AIAA SciTech Forum
8-12 January 2024, Orlando, FL
AIAA SCITECH 2024 Forum

10.2514/6.2024-1812

Autonomous Orbital Slot Maintenance with Impulsive Maneuvers and Reinforcement Learning

Yutong Cheng*, Max Z. Li†, and Oliver Jia-Richards‡
University of Michigan, Ann Arbor, Michigan, USA

This paper investigates impulsive maneuver-based control methods for satellite orbital slot maintenance in the presence of dynamic variations. The objective is to develop an efficient controller that ensures the satellite remains within a designated region around a reference trajectory, the slot, while minimizing propellant usage, facilitating autonomous onboard control. The controller is designed using Clohessy-Wiltshire terminal guidance for a linearized satellite dynamics model relative to the slot. The Q-learning algorithm is utilized in the decision-making process for selecting the optimal target point, after which the appropriate control input is determined through Clohessy-Wiltshire terminal guidance. The inclusion of atmospheric drag and gravity model differences adds complexity to determining the optimal target point for the maneuvers. The learning process produces a controller that favors the target point closest to the position where the satellite leaves the slot, reducing propellant costs by approximately 15.9% compared to when Q-learning was not used. These results underscore the effectiveness of Q-learning for maneuver optimization, consistently achieving greater rewards compared to returning to the slot center, implying reduced propellant consumption and enhanced position maintenance. Furthermore, the computational time remains within practical limits, making this approach a viable option for autonomous onboard control.

I. Introduction

To accommodate increasing amounts of space traffic with different missions and maneuvering capabilities, space traffic management and coordination (STM/C) principles consisting of "rules-in-orbit" must be developed, standardized, and adopted by a wide range of space-faring stakeholders to ensure on-orbit safety and efficient utilization of orbital capacities. Analogous to the role that the US Federal Aviation Administration (FAA) has over rules-of-the-sky and commercial air traffic, it will be critical for future traffic in an increasingly congested space environment—inclusive of commercial and scientific activities—to be managed and coordinated. However, the implementation of STM/C architectures for low-Earth orbit congestion control will increase the operational costs of satellites both in terms of monitoring by the satellite operators as well as potential propellant usage onboard the satellite in order to ensure conformance with a given STM/C architecture.

A recent research thrust is the design of low-Earth orbit constellations that can provide near-global coverage and would be applicable to a wide variety of missions such as terrestrial communications or remote sensing [1, 2]. In such constellations, satellites are assumed to follow prescribed trajectories that are designed to avoid inter-satellite collisions while maximizing the capacity of different orbits. For practical operation of individual satellites within a constellation, as well as to ensure adequate spacing between nearby satellites, slot-based architectures have been developed in parallel [3, 4]. In a slot-based architecture, each satellite is allocated, and must stay within, a fixed-size region of space around the reference trajectory, analogous to the notion of en-route slots and insertions into the en-route stream in air traffic control [5, 6]. The trajectory of the slot itself is propagated over time in order to conform to the specified constellation design. As long as each satellite stays within its slot, large low-Earth orbit constellations with a wide variety of customers and satellite mission types can be designed and implemented with relative ease.

Although slot-based architectures simplify the design of constellations, they shift the onus of responsibility for STM/C to the individual satellite operators. Each satellite must be continuously monitored to assess if it is leaving its slot and, if it is, carry an onboard propulsion system to correct its trajectory. Satellites may tend to drift out of their slots for a number of reasons. The most likely cause will be differences between the natural dynamics of the satellite

*Master's Student, Department of Aerospace Engineering, yutongch@umich.edu, AIAA Student Member.
†Assistant Professor, Department of Aerospace Engineering, Department of Civil and Environmental Engineering, Department of Industrial and Operations Engineering, maxzli@umich.edu, AIAA Member.
‡Assistant Professor, Department of Aerospace Engineering, oliverjr@umich.edu, AIAA Member.

and those used in simulation to propagate the position of the slot. The dynamics used to propagate the position of the slot will use a reduced-order model of the Earth gravity field and will likely not account for other accelerations (e.g., atmospheric drag, gravitational influence of the Sun and the Moon) that may act on individual satellites. In particular, the inclusion of atmospheric drag in the constellation design is often omitted because its effects vary significantly based on the unique characteristics of each satellite. Since constellations may be developed without specific satellite designs in mind or for multiple different users, effects that depend on the individual satellites are neglected.

It is therefore implicitly assumed during the design of these constellations that the individual satellites will have a way to correct the trajectory deviations of the satellite within its slot due to these dynamics differences through the use of an onboard propulsion system. These corrections will require the use of onboard propellant which will also reduce the satellite mass available for payload. In addition, the propellant cost of the corrections will depend on the magnitude of the difference between the satellite and slot dynamics. Some of these differences, such as atmospheric drag, may not be possible to actively control. However, the dynamics differences due to the model of Earth's gravity field present a potential cost trade-off—the use of a lower-fidelity model of the Earth's gravity field reduces the computational cost required to propagate the position of the slot, but increases the propellant cost required to maintain the satellite trajectory within the slot. Such trade-offs should be quantitatively assessed and taken into consideration when designing future STM/C system architectures.

The main purpose of this paper is to develop a reinforcement learning-based controller for computing impulsive maneuvers that maintain the satellite within the slot boundaries in the presence of atmospheric drag and Earth gravity model differences while minimizing the propellant cost. The method should be computationally simple such that the satellite can autonomously determine onboard its required control action without needing Earth-based resources. The position of the slot will be simulated with a low-fidelity model, while the satellite's position will be simulated using a high-fidelity model of the Earth's gravitational field as well as potential atmospheric drag. Over time, atmospheric drag and the differences in the Earth's gravitational field models between the satellite and the slot will cause the satellite to drift out of the slot. The controller's goal is to compensate for this drift assuming that the satellite has an onboard propulsion system capable of performing impulsive maneuvers in any desired direction.

II. Literature Review

The aerospace industry is witnessing a paradigm shift towards the development of large-scale satellite networks, a trend fueled by the global demand for high-speed internet and advanced communication services. Companies like SpaceX [7], OneWeb [8], and Telesat [9] are actively involved in developing mega-constellations, which will significantly increase the density of active satellites across various orbital altitudes. Such activities necessitate innovative approaches in satellite constellation architecture design. A prevalent design methodology is the Walker Constellation [10], which evenly spaces satellites in multiple orbital planes, ensuring uniform global coverage. In contrast, the street-of-coverage constellation [11], predominantly utilized in navigation systems such as GPS, arranges satellites in such a way that several satellites are always visible from any point on Earth, thereby guaranteeing continuous global service. Another concept is the Flower Constellation [12], where satellites' ground tracks form patterns akin to flower petals. This design offers enhanced flexibility in coverage and frequency of observation, as the overlapping orbits can be optimized for specific observational and communication needs. Such a configuration can be particularly advantageous in scenarios where targeted, repeated coverage over certain areas is required, making it a viable option for Earth observation and certain types of communication networks. Each constellation configuration, with its distinct orbital arrangement and coverage patterns, will influence orbital slot management. The selection of a particular design, encompassing factors such as satellite count and spatial distribution, directly impacts satellite density in specific orbital zones. Higher-density constellations necessitate more complex slot management strategies to avoid collisions and minimize signal interference. Furthermore, the inter-satellite spacing within the same orbital plane becomes a critical consideration, influencing how orbital slots are allocated and maintained.

Effective station-keeping plays a crucial role in maintaining the integrity of satellite constellation architectures, particularly when implementing active control techniques in low-Earth orbits. This task involves regular orbital adjustments to counteract perturbative forces like atmospheric drag and Earth's gravitational perturbations, which are more pronounced at lower altitudes. These forces tend to gradually decrease a satellite's altitude while increasing its velocity, necessitating continual monitoring and adjustment. A key method employed in station-keeping is orbit phasing. This technique specifically adjusts the satellite's position within its orbital plane, thereby maintaining appropriate spacing in constellation systems. Orbit phasing is particularly useful in dense constellations where precise inter-satellite distances are critical to avoid potential collisions. The effectiveness of orbit phasing hinges on robust orbit determination

systems, which provide accurate tracking and predictive modeling of the satellite's trajectory.

To counter deviations induced by gravitational perturbations and atmospheric drag, active impulsive maneuvers are executed. In conventional station-keeping, these maneuvers can be calculated based on analytical models and executed using finite thrust [13, 14]. However, the advent of low-thrust propulsion [15, 16] and the geometric constraints on satellites' relative state [17] have introduced additional complexity into the system dynamics. Consequently, innovative optimization-based methods have emerged to better tackle the nuances of station-keeping in such scenarios [18, 19]. These station-keeping maneuvers must be managed to conserve fuel reserves, a critical factor in extending the satellite's operational lifespan. The challenge lies in balancing regular orbital adjustments with finite fuel resources. This underscores the need for efficient, predictive station-keeping strategies within the dynamic and often unpredictable environment of low-Earth orbits.

Traditional control methods such as proportional-integral-derivative (PID) and linear quadratic regulator (LQR) techniques provide a simple approach to satellite trajectory control, particularly for continuous thrust systems. However, when decision-making in uncertain and complex environments is required, these conventional methods may not be adequate due to their limitations in handling complex system dynamics. In such scenarios, more advanced control strategies like reinforcement learning, adaptive control, and model predictive control may be needed. Adaptive control starts with a basic system model and dynamically adjusts its parameters in response to environmental changes [20]. Model predictive control, on the other hand, operates on a defined model of the system, optimizing control actions based on predictive analytics without an inherent learning capability [21]. In contrast, reinforcement learning, specifically model-free reinforcement learning, learns and evolves through direct interaction with the environment, making it suitable for systems where precise modeling is challenging [22].

Given the complexity of the system model in this research, due to dynamic differences in satellite behaviors, providing an analytical model is impractical. Consequently, reinforcement learning emerges as the preferred approach for achieving autonomous satellite control in this context. Its applicability in satellite control is increasingly recognized, with uses ranging from trajectory determination for autonomous spacecraft rendezvous [23] to coordination tasks in satellite constellations [24]. Moreover, reinforcement learning's potential extends to optimizing orbital maneuvers [25]. Its capability to learn and refine the most fuel-efficient maneuvers over time is particularly valuable, enhancing operational efficiency and sustainability. Reinforcement learning also demonstrates its versatility in navigating optimal orbital paths under dynamically changing conditions, further underscoring its suitability for this complex and evolving domain of satellite control.

While the potential of reinforcement learning in broader aspects of satellite control has been recognized, its application in the specific domain of orbital slot maintenance, particularly within the complex and dynamic context of low-Earth orbits, remains insufficiently explored. A critical area lacking in-depth research is the use of reinforcement learning for optimizing fuel efficiency in station-keeping maneuvers, a factor crucial to the sustainability and longevity of satellites. Moreover, the robustness and computational efficiency of these reinforcement learning-based control systems amidst dynamic environmental uncertainties are areas that have not been adequately addressed. This study aims to fill these gaps by developing a reinforcement learning-based controller. This controller is designed to ensure satellite positioning within its designated slot space, while minimizing both propellant use and computational cost. Through this approach, the study endeavors to enhance the accuracy, efficiency, and adaptability of satellite operations in the demanding conditions of low-Earth orbital environments.

III. Problem Statement

In the context of this research, the satellite orbits near-circularly around Earth, with the center of its designated slot serving as a reference point on this orbital path. While the slot could theoretically take any shape or size, for the purposes of this study, it is defined as a spherical region with a radius of 500 meters. The satellite's trajectory is subject to various perturbing influences, notably atmospheric drag and certain gravitational forces not fully accounted for in the slot's trajectory model. These additional gravitational forces, which are in excess of those considered for the slot trajectory propagation, can lead to deviations from the intended orbit. This discrepancy arises because the model used to predict the slot's path is of lower fidelity compared to the natural dynamics affecting the satellite. Consequently, to address these deviations and ensure the satellite remains within its assigned slot boundaries, targeted control efforts are essential. These efforts aim to realign the satellite with the center of its slot, thereby maintaining the overall efficiency of the space traffic management and coordination system.

The proposed controller is developed and evaluated through three analyses. Initially, a theoretical analysis is carried out using purely linearized dynamics, excluding any gravity model variations. The second analysis incorporates

nonlinear dynamics, considering potential gravity model differences, but focuses on a single target point. This phase includes two separate simulation scenarios: one accounting for gravity model differences and another discounting such differences. In this study, the term *gravity model degree* refers to the degree and order of the spherical harmonics gravity model used to represent Earth's gravitational field. In scenarios featuring gravity mismatches, the model's degrees are set differently: 20 for the reference model and 10 for the slot model. Conversely, in cases without such mismatches, both the satellite and slot dynamics are modeled using a uniform degree, set to 5. Throughout this second analysis, atmospheric drag is consistently factored in, assuming a constant atmospheric density and a typical value for atmospheric drag acceleration. The final analysis, incorporating reinforcement learning, builds upon the second and extends its focus to include multiple target points. The application of reinforcement learning is a key differentiator from other analyses, enabling autonomous control in this phase.

The controller will be designed based on Clohessy-Wiltshire (CW) terminal guidance for a linearized model of the satellite dynamics [26] relative to the center of a designated region around a reference trajectory, known as a *slot*. When the satellite detects that it is on the edge of the slot, a two-burn impulsive maneuver can be quickly calculated. This maneuver is designed to bring the satellite back to a target point chosen by the controller, using the maximum possible maneuver time length that keeps the satellite within the slot boundaries during maneuvering. The selected target point should minimize propellant use and maximize the time between maneuvers. At the end of maneuvering, a second maneuver can cancel out the relative velocity between the satellite and the slot center. This simple terminal guidance method is particularly effective over a few kilometers. Another advantage of this method lies in its relative simplicity, which facilitates more tractable calculations for real-time operations in space. Such attributes make it suitable for the precise and controlled maneuvering required in this study. The controller, while based on CW terminal guidance, is enhanced by a learning-based approach using techniques from reinforcement learning, specifically Q-learning. It is designed with the capability to select from a range of potential target points, with the objective of learning to choose the target point that results in minimized propellant cost. It refines the standard CW guidance by learning to identify target points and maneuver strategies that optimize propellant usage beyond the baseline efficiency offered by CW terminal guidance alone. The goal is to develop a policy that not only maintains the satellite's position within the slot but also achieves greater propellant cost reduction compared to traditional CW terminal guidance methods.

The CW guidance strategy is built upon the dynamics described by the CW frame. This frame is a linearized approximation of the two-body orbital dynamics about a circular reference trajectory. It includes a rotation of the frame itself to constantly align with the radial, along-track, and cross-track directions. This linearization and rotation facilitate a more simplified and intuitive representation of satellite motion during proximity operations such as rendezvous and docking maneuvers. The CW frame's mathematical framework simplifies the complex dynamics of orbital mechanics into a more manageable form, significantly enhancing the efficiency and precision of predicting and controlling relative positions and velocities.

Within the scope of three distinct analyses, it is crucial to distinguish the application of dynamical models. The first theoretical analysis employs the Hill-Clohessy-Wiltshire (HCW) equations for dynamics propagation. In contrast, the subsequent analyses utilize the full nonlinear dynamics of the slot and satellite. However, in these analyses, the representation of position errors is still framed within the CW frame. Let x, y, z represent the relative position of the satellite relative to the slot center. The origin of the CW frame is established at the slot center, with the x axis extending radially outward from this point, the y axis aligned with the along-track direction, and the z axis in the direction of the reference orbit's angular momentum vector (See Fig. 1). The center of Earth lies along the negative x axis and remains fixed within the CW frame. Derived from the CW frame, the HCW equations encapsulate the dynamics governing the relative motion of the satellite concerning the slot center within a circular orbit around Earth [27]

$$\begin{aligned} \ddot{x} &= 3n^2x + 2n\dot{y} + \Gamma_x, \\ \ddot{y} &= -2n\dot{x} + \Gamma_y, \\ \ddot{z} &= -n^2z + \Gamma_z, \end{aligned} \tag{1}$$

where $n = \sqrt{\mu/r^3}$ is the mean motion of the slot's reference orbit, μ stands for the standard gravitational parameter, r represents the spherical slot radius, and Γ is the thrust acceleration vector.

Throughout the simulation, the position and velocity vectors of the satellite are stored in a state vector $\delta\boldsymbol{s}(t)$, which details the position and velocity of the satellite relative to the slot center.

$$\delta\boldsymbol{s}^T(t) = \begin{bmatrix} x & y & z & \dot{x} & \dot{y} & \dot{z} \end{bmatrix}. \tag{2}$$

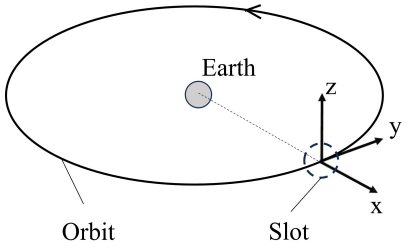


Fig. 1 The Clohessy-Wiltshire (CW) coordinate system.

When determining the velocity change required to bring the satellite to a target point selected by the controller, a 6×6 state transition matrix $[\Phi](t)$ can be used to map the six-dimensional state vector at one time to a six-dimensional state vector at another time.

$$\delta \boldsymbol{s}(t) = [\Phi](t)\delta \boldsymbol{s}(0). \tag{3}$$

This state transition matrix can be derived from the HCW equations,

$$[\Phi](t) = \begin{bmatrix} 4 - 3\cos nt & 0 & 0 & \frac{1}{n}\sin nt & \frac{2}{n}(1 - \cos nt) & 0 \\ 6(\sin nt - nt) & 1 & 0 & -\frac{2}{n}(1 - \cos nt) & \frac{1}{n}(4\sin nt - 3nt) & 0 \\ 0 & 0 & \cos nt & 0 & 0 & \frac{1}{n}\sin nt \\ 3n\sin nt & 0 & 0 & \cos nt & 2\sin nt & 0 \\ -6n(1 - \cos nt) & 0 & 0 & -2\sin nt & 4\cos nt - 3 & 0 \\ 0 & 0 & -n\sin nt & 0 & 0 & \cos nt \end{bmatrix}. \tag{4}$$

Partitioning the 6×6 state transition matrix into four 3×3 matrices is a common practice in orbital mechanics.

$$[\Phi](t) = \begin{bmatrix} [M](t) & [N](t) \\ [S](t) & [T](t) \end{bmatrix}. \tag{5}$$

The primary reason behind this is the distinct physical interpretations and applications of the different matrix sections. $[M](t)$ represents how the initial position affects the final position. $[N](t)$ shows how the initial velocity affects the final position. $[S](t)$ indicates how the initial position impacts the final velocity. Lastly, $[T](t)$ represents how the initial velocity affects the final velocity. For many practical problems in orbital mechanics, not all parts of the state transition matrix are necessary. By partitioning the matrix, it is possible to focus specifically on segments relevant to the data of interest. In this study, the focus will be on utilizing only $[M](t)$ and $[N](t)$ partitions for calculating the necessary velocity changes. Additionally, from a computational standpoint, handling smaller matrices can enhance efficiency, making this approach not only targeted but also more resource-effective.

The return maneuver initiates when the satellite reaches the slot's edge, serving as the initial step in realigning the satellite with its designated target point. The necessary velocity vector of the satellite at the initial time, $\boldsymbol{v}_{\text{req}}$, is calculated using Eq. (6),

$$\boldsymbol{v}_{\text{req}} = [N]^{-1}(t_f)(\boldsymbol{r}_{\text{rel}}(t_f) - [M](t_f)\boldsymbol{r}_{\text{rel}}(0)), \tag{6}$$

where $t = 0$ denotes the maneuver's start and $t = t_f$ denotes the maneuver's end [27]. $\boldsymbol{r}_{\text{rel}}$ is the relative position vector of the satellite with respect to the slot center. $[N]$ stands for the upper right 3×3 partition of the state transition matrix, while $[M]$ represents the upper left 3×3 partition of the state transition matrix. Subsequently, the required velocity change for the return maneuver, Δv_1, is determined using Eq. (7),

$$\Delta v_1 = \|\boldsymbol{v}_{\text{req}} - \boldsymbol{v}_{\text{rel}}(0)\|, \tag{7}$$

where $\boldsymbol{v}_{\text{rel}}$ represents the relative velocity vector of the satellite with respect to the slot center. After the return maneuver, as the satellite approaches its target point, a stopping maneuver becomes necessary to match the satellite's velocity with

that of the target point in the Clohessy-Wiltshire frame. The required velocity change for the stopping maneuver, Δv_2, is computed with Eq. (8).

$$\Delta v_2 = \| - \boldsymbol{v}_{\text{rel}}(t_f) \|. \tag{8}$$

The resulting Δv_1 and Δv_2 values are used to calculate the reward R for each return and stopping maneuver pair, as described by Eq. (9).

$$R = -\frac{(\Delta v_1 + \Delta v_2)}{t_f}. \tag{9}$$

The cumulative reward will be calculated as the total accumulation of R, serving as a qualitative metric to evaluate the performance of the controller developed using Q-learning techniques.

IV. Theoretical Analysis

The CW frame and the HCW equations are employed to describe the relative motion of the satellite in relation to the slot center within a circular Earth orbit. The primary objective is to provide a comprehensive theoretical assessment of expected outcomes under idealized conditions. The simulation takes place within an idealized model, omitting considerations of gravity model differences and conducting the entire simulation within the linearized CW frame. Atmospheric drag is excluded from consideration during the return maneuver, ensuring that the satellite returns precisely to its designated target point.

The primary objective is to thoroughly characterize the satellite's behavior as it operates within its designated orbital slots. This involves quantifying the expected total velocity change required, Δv, for the maneuver, the time intervals between these maneuvers, and the reward associated with different choices of target point locations, thereby better understanding the dynamics governing deviations from the slot center. Also, to comprehensively explore the influence of various factors, the location of the target point is systematically varied. The analysis aims to uncover the relationship between target point selection and the extent and rate of satellite deviations. In addition, by quantifying the expected Δv requirements, rate of Δv usage, and time intervals between maneuvers, the assessment intends to offer practical guidance for minimizing control effort while ensuring stability and precise positioning within the orbital slots.

A. Methodology in Theoretical Analysis

In the idealized scenario, the analysis focuses exclusively on target points with a zero radial position to ensure the satellite remains within its designated orbit. This approach is based on the principles of the linearized HCW equations in Eq. (1) where the acceleration in the radial direction ($\ddot{x}$) is influenced by the radial position (x). A nonzero radial position in the target point would result in a trajectory deviation, leading the satellite away from the slot center. Such deviations necessitate additional maneuvers and increased control efforts for realignment with the slot center. By focusing on target points without radial offsets, trajectory maintenance within the intended orbit becomes more stable and efficient.

Therefore, in the theoretical analysis within the idealized scenario, the choice is made to omit non-zero radial positions to maintain model simplicity and clarity. Nevertheless, it remains crucial to recognize the limitations inherent in this simplification. Notably, this approach may not fully encapsulate scenarios where non-zero radial positions exert a substantial influence, such as missions involving highly eccentric orbits or significant radial maneuvers. Consequently, while this approach offers valuable insights within the idealized framework, it is imperative to acknowledge its specific applicability constraints and potential discrepancies when compared to real-world scenarios.

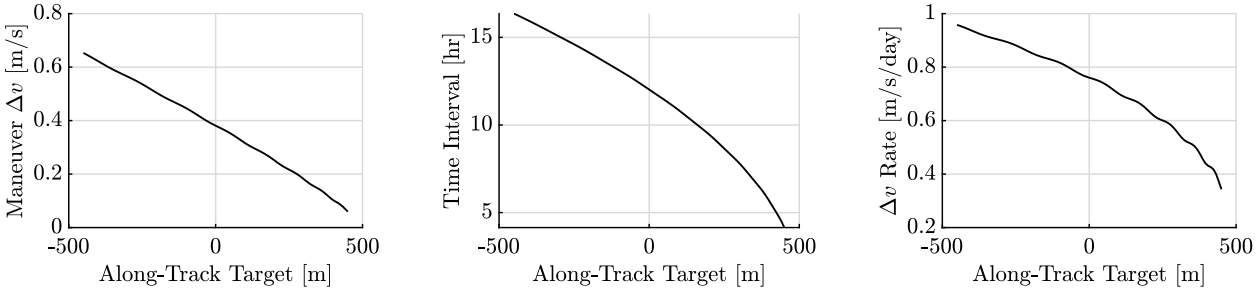
The sequence of the idealized simulation is structured as follows: The satellite initiates its trajectory from a specific target point location. Subsequently, satellite motion is simulated while accounting for a constant along-track drag acceleration. Upon reaching the edge of the slot, a return maneuver is executed in order to return the satellite back to the target point. During these return maneuvers, atmospheric drag is temporarily disregarded, ensuring that the satellite returns exactly to the desired target point. The required velocity change, Δv, expended during the maneuver, and the time interval between maneuvers, T, are recorded. The ratio of the total Δv to the total simulation time provides valuable insight into the expected cumulative reward for non-idealized simulations, shedding light on the effectiveness of this theoretical analysis.

In the idealized simulation, target points are distributed along the along-track axis, resulting in three sets of recorded results that encompass several key aspects of satellite behavior within the orbital slot. The first set of results focuses on

the Δv required for the two maneuvers: the return maneuver and the subsequent stopping maneuver, both essential for bringing the satellite to each along-track target point. These two values can be computed by using Eqs. 7 and 8. Then, the total Δv needed to return the satellite to each along-track target point is calculated as the sum of Δv_1 and Δv_2. The second set of results records the time interval between maneuvers for various along-track target points. This time interval encompasses both the duration of the satellite's return maneuver to the target point and the subsequent time taken to reach the edge of the slot from that target point. Determining this interval is pivotal in slot maintenance strategies, as it influences the frequency of control actions and propulsion resource utilization. A shorter time interval may necessitate more frequent maneuvers, while a longer interval allows for less frequent adjustments. The final set of results quantifies the rate of Δv for different along-track target points, measuring the rate at which velocity adjustments are applied to the satellite to bring it to a specific along-track target point.

B. Theoretical Analysis Results

The outcomes of the theoretical analysis are presented in Fig. 2. Fig. 2a displays the total Δv necessary to guide the satellite back to each specific along-track target point. Fig. 2b visualizes the recorded time intervals between maneuvers, representing the duration from the satellite reaching the edge of the slot to its subsequent arrival at the slot's edge. Lastly, Fig. 2c portrays the rate of Δv for various along-track target points.



(a) Total Δv **(b) Time intervals between maneuvers** **(c) Rate of Δv usage**

Fig. 2 Expected Δv requirements, time intervals between maneuvers, and rate of Δv usage for different along-track target points under idealized conditions.

According to the theoretical results depicted in Fig. 2, all three plots exhibit a decreasing trend. As the along-track position of the target points increases, a consistent decrease is observed in Δv, T (time intervals between maneuvers), and the rate of Δv usage. The trend in Δv is approximately linear, with subtle irregular fluctuations. As for T, it decreases at an accelerating rate with an increase in the along-track position of the target point. This results in rapidly decreasing time intervals between maneuvers, particularly for target points located closer to the front of the along-track axis.

These trends in Δv and T contribute to the observed pattern in the Δv rate, calculated as Δv over T. The Δv rate exhibits an exponential decrease with slight oscillations. This exponential trend shows that the efficiency benefit from each maneuver increases as the target point gets closer to the slot edge. The oscillations in the Δv rate may be attributed to the inherent dynamics of the satellite's movement. These findings imply that under idealized scenarios, target points near the point of the satellite's exit from the slot require less Δv, and the rate of Δv usage relative to time becomes increasingly efficient. The observed trend in the Δv rate offers insights into the anticipated rewards for different along-track target positions. A target point located nearer to the slot edge, where the satellite tends to exit the slot, is expected to yield a lower rate of Δv usage relative to target points located further from the slot edge.

V. Single-Target Point Analysis

The primary aim of the single-target point analysis is to offer a comprehensive assessment of satellite behavior and trajectory when subjected to non-idealized conditions. This analysis encompasses several specific objectives: Firstly, it strives to bridge the gap between theoretical idealizations and real-world scenarios. Unlike the theoretical analysis that simplifies conditions, the non-idealized simulation takes into account persistent atmospheric drag, a factor encountered in actual space missions. Secondly, this assessment aims to quantitatively assess the influence of non-ideal conditions on satellite behavior and trajectory parameters, including Δv usage, rate of Δv consumption, and the number

of maneuvers executed. This encompasses an exploration of how atmospheric drag perturbs the satellite's trajectory and how variations in gravity models impact its behavior. Additionally, the analysis seeks to draw comparisons between the outcomes of non-idealized simulations and those obtained under idealized scenarios.

This comparative analysis sheds light on the deviations and complexities introduced by real-world conditions, offering insights into the practical challenges of slot-based orbital maintenance. Lastly, another objective of this analysis is to anticipate the expected outcomes within the context of autonomous orbital slot maintenance. By examining the results derived from the single-target point analysis, including parameters like Δv utilization, the rate of Δv consumption, and the frequency of maneuver execution, valuable insights into satellite dynamics are gained, particularly when it consistently returns to a fixed target point. This examination provides a deeper understanding of the underlying principles governing the use of reinforcement learning in autonomous slot maintenance strategies. Furthermore, it sets the stage for comparisons with the reinforcement learning results that will be presented in a subsequent section. This comparative approach allows for a crosscheck, enabling the validation and calibration of reinforcement learning models against the real-world behavior of the satellite under non-idealized conditions.

A. Methodology for Single-Target Point Analysis

In the context of the non-idealized single-target point analysis, a key departure from the theoretical analysis is the utilization of the Earth-Centered, Earth-Fixed (ECEF) frame as the primary reference frame. In the ECEF frame, the origin is anchored at the center of the Earth, with its x axis oriented toward the Prime Meridian at the equator, the y axis directed to the equator at 90 degrees East longitude, and the z axis aligned with the North Pole. The choice of the ECEF frame in this analysis stems from its suitability for modeling real-world scenarios, where factors such as atmospheric drag and gravity model differences necessitate a more accurate representation of satellite behavior. This frame offers a stable and globally referenced coordinate system for addressing these complex conditions. Another significant departure from the theoretical analysis is that this analysis does not solely rely on the linearized dynamics typical of the CW frame. Instead, it incorporates the full nonlinear dynamics of the satellite and the slot. However, to leverage the benefits of linearized equations of motion and facilitate calculations, occasional transformations between the ECEF frame and the CW frame may be required. These transformations enable the integration of linearized dynamics into the analysis while preserving the accuracy provided by the ECEF frame in capturing the effects of non-idealized conditions.

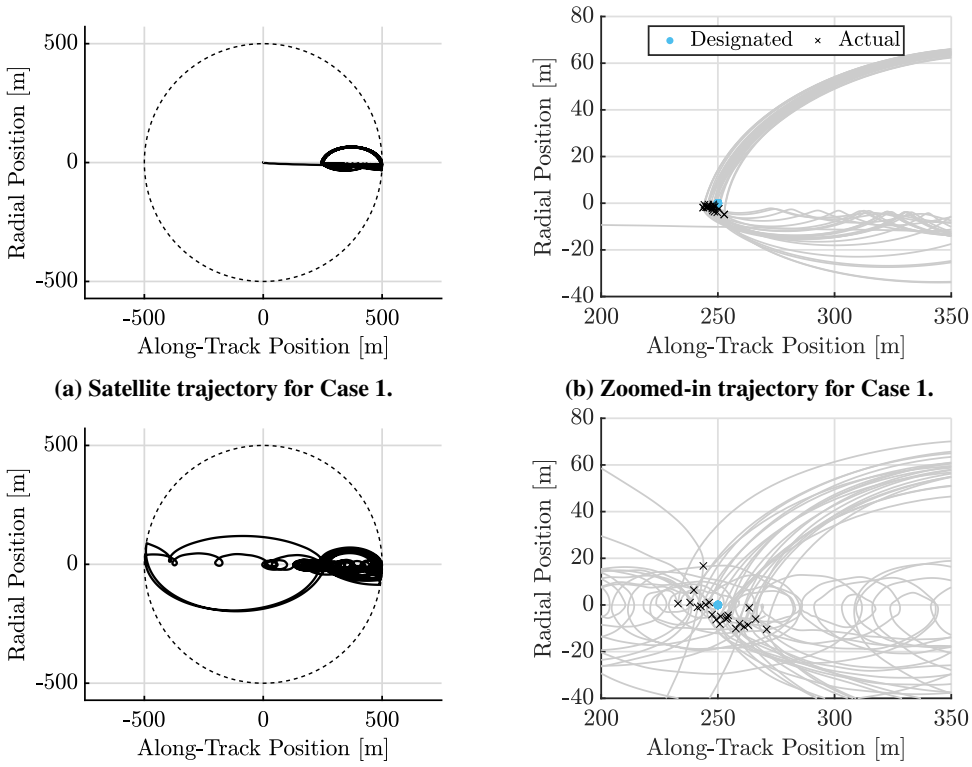
In this real-world context, it is highly probable that the satellite will not precisely return to its target point due to the complexities introduced by atmospheric drag and gravity model differences. This divergence from the idealized case in the theoretical analysis necessitates an adjustment in the control effort. In the idealized case, simulations were conducted for a single return maneuver followed by a stopping maneuver, as the results remained consistent regardless of the simulation's duration. In contrast, in the non-idealized scenario, where satellite trajectories may deviate from the target due to nonzero radial positions at the end of maneuvering, the simulation time is extended to a fixed duration of 5 days for all target points. Anticipating a higher number of maneuvers in the non-idealized scenario, this approach accounts for the possibility of varying Δv requirements each time the satellite returns to the same target point. It ensures that the analysis comprehensively captures the effects of real-world conditions.

The analysis comprises two parts: In the first part, the satellite's behavior is simulated with a single target point at $(0, 250, 0)$ m in the CW frame over a 5-day period in both cases. The satellite's trajectory is recorded to validate predictions related to non-idealized conditions, such as whether the satellite accurately returns to the target point and whether the Δv requirements exhibit variability across maneuvers. The second part parallels the idealized simulation but incorporates the impact of non-idealized conditions. Multiple target points are considered, and data such as average maneuver Δv, maneuvers performed, and cumulative reward is recorded. The average maneuver Δv represents the mean value of Δv_1 and Δv_2 for all return and stopping maneuver pairs. The number of maneuvers counts all return and stopping maneuver pairs. Meanwhile, the cumulative reward is the sum of all rewards R computed for each return and stopping maneuver pair, as described by Eq. (9). Exploration is conducted on how varying gravity models affect satellite behavior under the influence of atmospheric drag.

B. Single-Target Point Analysis Results

The results for the first part of the single-target point analysis, where the satellite consistently returns to the fixed target point $(0, 250, 0)$ m in the CW frame when it reaches the slot edge during a 5-day period, are presented in Fig. 3. Fig. 3a and 3b depict the satellite's 2D trajectory in the radial and along-track directions and a close-up view of the trajectory around the target point. These figures illustrate the satellite's behavior under the influence of atmospheric

drag alone. Additionally, Fig. 3c and 3d show the satellite's 2D trajectory and a zoomed-in view around the target point, respectively, taking into account both non-idealized effects. The results for the second part of the single-target point analysis are presented in Fig. 4. Fig. 4a to 4c display the mean value of the Δv for the return maneuver combined with the Δv for the stopping maneuver, number of maneuvers performed, and cumulative reward for various along-track target points under the influence of atmospheric drag. For comparison, Fig. 4d to 4f provide the same results under both non-idealized conditions.



(a) Satellite trajectory for Case 1.

(b) Zoomed-in trajectory for Case 1.

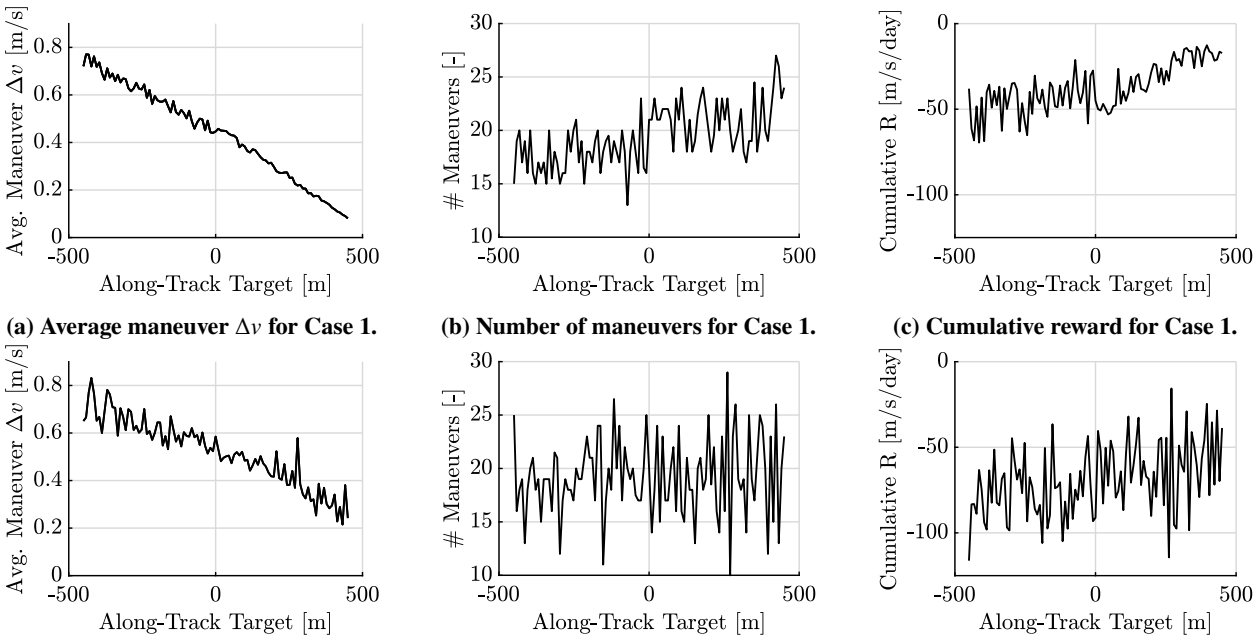
(c) Satellite trajectory for Case 2.

(d) Zoomed-in trajectory for Case 2.

Fig. 3 The satellite's trajectory when returning to $(0, 250, 0)$ m under atmospheric drag (Case 1), as well as the trajectory under both atmospheric drag and gravity model differences (Case 2).

The examination of the trajectory around the fixed target point (0, 250, 0) m in both cases shows that under non-idealized conditions, such as in Case 1 with atmospheric drag, the satellite does not return precisely to the target point. Starting from the slot center, it gradually shifts toward the positive along-track axis, returns near the target point, and then deviates toward the positive along-track direction again. This pattern reflects the influence of atmospheric drag, yet the deviation from the target point is relatively small. This indicates that even with the perturbative effects of atmospheric drag, the satellite's behavior remains qualitatively similar to the theoretical analysis. Although the drag causes a shift in the satellite's trajectory, it still manages to approximate the target point. This validates the effectiveness of the control approach under non-idealized conditions.

In Case 2, where the satellite experiences the combined effects of atmospheric drag and gravity model differences, its trajectory takes on a more chaotic character. The satellite not only exhibits a tendency to drift towards the positive along-track axis of the slot, as observed in Case 1, but also displays unpredictable excursions away from the target point. These excursions involve departures from the negative along-track direction, followed by lateral movement towards the positive along-track direction, and subsequent returns to the target point from its negative along-track and negative radial side. They necessitate additional return maneuvers to correct, consequently leading to an increase in the overall propellant cost. Furthermore, a striking feature of the trajectory in this scenario is the satellite's propensity for large spiral motions during its deviations from the designated path. This contrasts sharply with its behavior under the sole influence of atmospheric drag, where deviations appeared to follow a less chaotic pattern. The consideration of gravity model differences introduces an additional layer of complexity, resulting in the satellite's erratic deviations from its



(a) Average maneuver Δv for Case 1. **(b) Number of maneuvers for Case 1.** **(c) Cumulative reward for Case 1.**

(d) Average maneuver Δv for Case 2. **(e) Number of maneuvers for Case 2.** **(f) Cumulative reward for Case 2.**

Fig. 4 Average maneuver Δv, maneuvers count, and cumulative reward for various along-track target points under atmospheric drag (Case 1) and atmospheric drag with gravity model differences (Case 2).

intended path.

In contrast to the relatively smooth trends observed in the theoretical results, the average maneuver Δv, number of maneuvers performed, and cumulative reward exhibit more pronounced fluctuations under non-idealized scenarios. Notably, Case 2 results consistently show greater fluctuations compared to Case 1, which is in line with expectations as the presence of gravity model differences in Case 2 introduces added complexity to the satellite's dynamics. When comparing the Δv results between non-idealized scenarios and the idealized one, a decreasing trend is evident across all three sets of results. However, as the scenario becomes more complex, these trends show increased fluctuations. Additionally, the number of maneuvers for target points along the negative along-track axis under atmospheric drag does not exhibit a clear trend, while those on the positive along-track axis demonstrate an irregular and increasing pattern. Under both non-idealized effects, the number of maneuvers displays considerable fluctuations without a clear trend.

It is worth noting that the cumulative reward is directly influenced by both Δv and maneuver time interval, where the latter affects the number of maneuvers performed within a fixed time period. Consequently, cumulative reward represents a combined outcome of Δv and the number of maneuvers. In the context of both non-idealized scenarios, the cumulative reward shows an overall increasing trend, albeit with fluctuations, especially for Case 2. This trend suggests that, even though the values fluctuate, target points closer to the positive along-track side of the slot tend to yield higher rewards and lower propellant costs. However, it is important to recognize that real-world effects, such as atmospheric drag and gravity model differences, introduce complexities, challenging the straightforward relationship between target position and reward. Based on the results of the single-target point analysis under both non-idealized effects, it is observed that the target point associated with the lowest average maneuver Δv is located at (0,432,0) m, the target point with the fewest maneuvers is positioned at (0,270,0) m, and the target point yielding the highest cumulative reward is also situated at (0,270,0) m. However, these findings are subject to the substantial amount of noise in the data and cannot be interpreted as generalizable results. They provide initial insights into the potential outcomes of reinforcement learning when accounting for both atmospheric drag and gravity model differences.

VI. Maneuver Optimization with Reinforcement Learning

A. Q-Learning

Q-learning, an off-policy temporal-difference control algorithm, is a model-free reinforcement learning algorithm employed to determine the optimal action-selection policy within a finite Markov decision process. It is commonly used in situations where an agent interacts with an environment while remaining ignorant of the environment, subsequently learning optimal strategies through trial and error. The goal of Q-learning revolves around instructing the agent on selecting actions that maximize its cumulative rewards over time.

The key idea behind Q-learning is to maintain a Q-table that assigns a value to every possible state-action pair. During the learning process, the agent actively explores the environment, observes states, takes actions, receives rewards, and updates its Q-values based on the current reward and its new knowledge about the environment using Eq. (10) [22].

$$Q(S, A) \leftarrow Q(S, A) + \alpha[R + \gamma \max_{a} Q(S', a) - Q(S, A)], \tag{10}$$

where $Q(S, A)$ is the value of taking action A in the current state S. S' denotes the future state, and a is the action in state S' that is associated with the maximum Q-value based on current knowledge. α is the step-size parameter, and γ stands for the discount-rate parameter. Over time, these Q-values converge to represent the optimal state-action values, guiding the agent to choose actions that maximize cumulative rewards. The balance between exploration and exploitation in discovering better strategies and utilizing learned knowledge is determined by the policy selection strategy employed within the Q-learning architecture. Different policies can be implemented, ranging from those that favor exploration to those that are more exploitation-oriented. Once trained, the agent uses the Q-table to make informed decisions, selecting actions with the highest associated Q-values in each state.

B. Methodology for Maneuver Optimization

In this optimization phase, Q-learning is applied to optimize maneuvers, with the objective of minimizing propellant consumption. Within this context, the satellite is considered to be the agent, while its designated slot space constitutes the environment. The slot space discretizes into 18 states, each spanning a 20-degree sector. The available actions that the agent can take in each state correspond to along-track target points (See Fig. 5). This analysis incorporates three sets of actions: The first set involves five target points spaced evenly around the slot center with a 50-meter step; the second comprises eleven target points, while the third contains an expanded set of nineteen target points. The learning process involves a total of 10,000 episodes. Here, one "episode" in the context of Q-learning corresponds to a simulation conducted over a period of five days. Within each episode, a "step" is defined as the duration between the initiation of a return maneuver and the beginning of the subsequent return maneuver.

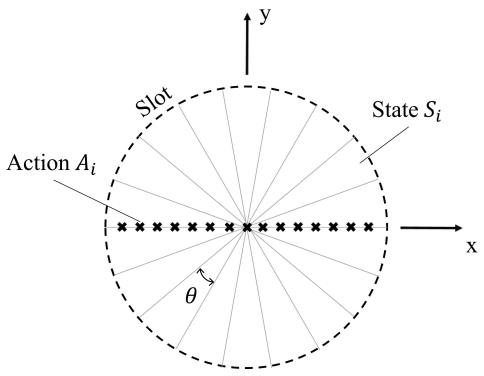


Fig. 5 The state and action spaces. Each state spans θ degrees, where θ is set to 20° in this work.

To evaluate the satellite's performance and guide the learning process, a reward function is employed based on Eq. (9). This function quantifies the satellite's efficient utilization of propellant, with the objective of minimizing Δv_1 and Δv_2), along with its ability to maintain its position within the slot space, thereby maximizing T. This corresponds to minimizing the absolute value of the reward. Given that the reward is negative, achieving this goal is equivalent to

maximizing the reward. Consequently, as the learning process progresses, the reward is expected to converge towards its optimal value. Note that "satellite" and "agent" are used interchangeably.

All of the Q-values in the Q-table initialize at zero, and Eq. (10) comes into play for Q-value updates as the satellite selects specific actions within certain states. The step size parameter α is a factor that decides the extent to which the newly acquired information will influence the existing Q-value. A constant α of 0.1 is maintained throughout the learning, striking a balance between utilizing prior knowledge and incorporating new information into the Q-value updates. The discount rate parameter γ influences how much the agent values future rewards compared to immediate ones. A constant γ of 0.9 encourages the satellite to more heavily consider future rewards, thereby promoting long-term planning and exploration to garner accurate estimates of rewards in the more distant future.

An *ϵ-greedy policy* is a strategy for action selection in a state where the agent must make decisions by balancing exploration and exploitation. The agent chooses an action based on the $\epsilon \in [0, 1]$ parameter, which is the probability of taking a random action in an ϵ-greedy policy. With probability ϵ, the agent explores by selecting a random action to discover potentially better options. With probability $1 - \epsilon$, the agent exploits by selecting the action with the highest estimated Q-value based on its current knowledge. As the agent gains more knowledge about the environment, it gradually reduces the value of ϵ to transition from an exploration-heavy strategy in the early stages to an exploitation-focused approach in the later stages. This adaptive adjustment of ϵ optimizes the agent's decision-making process over time. The calculation of ϵ for each episode is determined using Eq. (11).

$$\epsilon(k, \epsilon_0, L, p) = \epsilon_0 \exp\left(-\frac{1}{p}\left(\frac{k}{L}\right)^p\right), \tag{11}$$

where $k = 1, 2, 3, \ldots, 10000$ represents the episode number and $\epsilon_0 = 0.5$ is the value of ϵ for the first episode. This choice encourages more exploration at the beginning of the learning process. The parameter p controls the sharpness of the ϵ profile throughout the 10,000 episodes, and it is set to 2. Additionally, L governs the length of the profile, ensuring that the profile reaches the desired final ϵ value at the final episode. For this analysis, L = 3,500 is set, resulting in a final ϵ value of approximately 0.0084. This parameter configuration encourages convergence towards the optimal solution as the learning process progresses. The Q-learning algorithm is shown below in Alg. 1.

Algorithm 1 Q-learning algorithm

```
Algorithm parameters: α = 0.1, γ = 0.9
Initialize Q(S, A) = 0 for all state-action pairs
for each episode do
    Compute ϵ using Eq. 11
    for each step of episode do
        Observe S
        Choose A using ϵ-greedy policy
        Take action A, observe R and S′
        Update Q(S, A) using Eq. 10
        S ← S′
    end for
end for
```

Six distinct testing cases are undertaken. The first three solely consider the atmospheric drag effect, while the remaining three incorporate both atmospheric drag and gravity model differences. Testing Cases 1 and 4 employ a set of 5 along-track target points as actions. Testing Cases 2 and 5 utilize 11 target points, and Testing Cases 3 and 6 adopt 19 target points.

C. Optimization Results

Fig. 6 illustrates the satellite trajectories during the final episode of the learning process for six testing cases. The top row displays trajectories influenced solely by atmospheric drag, while the bottom row represents trajectories affected by both atmospheric drag and gravity model differences. In each row, the left column corresponds to scenarios with five evenly distributed target points, the middle column to eleven target points, and the right column to nineteen target points. The target points selected by the controller for each maneuver in every case are marked within the respective plots. Additionally, Fig. 7 depicts the moving average of the cumulative reward across the learning process for the same six

testing cases. This graph provides insights into how the reward evolves over time, reflecting the satellite's optimization progress under different conditions.

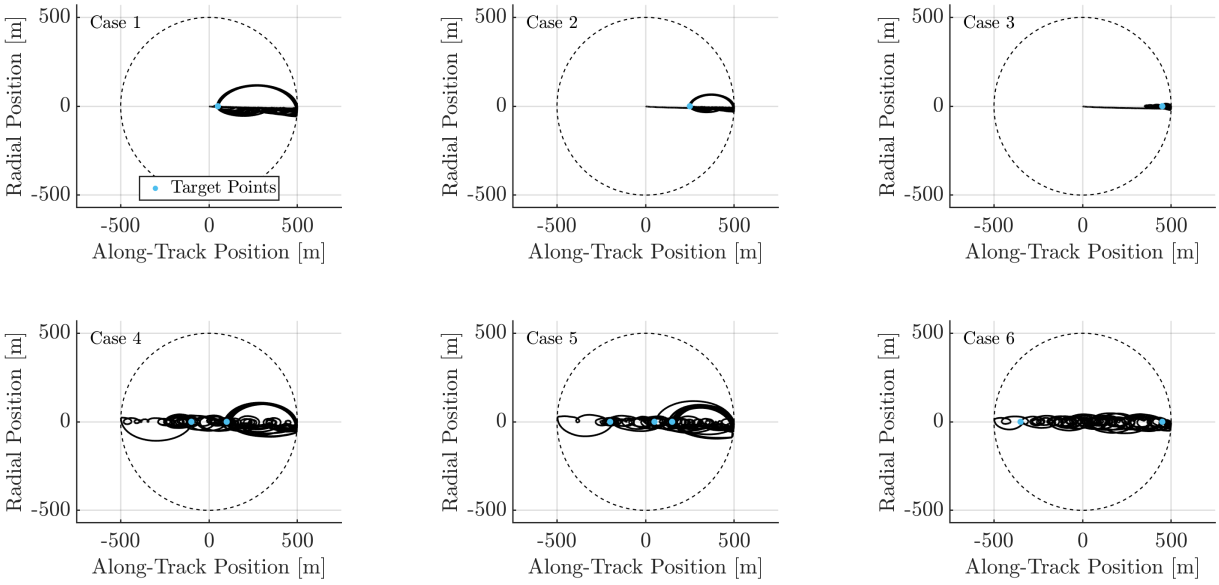


Fig. 6 Satellite trajectory during the final episode of the simulation for the six testing cases.

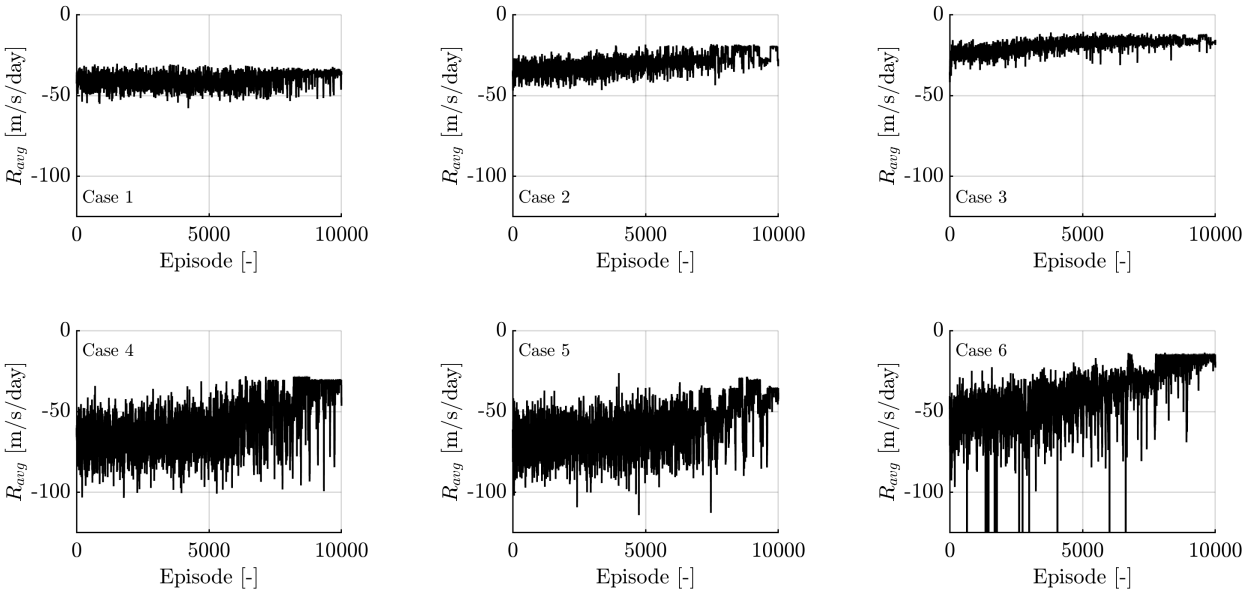


Fig. 7 Moving average of the cumulative reward over 10,000 simulation episodes for the six testing cases.

D. Discussion

The evaluation begins with Testing Cases 1, 2, and 3, which solely account for the impact of atmospheric drag. In Testing Case 1, the satellite's trajectory in the last episode reveals a return to the target point that is second from the end

on the positive along-track axis, while in Testing Cases 2 and 3, it returns to the target point at the farthest end on the positive along-track axis. Examining the moving average of cumulative rewards over 10,000 episodes for Testing Case 1 shows a consistent but fluctuating trend, with convergence to a higher value observed in the final 2,000 episodes. This pattern aligns with expectations, considering the initially high probability of selecting random actions that gradually decreases. Testing Cases 2 and 3 exhibit similar trends, albeit with an earlier onset of convergence as the number of target points increases. These outcomes are in line with the anticipated results, indicating that returning to a target point closer to the positive along-track side of the slot generally leads to reduced propellant costs and higher cumulative rewards. However, it is important to note that, for cases considering only atmospheric drag, reinforcement learning converges to a comparatively higher cumulative reward but not necessarily the highest possible value.

Next, the evaluation extends to the testing cases that encompass both atmospheric drag and gravity model differences, denoted as Testing Cases 4, 5, and 6. In contrast to the previous scenarios, the satellite's trajectory in the last episode reveals a more complex behavior. In the presence of gravity model differences, the satellite no longer consistently returns to a single target point, and the selected target points are not solely those situated on the positive along-track axis. As observed in the single-target-point analysis, the existence of gravity model differences introduces occasional deviations towards the negative along-track side of the slot, necessitating additional maneuvers to reposition the satellite within the slot space. Even when the satellite performs a return maneuver at the negative along-track side of the slot edge, it still tends to choose the nearest target point so that it minimizes Δv, mirroring its behavior at the right edge.

Specifically, in Testing Case 4, the satellite targets the farthest along-track positive axis point when it exits the slot in the positive along-track direction and the farthest along-track negative axis point when exiting in the negative along-track direction. Testing Case 5 presents slightly different results: the satellite returns to target points on the positive along-track axis from the positive edge of the slot, though not necessarily the farthest ones. From the negative along-track edge, it consistently returns to the farthest target point on the negative along-track axis. In Testing Case 6, the satellite returns to the farthest target point on the positive along-track axis when exiting from the positive side, and to the third farthest point on the negative along-track axis when departing from the negative side. The complexity introduced by gravity model differences leads to more diverse trajectory patterns, and further investigation is required to understand the underlying dynamics of the satellite.

The moving average of the cumulative reward exhibits a similar pattern across Testing Cases 4, 5, and 6. Initially, it displays a constant and fluctuating trend, which gradually converges to a higher value over the learning process. This convergence initiates earlier as the number of target points increases within each case. However, it is worth highlighting some peculiar observations in these cases. In both Testing Cases 4 and 6, an abrupt and substantial jump in cumulative reward occurs towards the end of the learning process. Conversely, Testing Case 5 presents a different behavior, with small consecutive jumps in cumulative reward, followed by a sudden decrease. At this juncture, the reasons behind these unexpected phenomena are unknown, warranting further investigation to elucidate the underlying dynamics and potentially refine the optimization process.

Upon completion of the learning process, the controller demonstrated a tendency to select the closest target point. This approach led to a significant reduction in propellant usage, evident from the enhanced reward values achieved at the end of the learning process. A reward value closer to 0 indicates lower propellant consumption. Notably, in Testing Case 6, the reward value converged to -14.4847 m/s/day, a substantial improvement over the -91.1616 m/s/day reward observed when consistently returning to the slot center. This translates to the propellant cost being approximately six times lower in comparison to scenarios without Q-learning implementation. Q-learning identified optimal solutions that align with the inclination to select the optimal target point, as suggested by both the theoretical analysis and single-target point analysis.

VII. Conclusion

This work aims to investigate control methods utilizing impulsive maneuvers to ensure a satellite remains within its designated orbital slot in the presence of dynamic mismatches. The objective is to design a controller that can effectively keep the satellite within the slot boundaries while minimizing propellant usage. Furthermore, the approach needs to be computationally efficient, allowing the satellite to autonomously determine its necessary control actions without relying on Earth-based resources.

Under the idealized scenario, the satellite consistently moves towards the positive along-track direction, resulting in a decreasing rate of propellant usage as the target point's along-track position increases. Under atmospheric drag conditions alone, the reward shows an overall increasing trend with the progression of the simulation, though with small oscillations. In scenarios with both atmospheric drag and gravity model differences, the satellite occasionally shifts

towards the negative along-track direction, introducing significant oscillations in the reward trend. In all six testing cases, the reward converges to a higher value by the end of the learning process compared to its initial value. However, unexpected sudden jumps in reward values are observed for Testing Cases 4 and 6, warranting further investigation. As a result of the learning process, the controller favored the nearest target point, which resulted in a propellant cost reduced to around one-sixth of what it would have been without Q-learning.

The results demonstrate the effectiveness of Q-learning in optimizing maneuvers for maintaining the satellite within its designated slot spaces. The Q-learning-based controller consistently achieves a higher reward compared to the strategy of returning to the slot center, signifying reduced propellant usage and more efficient position maintenance. In addition, while the computational time for this approach varies depending on the number of episodes set in the simulation, it remains within practical limits, making it a viable choice for autonomous onboard control.

This work relies on the assumption of a circular reference orbit for the applicability of the CW equations. Therefore, the controller designed in this study may not be suitable for satellites in highly eccentric orbits or missions involving large changes in altitude. For orbits with significant eccentricity or other complexities, additional modeling efforts may be required to accurately describe the relative motion. Furthermore, this research exclusively addresses two primary environmental factors contributing to deviation, namely, atmospheric drag and gravity model differences. It does not account for other potential causes of deviation, such as space debris and solar radiation pressure.

Future work should focus on investigating the reward fluctuations observed at the end of the learning process. Additionally, adapting the controller for missions with non-circular orbits, which introduce nonlinear dynamics, requires further study. This complexity may necessitate alternative strategies. Furthermore, exploring the adjustment of slot geometries to align with specific mission requirements may be essential.

Acknowledgments

This work was supported by funding from the University of Michigan Space Institute.

References

[1] Arnas, D., Casanova, D., and Tresaco, E., "2D Necklace Flower Constellations Applied to Earth Observation Missions," *Acta Astronautica*, Vol. 178, 2021, pp. 203–215. https://doi.org/10.1016/j.actaastro.2020.09.010.

[2] Jia, L., Zhang, Y., Yu, J., and Wang, X., "Design of Mega-Constellations for Global Uniform Coverage with Inter-Satellite Links," *Aerospace*, Vol. 9, No. 5, 2022, p. 234. https://doi.org/10.3390/aerospace9050234.

[3] Arnas, D., Lifson, M., Linares, R., and Avendaño, M. E., "Definition of Low Earth Orbit Slotting Architectures Using 2D Lattice Flower Constellations," *Advances in Space Research*, Vol. 67, No. 11, 2021, pp. 3696–3711. https://doi.org/10.1016/j.asr.2020.04.021.

[4] Lifson, M., Arnas, D., Avendaño, M., and Linares, R., "Low Earth Orbital Slotting: Implications for Orbit Design and Policy," *8th Annual Space Traffic Management Conference*, The International Academy of Astronautics, Austin, Texas, USA, 2022.

[5] Badrinath, S., Li, M. Z., and Balakrishnan, H., "Integrated Surface–Airspace Model of Airport Departures," *Journal of Guidance, Control, and Dynamics*, Vol. 42, No. 5, 2019, pp. 1049–1063. https://doi.org/10.2514/1.G003964.

[6] Saraf, A., Sui, V., Chan, K., Luch, N., Popish, M., Lohn, E., Huang, B., Levy, B., Rose, M., Balakrishnan, H., and Idris, H., "Benefits Assessment of Integrating Arrival, Departure, and Surface Operations with ATD-2," *36th Digital Avionics Systems Conference*, Institute of Electrical and Electronics Engineers, St. Petersburg, Florida, USA, 2017. https://doi.org/10.1109/DASC.2017.8101998.

[7] SpaceX, "Starlink," , 2023. URL https://www.starlink.com/, (last accessed on November 19, 2023).

[8] OneWeb, "Air," , 2023. URL https://oneweb.net/solutions/government/air, (last accessed on November 19, 2023).

[9] Telesat, "Telesat Lightspeed," , 2023. URL https://www.telesat.com/leo-satellites/, (last accessed on November 19, 2023).

[10] Walker, J. G., "Some Circular Orbit Patterns Providing Continuous Whole Earth Coverage," *Journal of the British Interplanetary Society*, Vol. 24, 1971, p. 369–384.

[11] Rider, L., "Analytic Design of Satellite Constellations for Zonal Earth Coverage using Inclined Circular Orbits," *Journal of the Astronautical Sciences*, Vol. 34, 1986, p. 31–64.

[12] Mortari, D., Wilkins, M. P., and Bruccoleri, C., "The Flower Constellations," *The Journal of the Astronautical Sciences*, Vol. 52, 2004, p. 107–127. https://doi.org/10.1007/BF03546424.

[13] Eckstein, M. C., "Geostationary Orbit Control Considering Deterministic Cross Coupling Effects," *International Astronautical Congress*, Dresden, Germany, 1990.

[14] Soop, E. M., *Handbook of Geostationary Orbits*, Springer Dordrecht, 1994, Chaps. 6, 7.

[15] Eckstein, M. C., "Optimal Station Keeping by Electric Propulsion with Thrust Operation Constraints," *Celestial Mechanics*, Vol. 21, No. 2, 1980, p. 129–147. https://doi.org/10.1007/bf01230889.

[16] Gazzino, C., Arzelier, D., Losa, D., Louembet, C., Pittet, C., and Cerri, L., "Optimal Control for Minimum-Fuel Geostationary Station Keeping of Satellites Equipped with Electric Propulsion," *20th IFAC Symposium on Automatic Control in Aerospace*, International Federation of Automatic Control, Sherbrooke, Quebec, Canada, 2016. https://doi.org/10.1016/j.ifacol.2016.09.065.

[17] de Bruijn, F. and Gill, E., "Analysis of Relative Motion of Collocated Geostationary Satellites with Geometric Constraints," *5th International Conference on Spacecraft Formation Flying Missions and Technologies*, German Aerospace Center Space Operations Center, Munich, Germany, 2013.

[18] de Bruijn, F. J., Theil, S., Choukroun, D., and Gill, E., "Geostationary Satellite Station-Keeping using Convex Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 39, No. 3, 2016, p. 605–616. https://doi.org/10.2514/1.g001302.

[19] Guelman, M. M., "Geostationary Satellites Autonomous Closed Loop Station Keeping," *Acta Astronautica*, Vol. 97, 2014, p. 9–15. https://doi.org/10.1016/j.actaastro.2013.12.009.

[20] Lavretsky, E., *Robust and Adaptive Control: With Aerospace Applications*, 1st ed., London: Springer Nature, 2012.

[21] Diehl, M., Ferreau, H. J., and Haverbeke, N., *Efficient Numerical Methods for Nonlinear MPC and Moving Horizon Estimation*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, Vol. 384, p. 391–417. https://doi.org/10.1007/978-3-642-01094-1_32.

[22] Sutton, R. S. and Barto, A. G., *Reinforcement Learning: An Introduction*, Adaptive Computation and Machine Learning, The MIT Press, 2018.

[23] Breger, L. and How, J. P., "SAFE Trajectories for Autonomous Rendezvous of Spacecraft," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 5, 2008, p. 1478–1489. https://doi.org/10.2514/1.29590.

[24] Kankashvar, M., Bolandi, H., and Mozayani, N., "Multi-agent Q-learning Control of Spacecraft Formation Flying Reconfiguration Trajectories," *Advances in Space Research*, Vol. 71, No. 3, 2023, p. 1627–1643. https://doi.org/10.1016/j.asr.2022.09.034.

[25] Sullivan, C. J. and Bosanac, N., "Using Reinforcement Learning to Design a Low-Thrust Approach into a Periodic Orbit in a Multi-Body System," *AIAA SciTech Forum*, American Institute of Aeronautics and Astronautics, Orlando, Florida, USA, 2020. https://doi.org/10.2514/6.2020-1914.

[26] Clohessy, W. H. and Wiltshire, R. S., "Terminal Guidance System for Satellite Rendezvous," *Journal of Aerospace Sciences*, Vol. 27, No. 9, 1960, pp. 653–658. https://doi.org/10.2514/8.8704.

[27] Prussing, J. E. and Conway, B. A., *Orbital Mechanics*, Oxford University Press, 2013, Chap. 10.