (expressed in terms of power-of-two number of samples[14]) and width $\Delta B$, where $\Delta B$ is related to the spreading $2m_{sp} + 1$. An upper bound for the annulus surface is $2\pi B \Delta B$. This surface is proportional to the number of output active butterflies of the radix $2 \times 2$ computational tree[15] (Byun, Park, Sun, & Ko, 2016). The number of overall radix $2 \times 2$ computational stages is $\log_2(2B)$, while an upper bound for the number of pruned stages can be easily calculated by assuming that, tracing back the tree, the number of active butterflies quadruplicates. Accordingly, the number of pruned stages is $0.5 \log_2(2B/(\pi \Delta B))$. An estimate of the computational saving is then well approximated by:

$$\frac{1}{2} - \frac{1}{2} \frac{\left[\log_2(\pi \Delta B) + \frac{2}{3}\right]}{\log_2(2B)}$$

Equation 36

14 More sophisticated schemes dealing with non-power-of-two number of samples can be exploited.

15 Approaches more efficient than radix 2x2 that however do not change the asymptotic complexity are known, but are outside the scope of this estimation.

Equation 36 shows that the saving is expected to logarithmically approach a 50% value.

## 3.5. Numerical results

The complexity and the accuracy of the proposed scheme are now assessed. For all the results, a customized version of the 2D pruned radix $2 \times 2$ "Decimation-In-Frequency" FFT algorithm was implemented where the memory occupancy was reduced by means of in-place calculations. The computational burden was evaluated in terms of number of performed multiplications, rather than giving particular focus on computation time. This is because the timing performance of actual implementations may strongly depend on software/hardware factors, like memory latencies or proper exploitations of the cache memories and computation pipelines, which are beyond the scope of this contribution, as in (Frigo & Johnson, 2005).

In Figure 84, the radiation by 2D volumetric sources having circular cross section with radius $a$, and radii comprised $5\lambda$ and $70\lambda$ was considered. The sources were discretized into $N = (2\beta a)^2$ points whereas the FFP was calculated in $K = 2\beta a$ points. Figure 84 then shows the number of multiplications required by the FFP computation. As it can be seen, the operations count grew as $K^2$, instead of $K^2 \log K$. This was due to the predominance of Step #1 for the considered sizes of the volumetric sources and to the fact that the complexity of Step #1 grew as $O(K^2)$. Such a predominance was, in turn, due to the employed Gaussian windows.

Other kinds of significantly more compact windows could be used, while achieving the same accuracy while drastically reducing $m_{sp}$ to values of about 3 or 6 for single or double precision, respectively, instead of 18. Nonetheless, note that single precision may be already satisfactory in many applications. Work along this direction is in

progress. From the bottom panel of Figure 84 illustrating the saving due to pruning, a 25% saving, coherent with Equation 36, is observed.

To assess the accuracy, the following scattering scenario was considered: a plane with unit amplitude, travelling in the positive direction of the x-axis, impinged on a homogeneous circular cylinder of radius $a = 5\lambda$ and relative dielectric permittivity $\varepsilon = 2.1$. The cylinder was discretized by a triangular mesh (Persson & Strang, 2004) of side equal to $\lambda/8$. Figure 85 shows the satisfactory agreement between the FFP as evaluated by the proposed approach and that computed by an exact evaluation of Equation 30. The root mean square error between the two calculations was $2.654 \cdot 10^{-11}$, practically approaching machine accuracy in double precision arithmetics.

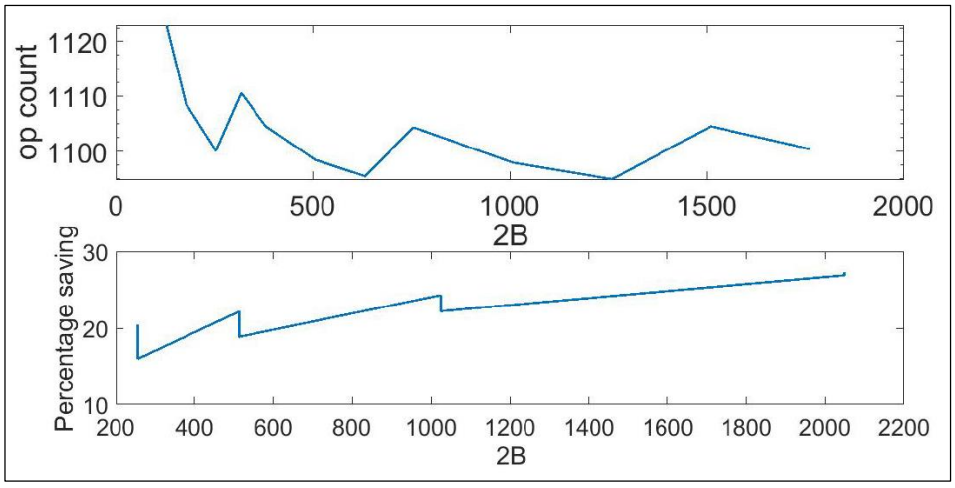



Figure 84: Computational performance. Upper panel: number of multiplication operations normalized by $N$ against $\sqrt{N}$. Lower panel: percentage operation saving due to the use of a pruned 2D FFT.

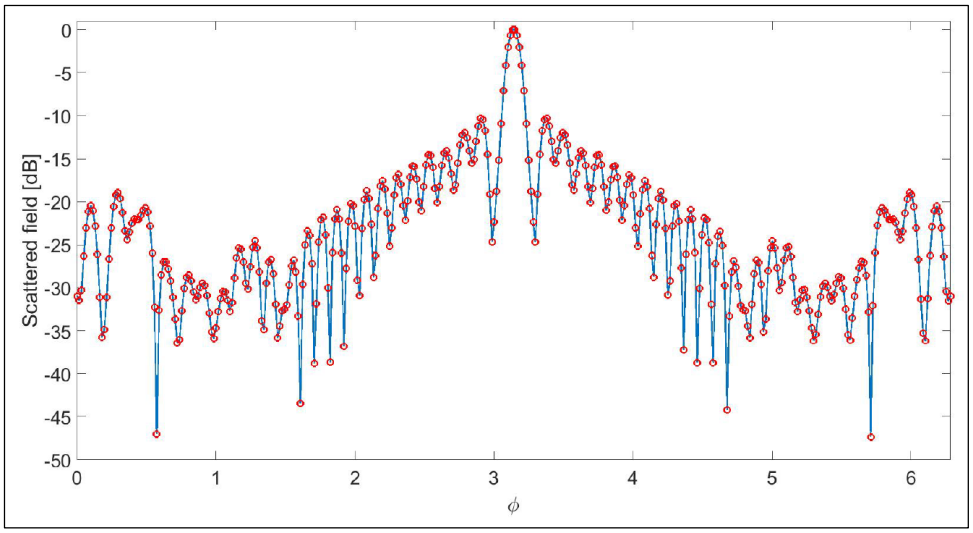



Figure 85: FFP scattered by a homogeneous dielectric cylinder with radius $a = 5\lambda$ and relative permittivity $\varepsilon = 2.1$ under unit plane wave incidence travelling along the positive direction of the x-axis. Blue solid line: proposed approach. Red circles: exact evaluation of Equation 30.

## 3.6. Future development: Domain Decomposition

As the pruning technique aims at reducing the computational complexity by limiting the transformed domain size, similarly the Domain Decomposition technique operates on the non-transformed domain.

The DD technique decomposes a large problem into several coupled sub-problems which adapt better to the original geometry, thus reducing the overall required domain. These sub-problems are then independently solved, and all the solutions combined in order to reach the global solution. Hence, the DD approach provides a considerable reduction in memory storage requirements and computational time. Domain Decomposition is particularly reliable when applied to a geometry whose subdomains are almost independent one from another, meaning that none of the subdomains experiences any considerable field return from other subdomains (Ozgun & Kuzuoglu, 2008).

When applying the DD, how to mesh the domain has to be wisely taken into account: if a global mesh is generated before applying the DD, then each sub-domain will have its sub-mesh with no problem associated; if instead the global domain is so large that requires the DD to be applied before the meshing, then what happens is that two neighboring sub-domain will have different meshes at their interface, resulting in the so-called “non-conforming subdomains”. To solve this unwanted situation, special DD techniques have to be developed which properly couple the sub-domains (Xue & Jin, 2015).

The described DD technique may be a valuable tool for future developments specific for the NUFFT algorithm, namely that portion of the GO-PO tool dedicated to the scattered Far Field radiation phase by means of the currents surface or volumetric integral. In fact, the considered 3D geometries such as planes and vessels are definitely electrically very large so that processors memory becomes a real bottleneck to deal with.

# References

Abubakar, A., Hu, W., Van den Berg, P., & Habashy, T. (2008). A finite-difference contrast source inversion method. *Inverse Problems, vol. 24, n. 6*, 1-17.

Aila, T., & Laine, S. (2009). Understanding the efficiency of ray trasversal on GPUs. *Proceedings of the Conference on High Performance Graphics*, (pp. 145-150). Saarbucken, Germany.

Aila, T., Laine, S., & Karras, T. (2012). *Understanding the efficiency of ray trasversal on GPUs - Kepler and Fermi addendum.* NVidia Technical Report.

Alexander, D. (2004). *Stealth Warfare, 1st ed.*

Boag, A., & Letrou, C. (2003). Fast radiation pattern evaluation for lens and reflector antennas. *IEEE Transactions on Atennas and Propagation, vol. 51, n. 5*, 1063-1068.

Breglia, A., Capozzoli, A., Curcio, C., & Liseno, A. (2015). Comparison of Acceleration Data Structires for Electromagnetic Ray-Tracing purposes on GPUs. *IEEE Antennas and Propagation Magazine* , 159-176.

Breglia, A., Capozzoli, A., Curcio, C., Liseno, A., & Piccinotti, J. (2015). GPU implementation of hybrid GO/PO BVH-based algorithm for RCS predictions. *2015 IEEE International Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting.*

Buddendick, H., & Eibert, T. (2010). Acceleration of ray-based Radar Cross Section predictions using the monostatic-bistatic equivalence. *IEEE Transaction on Atennas and Propagation, vol. 58, n. 2*, 531-539.

Burgisser, P., Clausen, M., & Shokrollahi, M. (1997). *Algebraic Complexity Theory.* Berlin: Springer-Verlag.

Byun, K., Park, C., Sun, J., & Ko, S. (2016). Vector Radix 2 OE 2 Sliding Fast Fourier Transform. *Mathematics Problems in Engineering, vol. 2106*.

Capozzoli, A., Curcio, C., D'Elia, G., & Liseno, A. (2010). Fast phase-only synthesis of conformal reflect-arreys. *IET Microwave Antennas and Propagation, vol. 4, n. 12*, 1989-2000.

Capozzoli, A., Curcio, C., Liseno, A., & Piccinotti, J. (2017). Efficient computing of Far-Field radiation in two dimensions. *IEEE Antennas and Wireless propagation letters, vol. 16*, 2034-2037.

Capozzoli, A., Curcio, C., Liseno, A., & Riccardi, A. (2013). Parameter selection and accuracy in Type 3 Non-Uniform FFTs based on Gaussian gridding. *Progress in Electromagnetic research, vol. 142*, 743-770.

Chaudhury, B., & Chaturvedi, S. (2005). Three-dimensional Computation of Reduction in Radar Cross Section using Plasma shielding. *IEEE Transactions on Plasma Science*, (pp. Vol. 33, No. 6).

Chew, W., Wang, Y., Otto, G., Lesselier, D., & Bolomey, J. (1994). On the inverse source method of solving inverse scattering problems. *Inverse Problems, vol. 10, n. 3*, 547-553.

Collin, R. (1985). *Antennas and Radiowave Propagation.* Singapore: International Student Edition, Mc Graw Hill.

Frigo, M., & Johnson, S. (2005). The deisgn and implementation of FFTW3. *Proceedings of the IEEE, vol. 46, n. 2*, 216-231.

Gao, P., Tao, Y., & Lin, H. (2013). Parallel shooting and boucing ray method on GPU clusters for analysis of electromagnetic scattering. *Progress in Electromagnetics Symposium*, (pp. 87-99).

Havran, V. (2001). *PhD Thesis on Heuristic Ray Shooting Algorithms.* Technical University of Prague.

Hema, S., Simy, A., & Mohan, J. (2016). *Plasma-based Radar Cross Section reduction, 1st ed.*

Jones, J., & Thurber, M. (1989). *Stealth Technology - The Art of Black Magic, 1st ed.*

Kline, M., & Kay, I. (1965). *Electromagnetic Theory and Geometrical Optics.* New York: Interscience Publisher, John Wiley & Sons.

Knott, E. (1985). A Progression of High-Frequency RCS Prediction Techniques. *Proceedings of the IEEE, vol. 73, n. 2*, 252-264.

Knott, E., Shaffler, J., & Tuley, M. (2004). *Radar Cross Section, 2nd ed.*

Knudsen, K., & Bruton, L. (1993). Recursive pruning of the 2D DFT with 3D signal processing applications. *IEEE Transactions on Signal Processing, vol. 41, n. 3*, 1340-1356.

Kouyoumjian, R. (1965). Asymptotic High-Frequency methods. *Proceedings of the IEEE*, 864-875.

Lee, J., & Greengard, L. (2005). The Type 3 Non-Uniform FFT and its applications. *Journal of Computational Physics, vol. 206, n.1*, 1-5.

Marchetto, A., Mercurio, A., Migliozzi, G., Piccinotti, J., & Risoldi, G. (2015). nEUROn UCAV: Develpment and Operational Assessment Campaign. *NATO STO - SCI 269.* Ottawa.

Morgenstern, J. (1973). Note on a lower bound on the linear complexity of the Fast Fourier Transform. *Journal of the ACM, vol. 20, n. 2*, 305-306.

Ozgun, O., & Kuzuoglu, M. (2008). A non-iterative Domain Decomposition method for Finite Element analysis of 3D electromagnetic scattering problems. *Antennas and Propagation Society International Symposium.* San Diego.

Persson, P., & Strang, G. (2004). A simple mesh generator in MatLab. *SIAM Review, vol. 46*, 329-345.

Pike, R., & Sabatier, P. (2002). Scattering: scattering and inverse scattering in pure and applied science. *Academic Press*.

Poisel, R. (2002). *Introduction to Communication - Electronic Warfare, 1st ed.*

Popov, S., Unther, J., Seidel, H., & Slusallek. (2007). Stackless KD-tree trasversal for high performance. *Proceedings of Eurographics, vol. 26, n. 3*, pp. 415-424.

Richmond, J. (1966). TE-wave scattering by a dielectric cylinder of arbitrary cross-section shape. *IEEE Transactions on Atenennas and Propagation, vol. 14, n. 4*, 460-464.

Schmitz, A., Rick, T., Karolski, T., Kuhlen, T., & Kobbelt, L. (2011). Efficient rasterization for outdoor radio wave propagation. *IEEE Tansaction on Visual Computer Graphic, vol. 17, n. 2*, 159-170.

Skolnik, M. (2008). *Radar Handbook, 3rd ed.*

Sorensen, H., & Burrus, C. (1993). Efficient computation of the DFT with only a subset of input or output points. *IEEE Transactions on Signal Processing, vol. 41, n. 3*, 1184-1200.

Stich, M., Friedrich, H., & Dietrich, A. (2009). Spatial splits in bounding volume hierarchies. *Proceedings of the Conference on High Performance Graphics*, (pp. 7-13). Saarbrucken, Germany.

Sun, R., Xie, J., & Zhang, Y. (2016). Simulation research of band-pass Frequency Selective Surface radome. *Symposium on Progress in Electromagnetic Research (PIERS).* Shangai.

Sweetman, B. (2001). *Lockheed Stealth, 1st ed.*

Tao, Y., Lin, H., & Bao, H. (2010). GPU-based shooting and bouncing ray method for fast RCS prediction. *IEEE Transaction on Antennas and Propagation, vol. 58, n. 2*, 494-502.

Trigub, R., & Belinsky, E. (2004). *Fourier Analysis and Approximation of Functions.* Dordrecht: Springer Science and Business Media.

Wachter, C., & Keller, A. (2006). Instant ray tracing: the Bounding Interval Hierarchy. *Proceedings of the 17th Eurographics Symposium on Rendering Techniques*, (pp. 139-149). Cyprus.

Weinmann, F. (2006). Ray tracing with PO/PTD for RCS modeling of large complex objects. *IEEE Transactions on Antennas and Propagation, vol. 54, n. 6*, 1797-1806.

Xue, M., & Jin, J. (2015). Robust Domain Decomposition methods for modeling of large phased arrays. *31st International Review of Progress in Applied Computational Electromagnetics* . Williamsburg.