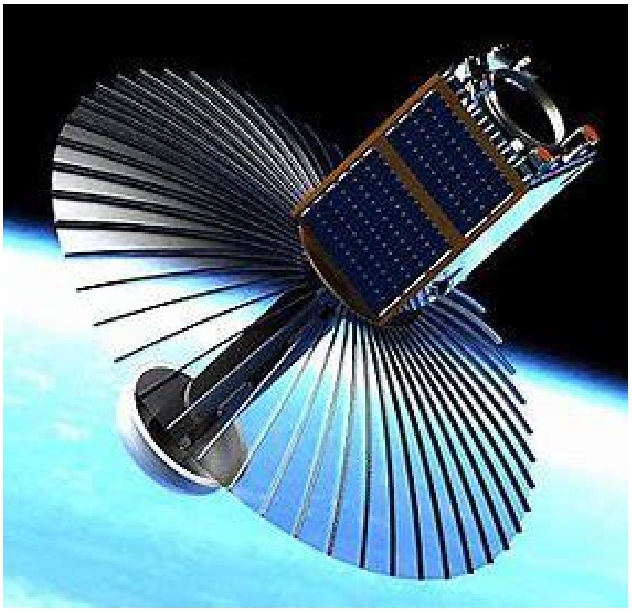
SAPIENZA
UNIVERSITÀ DI ROMA

Control of Communication and Energy Networks
Professor Delli Priscoli



Synthetic Aperture Radar attitude control using Backstepping controller and Deep Reinforcement Learning

Shayan Ahmadi - 1957130
Email : ahmadi.1957130@studenti.uniroma1.it

Ali Nosouhi Dehnavi - 1950716
Email: nosouhidehnavi.1950716@studenti.uniroma1.it

June - 2022

Contents

-

1 Introduction to Synthetic Aperture Radar

1.1 SAR mission Objectives

Space-based Synthetic Aperture Radar (SAR) is a remote sensing technique that is capable of providing 2-dimensional or 3-dimensional reconstructed images of target locations on the ground and has numerous remote sensing and mapping applications for Earth's surfaces. SAR is an active observation technique in which a huge satellite antenna transmits and collects microwaves simultaneously. As the emitted microwaves reflect off the ground, the echoes received contain amplitude and phase information that can be utilized to reconstruct images.
A coherent combining of the received signals permits the creation of a virtual aperture that is significantly longer than the physical antenna length. This fundamental characteristic of SAR is the origin of its name, "synthetic aperture," and gives it the imaging radar property.[2]
As SAR performs utilizing microwaves, it is able to conduct imaging throughout both the day and the night and can penetrate cloud cover and poor weather conditions that would render standard Earth Observation (EO) techniques ineffective. The surface area of the antenna is directly proportional to the possible resolution of ground-based imaging targets. Therefore, SAR missions have often been done with heavier spacecraft, such as the 2300 kg RADARSAT-2 from the Canadian Space Agency. On these platforms, it is possible to install a larger antenna than on a smaller satellite. In recent years, however, the number of SAR missions utilizing spacecraft weighing less than 100 to 200 kilograms has increased, as small satellite platforms offer a significant decrease in construction costs. In addition, the use of small satellites makes the construction of huge constellations, which can attain rapid worldwide coverage and revisit imaging areas within a few hours, more financially possible. Small satellite SAR missions are currently being conducted by companies such as ICEYE and Capella Space, who also offer their customers rapid imaging services. To deploy huge X-band antenna systems, they use satellite platforms weighing 85 kg and less than 100 kg, respectively. When operating in Spotlight mode, ICEYE and Capella Space are capable of image resolutions of 0.5 meters. Independent SAR payloads for small spacecraft are now being manufactured. EOS SAR offers a 50 kg payload with a 2 - 7 m antenna capable of obtaining a ground resolution of 1 m.

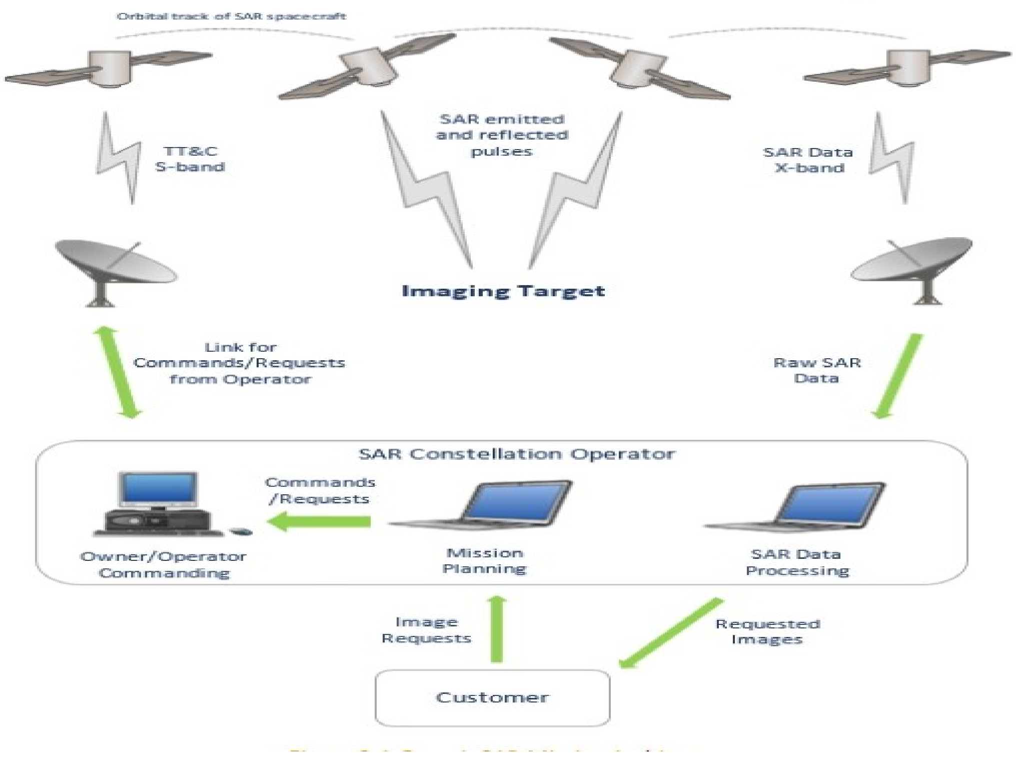


Figure 1: Generic SAR Mission Architecture

1.2 Mission Architecture

S-band and X-band transceivers are typically used for telemetry, command and control (TT & C), and payload data down-link in SAR missions. Therefore, communication with the spacecraft is possible from a variety of ground stations across the world. When a user requests imagery of a specific area or locations on Earth, mission analysis is performed to determine when a satellite in the constellation will next fly over that target. The customer is then provided with a scenario for the acquisition of the images for approval, as the quality of the images produced is dependent on a number of parameters, including imaging duration, area coverage, and the cross-track electronic beam steering required to scan the target. Upon client approval, a file containing the upcoming imaging events can be transferred to the appropriate spacecraft during its next ground station connection. The spacecraft can then perform SAR imaging of the specified location and downlink payload data via X-band at the subsequent ground station interaction. The encoded and processed SAR data is then used to generate the images requested by the user. Before being made available to the customer, the images are subjected to a quality check to ensure that they satisfy their specifications. This mission architecture is depicted in Figure1

1.3 Satellite Architecture

Figure 2 depicts an top-level diagram of the satellite architecture necessary for a SAR mission.This figure provides a basic overview of the types of equipment a SAR spacecraft would typically require, as detailed architectures of existing spacecraft are not readily available online. SAR missions require a highly stable platform for their Spotlight mode in which the deployed antenna electronically directs its beam to a specific location. Therefore, Attitude and Orbit Control Sub-system(AOCS) components including inertial measurement units (IMUs)/gyroscopes (GYRO), star trackers (ST), and reaction wheels (RW) are required to attain 0.1° pointing precision and 0.25 arcsec/s stability . Magnetorquers (MTQs) would also be deployed to facilitate desaturation of the reaction wheels and detumbling of the spacecraft following launch vehicle separation or while exiting safe mode.In addition, a separate payload data store is included. This is consistent with the OSSAT's existing plan, according to which payload data will not be stored in the main platform data storage. Following acronyms are used in the : Low-noise amp (LNA) and high-power amp (HPA). Figure 5-1 also includes the following extra abbreviations: attitude and orbit control system (AOCS), on-board computer (OBC), power management system (PM), transmitter (Tx), receiver (Rx), transceiver (TRx), sun sensor (SS), and magnetometer (MTM).

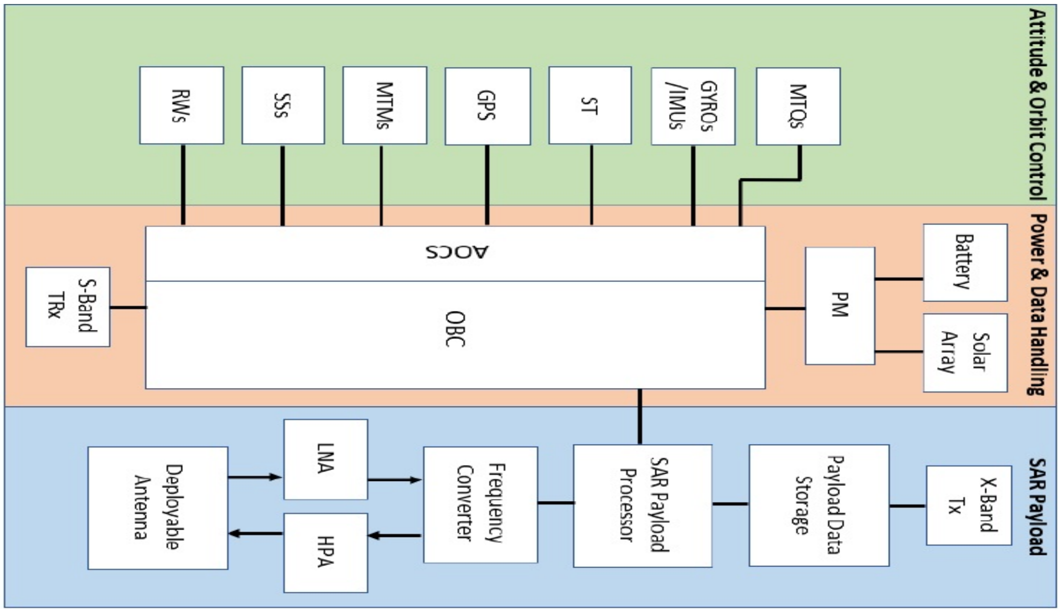


Figure 2: Top level satellite architecture

1.4 Attitude and Orbit control

- Accurate attitude control is essential for operation in the Spotlight mode. This mode requires directing and maintaining the antenna beam at the imaging location to increase the illumination time on the target. Current SAR payloads use electronic beam steering to do this tracking maneuver instead of slewing the satellite. This technology still requires precise attitude control to maintain the spacecraft's steady orientation while the antenna beam is electrically guided.
- Even when employing electronic beam steering, the Spotlight mode requires a pointing precision of 0.1°. York Space Systems' platform for ICEYE has an aiming precision of 10-36 arcsecs. This platform also delivers 0.25 arcsec/s of stability. In the absence of an antenna pointing device, perfect pointing will be necessary for the X-band payload data downlink regardless of the imaging technique employed
- A GPS unit is required to deliver a PPS signal to the SAR transceiver.Moreover, the GPS unit is required for exact orbit determination in order to improve the geolocation precision of images. See the section on Payload Commissioning below.

1.5 Power

- The incredibly high power requirements of SAR payloads are the primary reason why SAR satellites can only image for a few minutes per orbit.
- According to the Japan Aerospace Exploration Agency (JAXA), a SAR payload would require 1100 W of DC power.
- Although a value for the DC power consumption of Cappella Space's payload is unavailable, their solar panels are capable of producing 400 W, allowing them to image for 10 minutes per orbit, according to the their claim.

1.6 Structure

A large, deployable antenna is necessary for synthetic aperture radar operations.

- The ICEYE spacecraft is equipped with a 3.25-meter-long rectangular antenna.
- The EOS SAR payload is equipped with a circular antenna ranging in diameter from 2 to 7 meters.
- Capella Space employs an antenna with a diameter of 3.5 meters.
- It is probable that the payload will have electromagnetic noise requirements. As an OSSAT SAR missions are likely to observe in the X-band, which can cause an increase in noise in the receiver electronics due to the high frequency. The payload would require an electromagnetically "silent" environment to restrict the noise input of the platform.. The payload would likely require an electromagnetically "silent" environment to limit the platform's noise output. Consideration must also be given to the impact the payload's high RF output will have on the remainder of the platform.

Two type of SAR satellites are shown in fugures 3 and 4

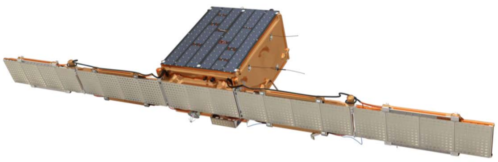


Figure 3: ICEYE-X1 First Generation Spacecraft with 3.25 m, deployable antenna

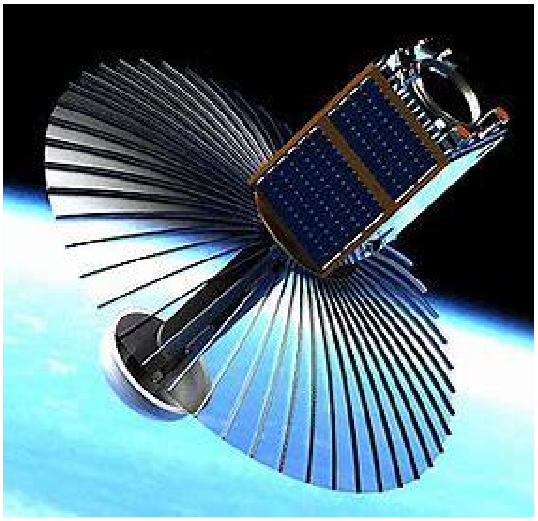


Figure 4: -SSTL's CARBSAR concept with a deployable antenna from Oxford Space Systems (currently in drployment

1.7 frequency bands

SAR systems are based on a pulsed radar placed on a platform that moves forward and have a side-looking imaging geometry. The radar technology sends out powerful electromagnetic pulses and sequentially gathers the echoes of the backscattered signal. For airborne and spaceborne systems, the typical pulse repetition frequency ranges from a few hundred to a few thousand Hertz. In both cases—airborne and spaceborne—the swath width ranges frequently between a few kilometers and 20 kilometers and between 30 and 500 kilometers. Only a portion of the transmitted pulse is backscattered to the receiving antenna, which may be the same as the transmit antenna (for a monostatic radar) or a different one. The transmitted pulse interacts with the Earth's surface (for a bi- or multi-static radar). The physical (i.e., geometry, roughness) and electrical (i.e., permittivity) characteristics of the imaged object determine the amplitude and phase of the backscattered signal. Depending on the frequency band, significant penetration can happen, requiring the modeling of imaged objects and media as volumes (such as plants, ice and snow, dry soil, etc.). Radar systems using longer wavelengths, which often have an amplified volume contribution in the backscattered signal, will experience more electromagnetic pulse penetration through medium. Table 1 lists the common SAR frequency bands and the corresponding wavelength ranges. To have a better understanding of how synthetic aperture radar improved the resolution of the imaging lets have a numerical example;

Frequency Band	Ka	Ku	X	C	S	L	P
Frequency[GHz]	40-25	17.6-12	12-7.5	7.5-3.75	3.75-2	2-1	0.5-0,25
Wavelength[cm]	0.75-1.2	1.7-2.5	2.5-4	4-8	8-15	15-30	60-120

Table 1: commonly used frequency bands for SAR system

The previous generation of the radar-based imaging satellites were denoted as SLAR (side-looking airborne radar) .The simplest SLAR system provides a 2-D reflectivity map of the imaged area, i.e., targets with high back-scattered signal are identified as bright spots in the radar images and flat smooth surfaces as dark areas. The flight direction is denoted as azimuth and the line-of-sight as slant range direction. This led to a moderate "azimuth" resolution which deteriorates as the range increases. For example, an X-band SLAR system with a 3-meter antenna d_a has an azimuth antenna beamwidth of 1:

$$\Theta_a = \frac{\lambda}{d_a} = \frac{0.03m}{3m} = 0.01rad \tag{1}$$

where λ is the wavelength. The azimuth resolution δ_a is given by the smallest separation between two point targets that can be detected by the radar. In the SLAR case this is given by the illumination length of the azimuth antenna beam on the ground. Assuming a range distance, r_0, from the antenna to the targets of 5 km yields azimuth resolution 2:

$$\delta a = (\frac{\lambda}{d_a}).r_0 = \Theta_a.r_0 = 0.01 * 5000 = 50m \tag{2}$$

In the following years, this concept was extended to the principle of the **synthetic aperture** as it is known today.

Consequently the resulting azimuth resolution becomes equal to half the azimuth antenna length ($\delta_a = d_a/2$) and is independent of the range distance. This means that the azimuth resolution in the previous example is equal to 1.5 m. It is more than 30 times better than conventional "SLAR". The above equation suggests that a short antenna yields a fine azimuth resolution. This appears surprising on the first view. However, it becomes immediately clear if one considers that a radar with a shorter antenna "sees" any point on the ground for a longer time (the illumination time can be approximated by Till . (mr_0/vd_a), which is equivalent to a longer virtual antenna length and thus a higher azimuth resolution.

1.8 SAR geometry

Figure 5 illustrates the typical SAR geometry, where the platform moves in the azimuth or along-track direction, whereas the slant range is the direction perpendicular to the radar's flight path. The swath width gives the ground-range extent of the radar scene, while its length depends on the data take duration, i.e., how long the radar is turned on. r_0 stands for the shortest approach distance, Θ_a

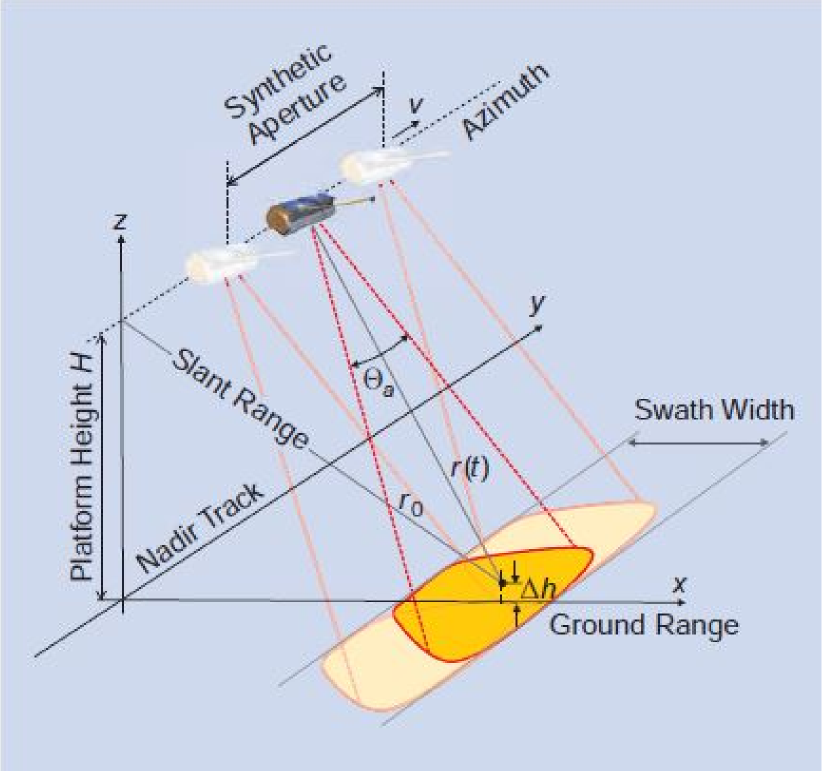


Figure 5: Typical SAR geometry

for the azimuth beam width and v for the sensor velocity. given any time t, the distance between the radar moving at constant velocity v and a point on the ground, described by its coordinates $(x, y, z) = (x_0, 0, \Delta_h)$, is easily obtained applying Pythagoras' theorem:

$$r(t) = \sqrt{r_0^2 + (vt)^2} \approx r_0 + \frac{(vt)^2}{2r_0} \quad for \ \ vt/r_0 \ll 1 \tag{3}$$

where for r_0 we can assume that $t = t_0 = 0$ which is corresponding to time of closest approach,in this time the distance is minimum and it is equal to: $r(t_0) = r_0 = \sqrt{(H - \Delta h)^2 + x_0^2}$

H is the platform height. In general, r_0 is significantly greater than vt during the illumination period. This permits expanding $r(t)$ into a Taylor series and ignoring all but the first two terms, yielding the approximation on the right-hand side of equation 3. In the preceding formula, the time, denoted by the variable t, is related to the movement of the platform and is hence sometimes referred to as *slowtime*. The range variation of a point target over time is closely connected to the azimuth phase by $\varphi(t) = -4\pi r(t)/\lambda$, i.e., the phase variation similarly exhibits a parabolic tendency (the factor 4r is a result of the SAR system's two-way range measurement). Note that the quadratic approximation in 3 is employed for convenience. Without approximation, accurate SAR data processing takes into consideration the exact phase history.

1.9 Imaging modes

By adjusting the antenna radiation pattern, current SAR systems are capable of operating in multiple imaging modes. This is accomplished by separating a planar antenna into sub-apertures and controlling the phase and amplitude of each sub-aperture with transmit/receive modules (TRM). Typically, a few hundred TRMs are deployed, with their settings controlled by software. The most fundamental mode is the Stripmap operation, in which the pattern is fixed to a single swath, thereby imaging a single continuous strip as depicted in Fig.6 (a). The system can be operated in ScanSAR mode if a greater field of view is necessary. As indicated in Figure 6, the antenna elevation pattern is guided repeatedly to varied elevation angles corresponding to several sub-swaths 6 (b). Each subswath is illuminated by numerous pulses, but for a shorter duration than in .Stripmap. Adjustments are made to the timing so that time-varying elevation patterns repeat cyclically the imaging of several continuous subswaths. After suitable processing, a wide-swath SAR image is produced; however, the azimuth resolution is degraded in comparison to the Stripmap mode. If a higher resolution in azimuth is desired, the Spotlight mode is employed. As seen in Figure 6, the antenna pattern is guided in azimuth towards a set point to illuminate a given region 6 (c). The extended illumination time leads in a longer synthetic aperture and, hence, a higher resolution. Along the radar flight path, the Spotlight mode does not scan a continuous swath, but rather separate patches. It turns out that single-channel SAR has inherent limitations such that increasing azimuth resolution reduces swath width, and vice versa.

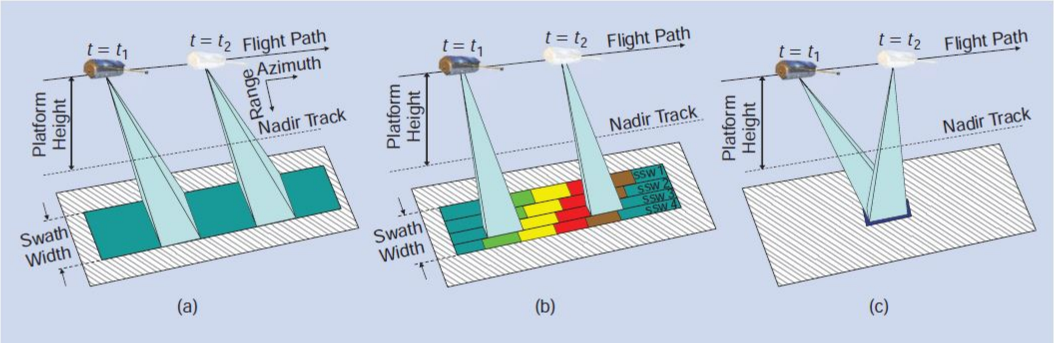


Figure 6: Typical SAR geometry

2 SAR Attitude Dynamic

2.1 Flexible satellite attitude model

The flexible satellite consists of a rigid body and several flexible appendages. Two sets of equations can be used to explain its attitude: the kinematic equation and the dynamical equation. The kinematics of the satellite describes the attitude of the main body, and the parameterizations for the kinematics consist mostly of Euler angles, Cayley-Rodrigues parameters, modified Rodrigues parameters, the quaternion, etc. Noting that the quaternion employs the smallest number of parameters (four parameters) to represent the orientation globally, the quaternion is employed to describe the attitude of the satellite in this study. Following is an illustration of the **kinematics** of the flexible spaceship in terms of quaternions.

$$\dot{q}=\frac{1}{2}E(q)w \tag{4}$$

where $q=(q_0,q_1,q_2,q_3)^T=\left(q_0,q_V^T\right)^T$ is the quaternion, $\omega=(\omega_1,\omega_2,\omega_3)^T$ is the angular velocity in the body fixed frame.
The matrix $E(q)$ is:

$$E(q)=\left[\begin{array}{c}-q_v^T\\ -s\left(q_v\right)+q_0I_3\end{array}\right]$$

with I_3 being a 3×3 identity matrix and $s(\cdot)$ being a skew matrix, i.e., for $\forall x=(x_1,x_2,x_3)^T\in\mathbb{R}^3$

$$s(x)=\left[\begin{array}{ccc}0 & x_3 & -x_2\\ -x_3 & 0 & x_1\\ x_2 & -x_1 & 0\end{array}\right] \tag{5}$$

For a skew matrix, we also have the following properties:

$$x^Ts(x)=0 \text{ and } \|s(x)\|=\|x\|,\forall x\in\mathbb{R}^3, \tag{6}$$

Actually, let Φ denote the principal angle rotated around the Euler axis and $\ell=(\ell_1,\ell_2,\ell_3)^T$ denote the Euler axis coordinates associated with Euler's Theorem with $\ell_1^2+\ell_2^2+\ell_3^2=1$. Then the quaternion can be defined as,

$$\begin{aligned}q_0&=\cos\frac{\Phi}{2}\\ q_i&=\ell_i\sin\frac{\Phi}{2},i=1,2,3\end{aligned} \tag{7}$$

By (7), it can be easily verified that $q_0^2+q_1^2+q_2^2+q_3^2=1$.
Assuming the desired attitude and attitude angular velocity are q_{d} and ω_{d}, then the error attitude is defined as,

$$\begin{cases}q_e=\left[\begin{array}{cc}q_0 & q_v^T\\ -q_v & q_0I_3-[q_v\times]\end{array}\right]q_d\\ \omega_e=\omega-A\left(q_e\right)\omega_d\end{cases} \tag{8}$$

where $q_{\mathrm{e}}=[q_{\mathrm{e}0},q_{\mathrm{e}1},q_{\mathrm{e}2},q_{\mathrm{e}3}]^{\mathrm{T}}$ and $\omega_{\mathrm{e}}=[\omega_{\mathrm{e}0},\omega_{\mathrm{e}1},\omega_{\mathrm{e}2}]^{\mathrm{T}}$ denote the error quaternion and error angular velocity, respectively; $\boldsymbol{q}_{\mathrm{v}}=[q_1,q_2,q_3]^{\mathrm{T}}$ is the vector part of the quaternion $\boldsymbol{q};\mathbf{I}_3\in\mathbb{R}^{3\times3}$ denotes the 3×3 identity matrix; $\mathbf{A}\left(\boldsymbol{q}_{\mathrm{e}}\right)\in\mathbb{R}^{3\times3}$ represents the rotation matrix based on the error quaternion $\boldsymbol{q}_{\mathrm{e}}$, which is expressed as

$$A\left(q_e\right)=\left[\begin{array}{ccc}q_{e1}^2-q_{e2}^2-q_{e3}^2+q_{e0}^2 & 2\left(q_{e1}q_{e2}+q_{e3}q_{e0}\right) & 2\left(q_{e1}q_{e3}-q_{e2}q_{e0}\right)\\ 2\left(q_{e1}q_{e2}-q_{e3}q_{e0}\right) & -q_{e1}^2+q_{e2}^2-q_{e3}^2+q_{e0}^2 & 2\left(q_{e1}q_{e0}+q_{e2}q_{e3}\right)\\ 2\left(q_{e1}q_{e3}+q_{e2}q_{e0}\right) & 2\left(q_{e2}q_{e3}-q_{e1}q_{e0}\right) & -q_{e1}^2-q_{e2}^2+q_{e3}^2+q_{e0}^2\end{array}\right] \tag{9}$$

Substituting Eq. (8) into Eq. (4) the tracking kinematic model is figured out as,

$$\dot{q}_e=\frac{1}{2}\left[\begin{array}{c}-q_{ev}^T\\ q_{e0}I_3+[q_{ev}\times]\end{array}\right]\omega_e \tag{10}$$

The **dynamical equation** describes both the overall rigid body motion and essential motions of flexible appendages.Using Eulor's equation of motion, it can usually be modeled as,

$$\begin{aligned} J\dot{\omega} + \delta^T\ddot{\eta} &= s(\omega)\left(J\omega + \delta^T\dot{\eta}\right) + u + d(t), \\ \vec{\eta} + C\dot{\eta} + K\eta &= -\delta\dot{\omega}, \end{aligned} \tag{11}$$

where $J = J^T$ is a 3×3 symmetric inertia matrix, $u = (u_1, u_2, u_3)^T$ is the control torque, δ is the coupling matrix between the flexible and rigid structures and it is exactly where the flexibility of SAR is taken into account, η is the modal coordinate vector, $C = \operatorname{diag}\{2\xi_i\omega_{ni}, i = 1, \cdots, N\}$ is a damping (diagonal) matrix, $K = \operatorname{diag}\{\omega_{ni}^2, i = 1, \cdots, N\}$ is a stiffness matrix, N is the number of flexible modes, and $\omega_{ni}, i = 1, \cdots, N$ and $\xi_i, i = 1, \cdots, N$ are the natural frequencies and the corresponding damping, respectively. In general, it is not possible to model the attitude dynamic system completely. There may be some errors when measuring or estimating the inertia matrix J, coupling matrix δ, damping matrix C, and stiffness matrix K. Under this circumstance, $d(t)$ can be regarded as disturbances including possible parameter variations and external disturbance torques. Here $d(t)$ is assumed to satisfy the following assumption.

Assumption - There exists a positive constant d such that $\|d(t)\| \leq d$.
It seems more natural to assume that $|d_i(t)| \leq d_i$ with positive constants $d_i, i = 1, 2, 3$. Note that $\|d(t)\| = \sqrt{|d_1(t)|^2 + |d_2(t)|^2 + |d_3(t)|^2}$. By letting $d = \sqrt{d_1^2 + d_2^2 + d_3^2}$, it is easy to obtain $\|d(t)\| \leq d$ which is just the assumption introduced above.

If we let $\psi = \dot{\eta} + \delta\omega$ and $J_m = J - \delta^T\delta$, then combining (1) and (5), the mathematical model for the flexible satellite attitude control system can be rewritten as

$$\begin{aligned} \dot{q} &= \frac{1}{2}E(q)\omega \\ \dot{\eta} &= \psi - \delta\omega \\ \dot{\psi} &= -(C\psi + K\eta - C\delta\omega) \\ J_m\dot{\omega} &= s(\omega)\left(J_m\omega + \delta^T\psi + h_\omega\right) + \delta^T(C\psi + K\eta - C\delta\omega) + N_c + d(t). \end{aligned} \tag{12}$$

Now based on the flexible satellite attitude model (12),the control objective of the paper is to design a controller for tracking the desired attitude and angular velocity in the minimum time and damping out the vibrations caused by flexibility.

Note that if q_v is required to be stabilized to zero, it follows from $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$ that $q_0 = \pm 1$. It should be pointed out that mathematically speaking, the equilibria $q = (-1, 0, 0, 0)$ and $q = (1, 0, 0, 0)$ are different points. However, according to (7), by a simple calculation, it is clear that $q_0 = 1$ corresponds to $\Phi = 4m\pi, m \in \mathbb{Z}$ and $q_0 = -1$ corresponds to $\Phi = 4m\pi + 2\pi, m \in \mathbb{Z}$. Thus, the two equilibria in reality represent the same equilibrium position. Usually, the desired quaternion is designed to be $(1, 0, 0, 0)$ while $(-1, 0, 0, 0)$ is regarded as an unstable equilibrium. Then it is said that since in the physical space both of the equilibria are the same, global asymptotical stability can be obtained. However, under those controllers, all the states have to be driven to $(1, 0, 0, 0)$ even if they are located very close to $(-1, 0, 0, 0)$. Apparently, such a control method is not energy-efficient, which is actually the "unwinding phenomenon" . The "unwinding phenomenon" can be avoided if a so-called "set control scheme" is adopted, which will make both of the equilibria stable so that the quaternion variables can be stabilized to the equilibrium that is closer to them

To simplify the notation, for $x \in \mathbb{R}^n$, we denote

$$\operatorname{sign}(x)|x|^\alpha = \left(\operatorname{sign}(x_1)|x_1|^\alpha, \cdots, \operatorname{sign}(x_n)|x_n|^\alpha\right)^T$$

with $\operatorname{sign}(\cdot)$ being a sign function and $\alpha > 0$.

Remark - Practically speaking, the thruster, piezoelectric actuators, moment gyros, momentum wheels, and others are the major actuators of attitude control systems. The reaction wheel and thruster, however, are the most often used actuators for a 3-axis stabilized satellite. From a theoretical perspective, momentum wheels and reaction wheels' actuator dynamics are typically represented as a first- or second-order system, and their time constants are relatively short. In this scenario, the actuators will typically make sure that input commands are interpreted satisfactorily. Since the actuator's dynamics are faster than those of the plant, they are typically not taken into account when designing controllers. As a result, the actuator dynamics is frequently overlooked in the majority of studies on satellite attitude control where the controller is designed for control torque.

2.2 Controller Design

System 12 may clearly be categorized as a cascaded system based on its structure. As a result, the finite-time control technique will be used in this part to propose a cascaded design approach for the attitude control problem. First, a two-part description of the fundamental idea behind our control strategy is given.
STEP 1:

Create a stabilizing controller for the kinematic subsystem and modal dynamics of the following system:

$$\begin{aligned} \dot{q} &= \frac{1}{2}E(q)\omega_d \\ \dot{\eta} &= \psi - \delta\omega_d \\ \dot{\psi} &= -\left(C\psi + K\eta - C\delta\omega_d\right) \end{aligned} \tag{13}$$

The virtual controller is the angular velocity indicated by ω_d.
To stabilize the system 12, we first create the virtual controller $\omega_d(q, \eta, \psi)$. However, in practice, the modal variables (η, ψ) are rarely directly measured. To that purpose, a modal observer is built in order to acquire the estimated modal variables and the virtual controller $\omega_d(q, \eta, \psi)$ can then be utilized to stabilize the subsystem 13.
For simplicity in this study we neglect this requirement and assume that we have access to modal variable.

STEP 2:

Develop a non-smooth controller u such that the angular velocity ω quickly approaches the virtual angular velocity $\omega_d(q, \eta, \psi)$ obtained in the previous section. Under the proposed controller, the attitude and the modal variables will be stabilized first, followed by the angular velocity. In the absence of disturbances, the state variables will asymptotically stabilize at the origin. In the presence of perturbations, the state variables will stabilize to a narrow region close to the origin.

A. Design Virtual Angular Velocity ω_d
Choosing virtual control action as follow:

$$\omega_d(q, \eta, \psi) = -k_1\left[q_v + \delta^T(C\psi - 2K\eta)\right] \tag{14}$$

with $k_1 > 0$, system 13 can be almost globally asymptotically stabilized. Global Asymptotic stability of system can be proved if we consider a positive definite Lyapunov function as 15.

$$\begin{aligned} V_1(q, \eta, \psi) = (q_0 - 1)^2 + q_v^T q_v + \frac{1}{2}\psi^T\psi + \eta^T K\eta \\ + \frac{1}{2}(\psi + C\eta)^T(\psi + C\eta) \end{aligned} \tag{15}$$

The derivative of $V_1(q, \eta, \psi)$ along system 13 is negative semi definite. Therefore the virtual controller ω_d almost stabilizes the subsystem 13 globally asymptotically, excluding unstable equilibrium point (-1, 0 , 0 , 0,0....0,0....0). Last two sets of zeros have N values(the number of flexible mode) Virtual controller 14 will drive the system towards the equilibrium (-1 0 0 0) .
Another concern in practice is the unwinding phenomenon caused by duplicate attitude representations.

A quaternion-based attitude control system with two equilibria can be created by parameterizing the satellite attitude by two antipodal unit quaternions (both Q and -Q reflect the same physical orientation in the quaternion context). However, only one of the two equilibria is considered stable in the majority of the existing literature, while the other is viewed as unstable, leading to the unwinding phenomenon (spacecraft tumbles through an unnecessary large rotation). The quaternion should be stabilized to the equilibrium point that is most similar to the original quaternion in order to theoretically avoid the unwinding problem.

For instance, it will be preferable to stabilize the quaternion to (-1,0,0,0) if q(0) is close to (-1,0,0,0), whereas (1,0,0,0) will be preferable if q(0) is close to (1,0,0,0). In this case, a suitable controller can be selected as 16:

$$\omega_d(q, \eta, \psi) = -k_1 \left[\operatorname{sign}\left(q_0(0)\right) q_v + \delta^T(C\psi - 2K\eta)\right] \tag{16}$$

B. Design Controller u

we want to have a nonsmooth controller u to guarantee that the virtual controller $\omega_d(q, \eta, \psi)$ will be tracked by the angular velocity *omega*.

By defining controller u as 17:

$$\begin{aligned} u = & -s(\omega)\left(J_m\omega + \delta^T\psi\right) - \delta^T(C\psi + K\eta - C\delta\omega) \\ & + J_m\omega_d - \beta_1 s\left(\alpha, \omega - \omega_d\right), \end{aligned} \tag{17}$$

where $0 < \alpha < 1, \beta_1 > \max[2, d]$

and $s(\alpha, x)$ is:

$$s(\alpha, x) = \begin{cases} x, & \text{for } |x| > \epsilon \\ \frac{\operatorname{sign}(x)|x|^\alpha}{\epsilon^{\alpha-1}}, & \text{for } |x| \leq \epsilon \end{cases} \tag{18}$$

with $x \in \mathbb{R}, \alpha > 0$ and $\epsilon \geq 1$. In addition,

for $x \in \mathbb{R}^n$, we denote $s(\alpha, x) = \left(s\left(\alpha, x_1\right), \cdots, s\left(\alpha, x_n\right)\right)^T$. It can be easily verified that $s(\alpha, x)$ is continuous.

then we have the following statements :

1. In the absence of disturbances, system 4 and 11 can be almost globally asymptotically stabilized.
2. In the presence of disturbances, the states of system 4 and 11 will converge into some bounded regionas.

When $\alpha = 1$, $s(\alpha, x) = x$. Then, controller 17 reduces to:

$$\begin{aligned} u = & -s(\omega)\left(J_m\omega + \delta^T\psi\right) - \delta^T(C\psi + K\eta - C\delta\omega) \\ & + J_m\omega_d - \beta_1(\omega - \omega_d), \end{aligned} \tag{19}$$

which can be considered as a conventional **backstepping** controller. It seems that through adjusting control gains k_1 and β_1 of controller 19 to be large enough, both convergence regions (i.e., for the nonsmooth control17 case and for the conventional backstepping control case19) can be rendered to be as small as desired, which may mean that the nonsmooth control method does not have any prominent advantages over the conventional backstepping control method on disturbance rejection. However, high-gain feedback control systems often tend to become unstable in practical operation. From the considerations of stability as well as control saturation constraint, it is advisable not to select k_1 and β_1 to be high control gains. In this case, for the nonsmooth control method given here, an additional parameter (i.e., the fractional power α) can be adjusted to enhance the disturbance rejection.

Having dynamic of the system and controller we are able to implement control scheme for SAR satellite .As shown in figure7 includes SAR dynamics+kinematics 12 which is preceded by controller 16 and 17. There is also feedback error term block 8 that provides error terms for quaternions and angular velocities.
In addition to above-mentioned blocks there are also two Deep Reinforcement Learning Agents which aim to optimize the coefficients used in controller.

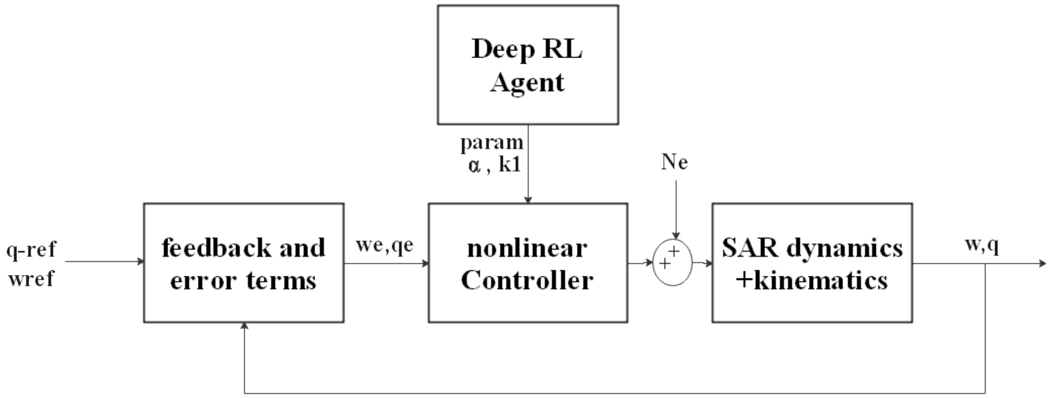


Figure 7: Block diagram of SAR satellite attitude control

3 Parameter Optimization using Deep Reinforcement Learning

Deep reinforcement learning, a real-time decision-making algorithm by iterative training to maximize the reward in a dynamic environment, has been extensively used in video games, robots, drones, and other systems. Deep reinforcement learning has steadily been employed in the field of satellite attitude control to introduce the autonomous decision-making capability as the processing capacity of satellites has increased, leading to the development of increasingly sophisticated software for satellite systems. To improve the adaptability and autonomy of satellite control systems, Xu et al. (2019) developed a model-based deep reinforcement learning with heuristic search approach. In order to simplify the labor of parameter adjustment and establish the ability to adapt to the space environment, Zhang et al. (2018) developed an attitude control approach based on the deep deterministic policy gradient (DDPG) to learn attitude control strategy in orbit. To obtain a highly precise attitude control in a dynamically uncertain environment, Elkins et al. (2020) developed an adaptive spacecraft attitude controller utilizing the twin delayed deep deterministic policy gradient (TD3). For planetary powered descent and landing at six degrees of freedom, Gaudet et al. (2020) proposed a deep reinforcement learning-based method. Due to the deep learning's lack of interpretability, the current attitude control methods, which go from the sensor signal to the control torque, are end-to-end methods and cannot guarantee the stability and reliability of the satellite attitude control system. Traditional satellite attitude controllers, like the PD controller (Wen and Kreutz-Delgado, 1991), H1 controller (Liu et al., 2018), sliding mode controller (Amrr and Nabi, 2020), backstepping controller (Huo et al., 2018), and others, have theoretical completeness and reliability, which helps the satellite attitude control system be reliable and stable. In order to improve the autonomy of attitude tracking control and maintain the dependability of the satellite attitude control, it is important to mix deep reinforcement learning with classical satellite attitude controller

Considering the continuous state and action spaces of the satellite attitude dynamics, we use deep deterministic policy gradient (DDPG) algorithm to train a policy in order to optimize parameter gains for the backstepping controller. DDPG is a Q based method that combines elements of actor- critic framework by maintaining specific actor and critic networks. To that end, we employ the Reinforcement learning toolbox in Matlab. RL toolbox helps us in the process of building and training a policy using RL algorithms. it also allows you to represent policies and value functions using deep neural networks or look-up tables and to train them through interactions with environments modeled in Matlab

or Simulink.
So first, we set up the environment in which we want to train our agent by specifying observations (states), actions, and reward variables from the Simulink model of the satellite dynamics and the controller.

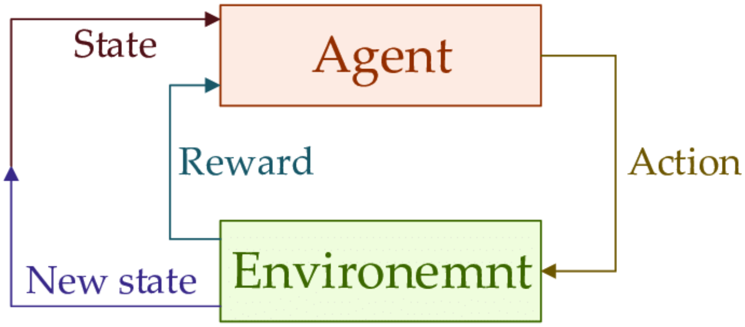
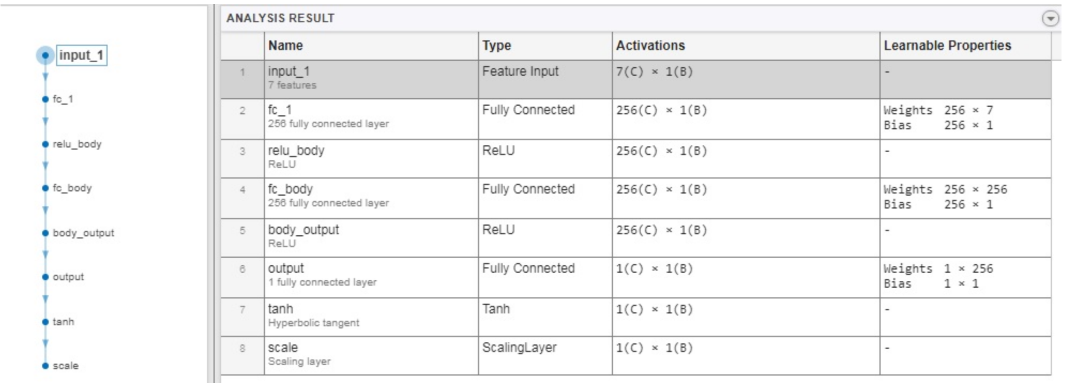


Figure 8: Block diagram of RL agent and environment

The way our trained agents interacts with the environment in Simulink is through the RL agent block which inputs states and the reward function, and outputs the action to the environment.
In our case the states for the two trained agents are the attitude quaternion and the angular velocity vectors $s_t = \{q_t, \omega_t\}$. The actions are the parameters gain for the controller $K_1, alpha$ each optimized by one of the agents.
We choose the reward function to minimize the quaternion error w.r.t the reference values in the minimum time.

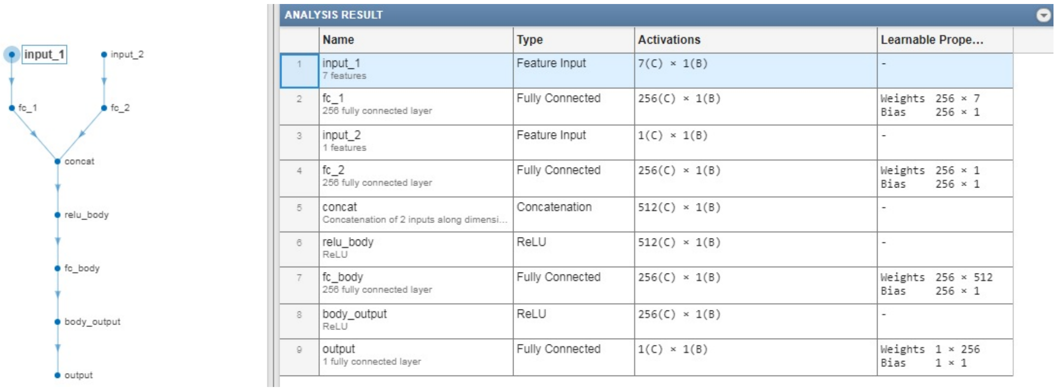
Next, we import our environment to the Reinforcement learning designer app. Using this app we can create the agent, which in this case is a deep neural network.
The architecture of the actor-critic network is proposed by the RL designer app itself after analysing the data received from the environment.



ANALYSIS RESULT

	Name	Type	Activations	Learnable Properties
1	input_1 7 features	Feature Input	7(C) × 1(B)	-
2	fc_1 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 7 Bias 256 × 1
3	relu_body ReLU	ReLU	256(C) × 1(B)	-
4	fc_body 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 256 Bias 256 × 1
5	body_output ReLU	ReLU	256(C) × 1(B)	-
6	output 1 fully connected layer	Fully Connected	1(C) × 1(B)	Weights 1 × 256 Bias 1 × 1
7	tanh Hyperbolic tangent	Tanh	1(C) × 1(B)	-
8	scale Scaling layer	ScalingLayer	1(C) × 1(B)	-

Figure 9: Actor network structure of the proposed DDPG

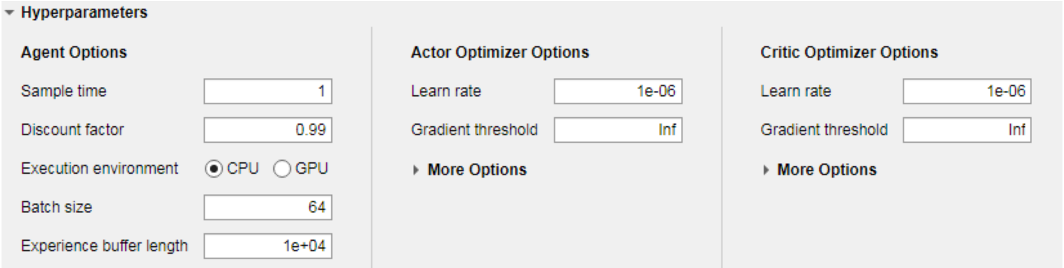


ANALYSIS RESULT

	Name	Type	Activations	Learnable Prope...
1	input_1 7 features	Feature Input	7(C) × 1(B)	-
2	fc_1 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 7 Bias 256 × 1
3	input_2 1 features	Feature Input	1(C) × 1(B)	-
4	fc_2 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 1 Bias 256 × 1
5	concat Concatenation of 2 inputs along dimensi...	Concatenation	512(C) × 1(B)	-
6	relu_body ReLU	ReLU	512(C) × 1(B)	-
7	fc_body 256 fully connected layer	Fully Connected	256(C) × 1(B)	Weights 256 × 512 Bias 256 × 1
8	body_output ReLU	ReLU	256(C) × 1(B)	-
9	output 1 fully connected layer	Fully Connected	1(C) × 1(B)	Weights 1 × 256 Bias 1 × 1

Figure 10: Critic network structure of the proposed DDPG

As shown in the above figures no convolutional layers are necessary. Since the input states are continous, feature input is used as the input layer. For the output, scaling layer is used and the Tanh function in the output layer of the actor network ensures the boundedness of the actions to satisfy the constraints of the parameters gain of the controller.
Now we set the hyper-parameters and the training options for our agent.



Agent Options

Sample time: 1

Discount factor: 0.99

Execution environment: ● CPU ○ GPU

Batch size: 64

Experience buffer length: 1e+04

Actor Optimizer Options

Learn rate: 1e-06

Gradient threshold: Inf

Critic Optimizer Options

Learn rate: 1e-06

Gradient threshold: Inf

Figure 11: The hyper-parameters chosen for the training of the actor-critic networks

The optimization of hyperparameters is one of deep reinforcement learning's most challenging and time-consuming processes. These factors, such as the learning rate or discount factor, can significantly impact how well your agent performs. After some trials and iterations we chose the values above for the training of our agent. DDPG algorithm in this app uses Adam optimizer by default.
Next, we set the training options and let the algorithm train our agents.

4 Simulation Results

4.1 Model overview

Fig 12 illustrates the overall scheme implemented in simulink. All the blocks are discussed in previous

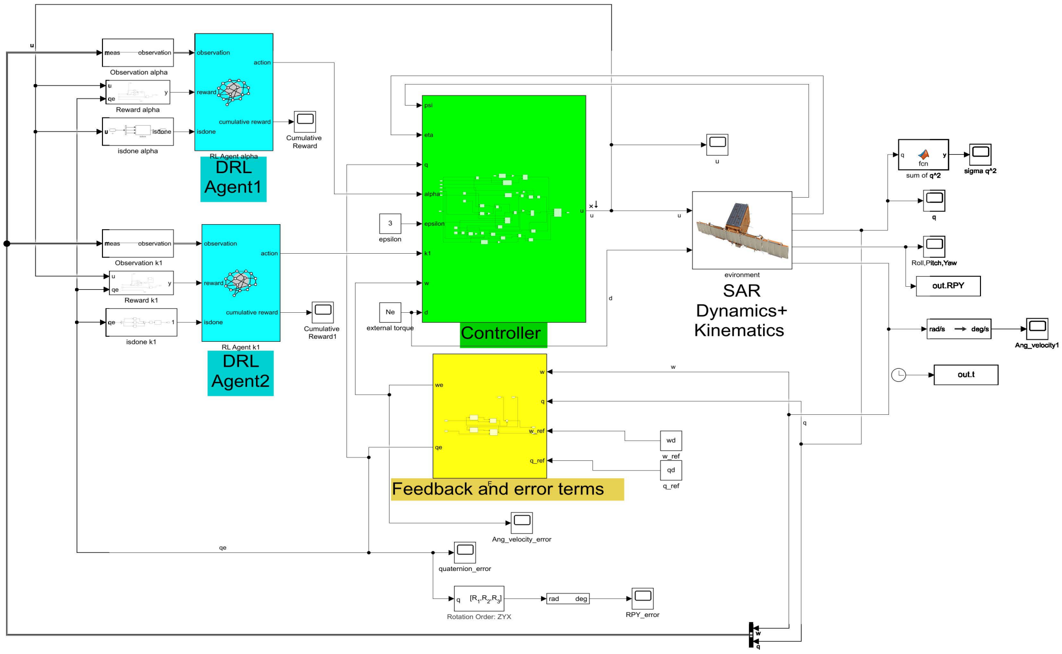


Figure 12: Simulink general layout

sections. There are SAR dynamics and kinematics block which includes all the sub-blocks related to a flexible SAR satellite dynamics(12). The controller in green(16 and 17).The feedback & error terms in yellow (8). There are also 2 Deep Reinforcement Learning agents in cyan to optimize some parameters of controller. The model parameters of a flexible spacecraft are chosen as,

$$J = \begin{bmatrix} 10 & 1 & 1 \\ 1 & 8 & 0 \\ 1 & 0 & 6 \end{bmatrix} \mathrm{Kg}^2 \cdot \mathrm{m}^2 \qquad \delta = \begin{bmatrix} 1.3523 & 1.2784 & 2.1553 \\ -1.1519 & 1.0176 & -1.2724 \\ 2.2167 & 1.5891 & -0.8324 \\ 1.23637 & -1.6537 & 1.2251 \end{bmatrix} \mathrm{Kg}^{1/2}\ \mathrm{m/s}^2$$

$$\omega_{\mathrm{n}1} = 1.5973, \omega_{\mathrm{n}2} = 2.2761, \omega_{\mathrm{n}3} = 1.9538, \omega_{\mathrm{n}4} = 2.4893\ \xi_1 = 0.056, \xi_2 = 0.086, \xi_3 = 0.08, \xi_4 = 0.025$$

This implies $C = \mathrm{diag}\{0.1789, 0.3915, 0.3126, 0.1245\}$ and $K = \mathrm{diag}\{2.5514, 5.1806, 3.8173, 6.1966\}$.

4.2 simulation without agent

In this section first we intend to show that the proposed controller stabilizes the system and enables it to track any desired reference for quaternions from any given initial angles and angular velocity.

The initial values for control parameters and attitude dynamics are shown in table 2

parameters	k_1	α	ϵ	$\theta(0),\phi(0),\psi(0)$	$w1(0)$,$w2(0)$,$w3(0)$
values	10	0.2	3	60°,30°, 15°	$\pi/12$,$\pi/24$,0 deg/s

Table 2: simulation parameters without agent

As shown in figure 13 satellite dynamics is stabilized and tracks given reference and initial values in table 2.In transient time the behaviour of system is not smooth.In the next chapters a comparison is made considering settling time and peaks with the system equipped by DRL-agents .This is also the case for the angular velocities in Figure14. From practical point of view the control action needs to be taken into account. Since in this study we neglected the actuator model(ex:reaction wheels) the magnitude and change rate of control torque must be within a reasonable range. Otherwise the demanded torque can not be transferred to the system. Fig 15 shows the performance of the control action. It is obvious that the frequency of control action is relatively high(approximately 400Hz)and it might not be suitable to be followed by an actuator,even with high bandwidth.

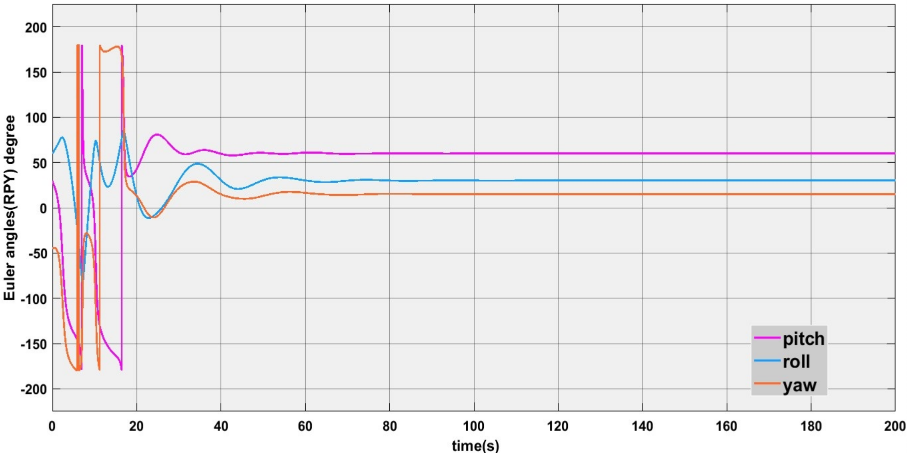


Figure 13: Euler angles Roll,Pitch,Yaw with references: 60°,30° 15°

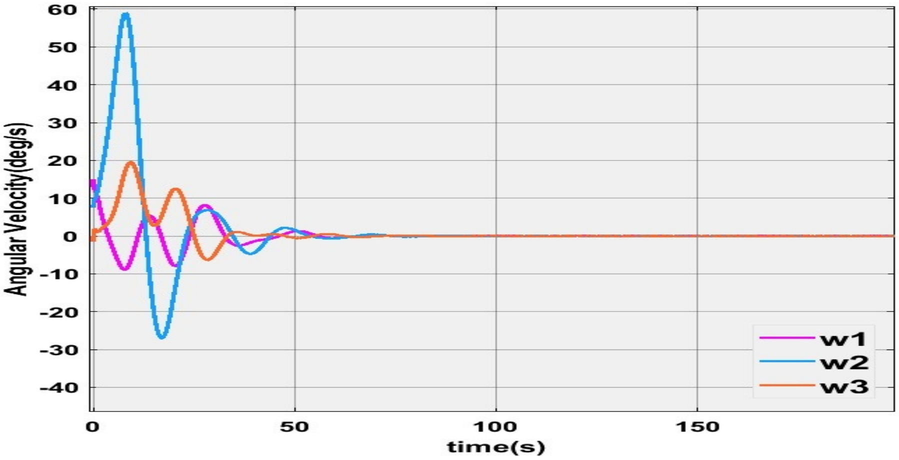


Figure 14: Angular velocities initializing from : $\pi/12$,$\pi/24$,0 deg/s

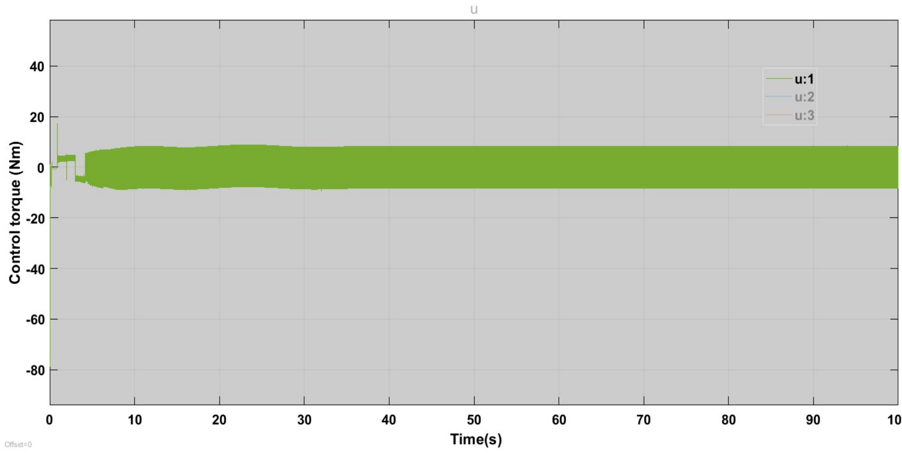


Figure 15: control torque (Nm)

4.3 Simulation with agents

In this section first we describe how the two agents are trained in order to provide the controller with optimized parameters k_1 and α and then simulation results are shown.

4.3.1 DRL agents training

After setting up the environment, the structure of the agent and choosing the proper hyperparameters and training options, now we let the DDPG algorithm train our agent for 2000 episodes with each episode having a maximum length of 50 steps.
To make the training more efficient, we put a constraint on the control action through the RL agent block in Simulink. The result of the training shows the agent converged to its optimum value after almost 1000 episodes. sate and action space and reward function are as follow:

$$\begin{aligned} S_t &= \{w_t, q_t\} \\ A_t &= \{\alpha\} \\ R(s_t) &= -\left|q_e\right| - 0.4 \end{aligned} \tag{20}$$

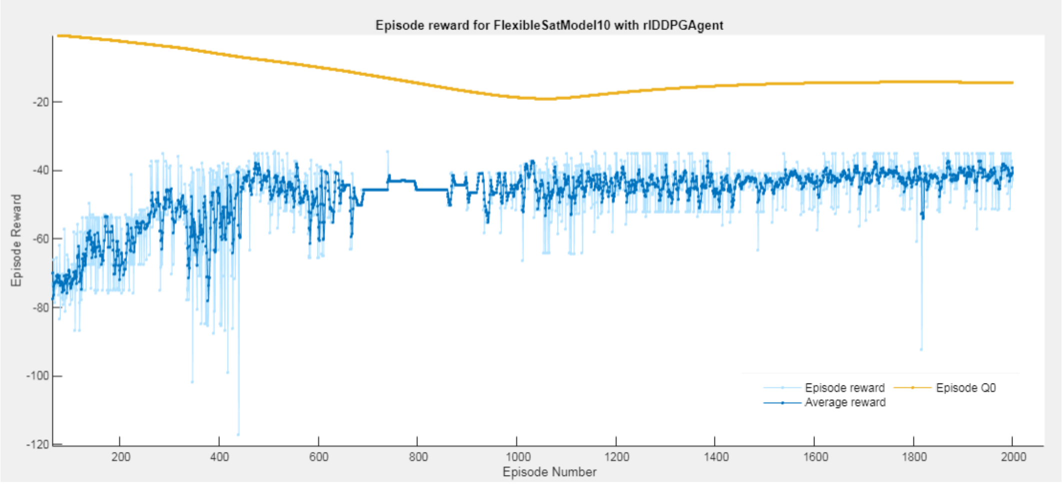


Figure 16: Reward curve and average reward after 2000 episodes of training of the agent using DDPG algorithm

The faded blue curve in the figure above represents the episode reward which is the reward of every simulation or episode, whereas the other blue curve is a moving average and it's useful because in deep

RL, rewards can get very noisy due to aggressive exploration strategy. Also it helps us to follow where the overall trend of the rewards is going. The orange line is the episode Q0 and it defines the initial value or Q function that the critic is estimating based on our initial condition. Similarly second agent to optimize $k_1$is trained by following characteristics:

$$\begin{aligned} S &= \{w_t, q_t\} \\ A &= \{k_1\} \\ R(s_t) &= -(1.4 * 1e - 5 - q_e)^2 - 0.4 \end{aligned} \tag{21}$$

the constant terms in 20 and 21 penalize the converging time.This improves the settling time and it is shown numerically in the next section.

4.3.2 Results

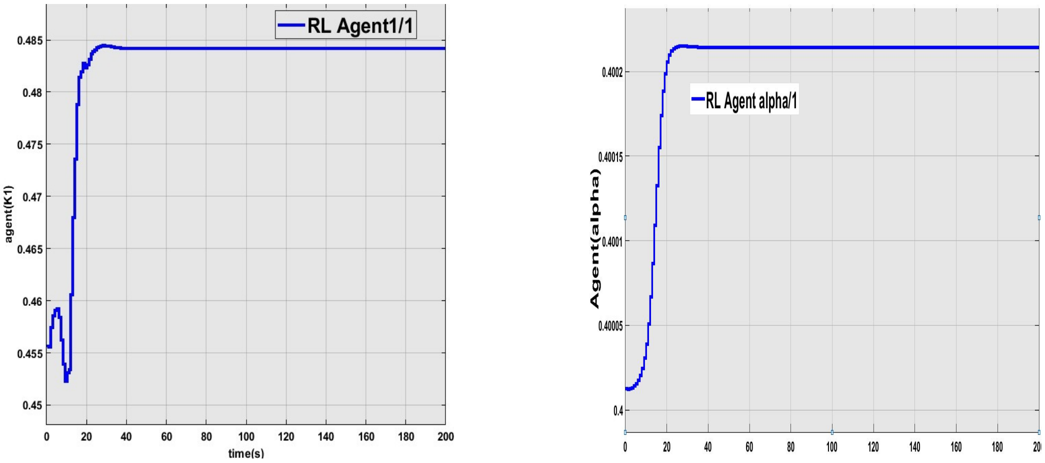
Having two agents trained, we are able to deploy them instead of constant coefficients for k_1 and α.Therefore the new setting is as shown in table 3

parameters	k_1	α	ϵ	$\theta(0), \phi(0), \psi(0)$	$w1(0)$,$w2(0)$,$w3(0)$
values	agent2	agent1	3	60°,30°, 15°	$\pi/12$,$\pi/24$,0 deg/s

Table 3: simulation setting with agent

Figure 17 depicts the evolution of k_1 and α during simulation time.Both converges to a constant value which let the system behave faster and smoother(this comparison is made in section 4.4).

Both parameters tend to a constant.



(a) agent k_1

(b) agent α

Figure 17: k_1 and α evolution

The following figures show Euler angles18, angular velocity 19 and control torque evolution 20.
It is clear that on the one hand the transient time behaviour and the settling time improve when agents take the control of controller parameters. On the other hand the control action 20 is not harsh anymore and it shows significant improvement with respect to the system with constant coefficients.

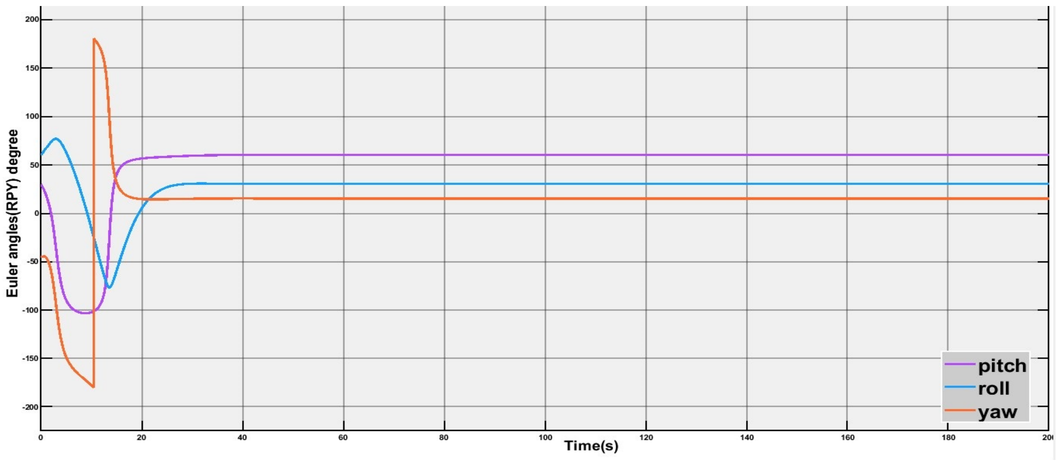


Figure 18: Euler angles Roll,Pitch,Yaw with references: 60°,30° 15°

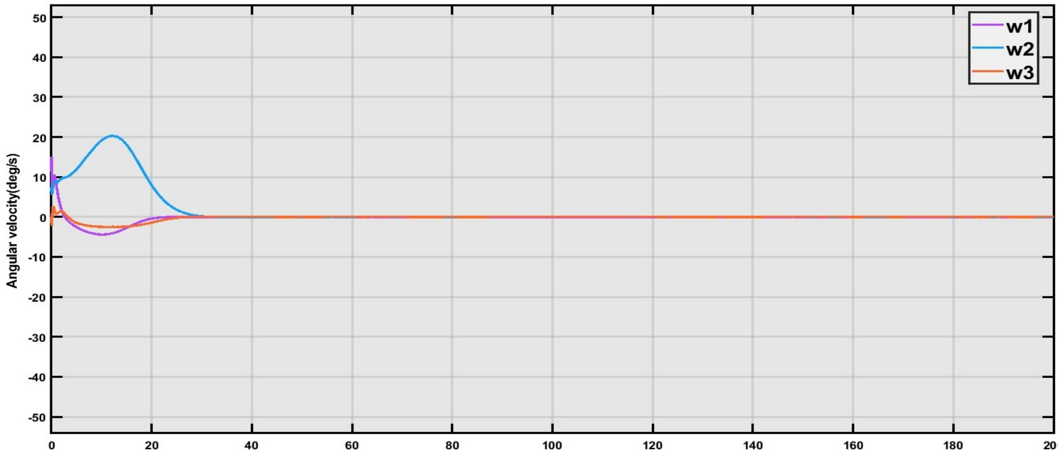


Figure 19: Angular velocities initializing from : $\pi/12$,$\pi/24$,0 deg/s

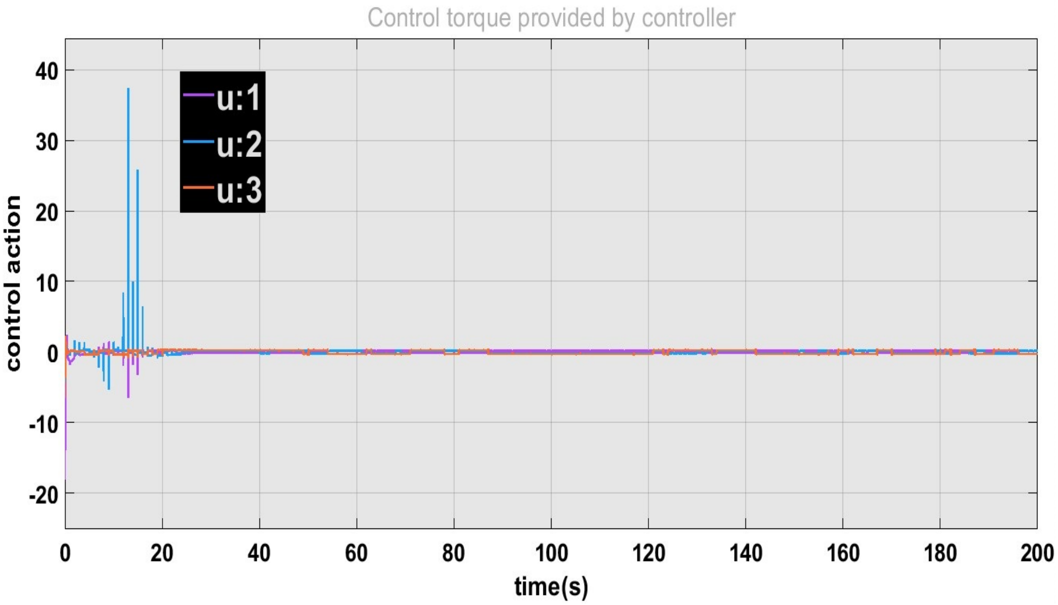


Figure 20: control torque Nm

4.4 comparison

In the table 4 settling time,transient time, peak values are compared for two controlled system:the first one without agent and the second one with trained agent included. For simplicity the results shown in the table 4 is related to the roll angle(θ).The other angles have similar behaviour.

Roll angle(θ)	Transient Time	Settling Time	Overshoot	peak
without DRL agent	28.98 s	45.32 s	199.9	179.9
with DRL agent	21.09 s	28.13 s	0.0022	102.69

Table 4: comparison between two cases:a(no agent) b(with agent)

This table shows the superiority of deploying DRL agents in the control system and in particular in SAR attitude control system.

5 Conclusion

In this study we focus on a specific type of imaging satellites ”SAR” which illuminate earth surface by sending and receiving waves. Flexibility must be addressed in modeling phase due to the antenna and flexible appendages.Using an improved version of backstepping controller lets the satellite converge to one of two equilibria depending on initial values[1].This controller not only stabilizes the system but also enables reference tracking problem. Deep reinforcement learning provides the controller with optimized parameters and improves the temporal characteristics of the system.Finally we have a controlled system which is able to track desired Euler angles(or equivalently ’quaternions’)references starting from any initial angles and angular velocity.

6 Further studies

One of the most challenging area that has not been so far addressed properly (at least to the best of our knowledge) is a suitable attitude profile generator for Synthetic Aperture Radars.Although there are some concepts in literature such as minimization of ”Centroid Doppler variation ”[4] in order to have the optimum attitude during imaging phase,it has not been deployed using powerful techniques such as reinforcement learning. Of course for optical imaging satellite this profile generator is already implemented using DRL agent[3].

References

[1] Shihong Ding and Wei Xing Zheng. Nonsmooth attitude stabilization of a flexible spacecraft. *IEEE Transactions on Aerospace and Electronic Systems*, 50(2):1163–1181, 2014.

[2] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013.

[3] Zhong Shi, Fanyu Zhao, Xin Wang, and Zhonghe Jin. Satellite attitude tracking control of moving targets combining deep reinforcement learning and predefined-time stability considering energy optimization. *Advances in Space Research*, 69(5):2182–2196, 2022.

[4] Shuo Zhao, Yunkai Deng, and Robert Wang. Attitude-steering strategy for squint spaceborne synthetic aperture radar. *IEEE Geoscience and Remote Sensing Letters*, 13(8):1163–1167, 2016.