



Multi-UAV Tactic Switching via Model Predictive Control and Fuzzy Q-Learning

Original
Article

Ahmed T. Hafez¹, Sidney N. Givigi², Shahram Yousefi³, Mohamad Iskandarani²

¹Military Technical College, ²Royal Military College of Canada, ³Queen's University

Key words:

Model Predictive Control, Autonomous robotics, Cooperative robotics, Formation tactic, Dynamic Encirclement tactic

Corresponding Author: Hafez, Ahmed, Taimour, PhD Head of Aircraft Armament Department, **Tel :** +201005115209 **Email:** a.taimour@mtc.edu.eg

Abstract

Teams of cooperative Unmanned Aerial Vehicles (UAVs) require intelligent and flexible control strategies to allow the accomplishment of a multitude of challenging group tasks. In this paper, we introduce a solution for the problem of tactic switching between the formation flight tactic and the dynamic encirclement tactic for a team of N cooperative UAVs using a predictive decentralize control approach. Decentralized Model Predictive Control (MPC) is used to generate tactics for a team of N UAVs in simulation and real-world validation. A high-level Linear Model Predictive Control (LMPC) policy is used to control the UAV team during the execution of a desired formation, while a combination of decentralized LMPC and Feedback Linearization (FL) is applied to the UAV team to accomplish dynamic encirclement. The decision of switching from one tactic to the other is derived by a fuzzy logic controller, which, in its turn, is derived by a Reinforcement Learning (RL) approach. The main contributions of this paper are: (i) solution of the problem of tactic switching for a team of cooperative UAVs using LMPC and a fuzzy controller derived via RL; (ii) simulations demonstrating the efficiency of the method; and (iii) implementation of the solution to on-board real-time controllers on Qball-X4 quadrotors.

I. INTRODUCTION

Recently, the increase in cooperative UAVs' capabilities and flexibility provides an opportunity for new operational paradigms. These vehicles are developed to be capable of working in different circumstances and weather with some assistance of human control and have the ability to handle different complicated or uncertain situations. Therefore, the efforts in research and development have gained great prominence throughout the world [1, 2, 3].

UAV tactics are used to ensure the success of the cooperative UAVs in performing their required applications. These tactics are defined as the general strategies used by the members of a UAV team to achieve a desired outcome, and they can be either centralized, where a group of UAVs receives their commands from one centralized decision maker (i.e., a leader UAV in the fleet or a ground control station), or decentralized, in which each member in the UAV team is responsible for making its own decisions guarantying the execution of the whole team operation [4].

Research and experiments have dealt with the different strategies that can be carried out by a team of cooperative

UAVs in different operations, and a wide variety of approaches to effectively implement these tactics have been proposed. From different point of views, UAVs' tactics are generally classified into four main categories [4]:

- Swarming
- Formation Reconfiguration
- Task/Target Assignment
- Dynamic Encirclement

In this paper, a solution for the problem of tactic switching for a group of cooperative UAVs from a desired formation configuration tactic to a required dynamic encirclement tactic around an invading target.

Formation reconfiguration is defined as the dynamic ability of a UAV team to change their formation according to the surrounding circumstances and due to the response to different external factors [5].

The new UAVs' formation must guarantee safety, compatible with the UAV dynamics and governed by a time constraint to pass between obstacles (Collision avoidance). Formation reconfigurations have different behaviors according to the following factors affecting the UAV team [5]:



- Changing the position of UAVs during mission execution.
- Combining small groups to form a large group according to the required mission.
- Breaking a large group into smaller groups to perform more than one application at the same time.

Some of the main formation structure approaches are:

1) Leader Follower Approach: In this approach, some UAVs in the team are designed as leaders while others are followers. Its main advantage is that it can be easily understood and implemented but its main disadvantage is that it is not robust in the case of leader failure [6, 7, 8].

2) Virtual Structure Approach: In this approach, the whole team is treated as a single rigid body and instead of following a certain path, each UAV follows a certain moving point which allows them to be attached to each other. In this case, the formation is treated as a single object which increases robustness. On the other hand, it can only perform synchronized maneuvers and it cannot deal with obstacle avoidance constraints or reject external disturbances [9].

3) Behavior-Based Approach: A desired behavior is designed for each UAV, including the required information for mission, goal seeking and collision avoidance. The control action of each UAV is a weighted average of the control for each behavior and it is suitable for uncertain environments but it lacks theoretical guarantees of stability [10].

There is an extensive amount of research in the field of formation reconfiguration for multiple cooperative UAVs. In the presence of bad weather conditions, obstacles, aerial jamming of communication channels between UAV team members and external threats such as Early Warning Radar (EWR) and Target and Fire System (TF) which prevent the team from performing their required mission, formation reconfiguration appears as an optimal solution for the UAV team to fulfill their mission by switching from a particular geometric pattern to another [11, 12, 13, 14, 15, 16].

For instance, a formation of three UAVs can be controlled by a centralized multi-layer control scheme to follow a pre-determined trajectory in tracking missions. Each layer in the control scheme is responsible for generating the required control input for each member in the formation [15]. In [17], a consensus based formation control strategy is used along with a collision avoidance strategy based on artificial potential approach to control the formation of a group of cooperative UAVs during flight to a desired position in the presence of obstacles.

In [18], a hybrid approach based on the Lyapunov Guidance Vector Field (LGVF) and the Improved

Interfered Fluid Dynamical System (IIFDS) is applied to solve the problems of target tracking and obstacle avoidance in 3-D cooperative path planning for cooperative UAVs. Moreover, in [11], based on Lyapunov functional approach and algebraic Riccati equation technique, an approach to determine the gain matrices in the formation protocol was proposed solving the time-varying formation control problems for cooperative UAVs swarm systems with switching interaction topologies.

In [19], a formation of cooperative UAVs is controlled in a decentralized manner to solve the problem of multi-UAV tracking effectively and reasonably, while in [16], a robust control algorithm accompanied with a higher level path generation method is used to control the structure of a group of cooperative UAVs. The goal is to perform formation change maneuvers with a guaranteed safe distance between the different members of the team throughout the whole mission. The robust control ensures the stability of the formation during maneuver while the path generation method provides the vehicle with the safe paths.

On the other hand, a dynamic encirclement tactic is defined as the situation in which an invading target is isolated in order to maintain awareness and containment of it by a group of multiple cooperative UAVs. During the encirclement phase, cooperative UAVs assume desired positions around the target to restrict its movements [20, 21].

Much of the current literature on multiple cooperative UAVs pays particular attention to dynamic encirclement. In real life, dynamic encirclement has been used in defending a secure airspace against an invading aircraft, maintaining surveillance over a ground target and in protecting the borders against invading targets [20, 22, 23, 24].

For instance, in [25], the problem of capturing a target with a group of cooperative UAVs was solved using a complete dynamic network topology based on the relative distance between the members of the team. Moreover, at least one member of the team acts as a leader and can acquire the target information sending it to the other members. In [26] a distributed cooperative control method is proposed to solve the encirclement problem based on a cyclic pursuit strategy which requires the ability of all the team members to have complete target information which can be considered as a limitation constraint of the real life scenario.

Recently, researchers represent and solve the problem of dynamic encirclement as an optimal control problem allowing for states and control inputs constraints.

In [27], a linear quadratic regulator control technique was used to solve the problem of capturing a ground target with a team of UAVs in the presence of wind disturbances with known velocity and direction. In [28], a nonlinear model predictive controller was used to solve the problem of dynamic encirclement around moving targets in the

presence and absence of communications. Also, the analysis of the system stability was introduced supporting the designed controller. In [29, 30], an approach based on the combination of linear model predictive control with feedback linearization technique is used to control a team of N cooperative UAVs performing a dynamic encirclement tactic against an invading stationary and movable targets in simulation and real time.

The design of a controller to modify the response and behavior of a non-linear system to meet certain performance requirements is a challenging problem in many control applications. Model Predictive Control (MPC) is characterized by its ability to handle difficult constraints on controls and states (Inputs-Outputs) while dealing with multi-variable systems. It can easily adapt to new contexts due to rolling horizon [31, 32]. Different types of MPC were used to solve the optimization control problem for a group of autonomous vehicles in different applications and missions [33, 34, 35, 36, 37].

For instance, Nonlinear Model Predictive Control (NMPC) can be used to control the formation of a fleet of UAVs in the presence of obstacles and collision avoidance [35], and to develop the formation guidance for a team of UAVs, where the controllers predict the behavior of the system over the receding horizon and generate the optimal control commands over the horizon for flight formation and inter-collision avoidance in the presence of control input and state constraints [38, 39].

In [40], a combination of decentralized linear MPC and Taylor series linearization was successfully applied to a team of UAVs to solve the problem of the dynamic encirclement around a stationary target. Also, the problem of dynamic encirclement for a team of cooperative UAVs surrounding a stationary and movable target was solved using decentralized NPMC in [28] and [41]. A theoretical analysis for the stability of the designed NMPC was introduced proving the ability of the designed controller to execute the required mission.

Recently, control designers started to investigate the effect of adding a learning algorithm to the predictive decentralized approach especially MPC. Applying the learning algorithm to the MPC will improve the performance of the system and guarantee safety, robustness and convergence in the presence of states and control inputs constraints [42]. In [43], the problem of formation reconfiguration for a group of multiple cooperative UAVs using a combination between state transformation technique and decentralized Learning Based Model Predictive Control (LBMPC) controller in the presence of uncertainties and in an obstacle-loaded environment was solved. The introduced controller succeed to learn the unmodeled dynamics using the learning approach and solve the optimization control problem using the MPC. Moreover, in [44], a team of

cooperative UAVs succeeded in tracking a required target with a desired formation using LBMPC technique.

In this paper, the decision of switching from the formation tactic to the encirclement tactic is controlled by fuzzy controller in a decentralized manner. Each UAV has the ability to decide whether to switch or continue in the formation tactic according to the surrounding constraints. Fuzzy systems have been used in switching systems for a long time, almost since its inception [45]. However, one of the drawbacks of using fuzzy systems is the difficulty of tuning the inference systems as it may be very time consuming [46]. Therefore, in order to make such systems more useful, a way of tuning the controllers automatically may be explored.

Our main contribution in this paper is solving the problem of UAV tactic switching for a team of cooperative UAVs in simulation and real-time implementation on-board Qball-X4 quadrotors. An LMPC is used to control the UAV team during formation flight, while a combination of decentralized LMPC and FL is used to solve the problem of dynamic encirclement. The switching decision is controlled by a fuzzy logic controller derived using a fuzzy Q-learning approach [47, 48] according to the surrounding factors. We occupy ourselves with a decentralized high-level controller, where each team member generates the required path necessary to respect the line-of-breast formation and encirclement conditions. All formation and encirclement tasks are accomplished autonomously.

This paper is organized as follows. In section II, the problems of formation and dynamic encirclement tactics are formulated, then, specific control objectives are defined and the dynamic feedback linearization is introduced. In section III, the outline for the designed control policy and all the system constraints are presented, while in section IV, the fuzzy logic controller used in taking the decision of switching is represented. The simulation results are presented in section V, while in section VI, the experimental setup and the experimental results are introduced. The discussion about the simulation results and the experimental results are presented in section VII, finally, the introduced work is concluded and some of the future objectives are presented VIII.

II. PROBLEM FORMULATION

We consider the problem of tactic switching for a team of N UAVs. The members of the team form a linear formation and advance in a leader-follower manner then switch to dynamic encirclement. In this paper, the problem of tactic switching is limited to two dimensional space, although the problem could be extended to the three dimensional case with increased computational demands and minimal modification to the controllers [20]. The problem consists of two main parts: the first one deals with the formation tactic using LMPC, while the second part

deals with the dynamic encirclement tactic using a combination of FL and LMPC. In this section, the two main problems will be presented along with the system identification technique used to find vehicle dynamics.

A. System Description

Before starting to tackle the problem of tactic switching, our system need to be described. The system under study consists of N cooperative UAVs characterized by their radius of communication (r_c) and sensing radius (r_s). Each UAV in the fleet has the possibility to communicate only with its neighbor if the separating distance between them is less than r_c . On the other hand, each UAV in the team has the ability to sense the presence of a target or obstacle if the separating distance between the UAV and the target or obstacle is less than r_s . One should also notice that usually in real applications $r_c > r_s$. Furthermore, if a UAV wants to communicate with another vehicle that is outside its communication radius, but there is a vehicle that is a neighbor for both, an *ad-hoc* network may be defined and communication established. Fig.1 represents the different radii of communication and sensing for two UAVs.

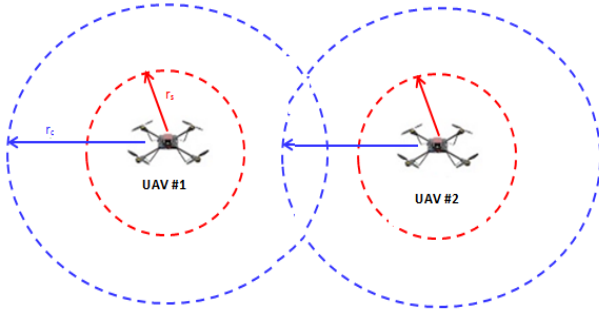


Fig.1: Radii of communication and sensing for two UAVs.

B. System Identification

We start our problem formulation by identifying a linear model for each UAV in the team using a system identification technique based on a least-squares algorithm. A linear second order process model was found from collected flight data with small steady-state error [40].

Each UAV in the team will be characterized by the following state-space representation [40]:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1.00 & 0 \\ 0 & 0 & 0 & 1.00 \\ -2.60 & 0 & -1.70 & 0 \\ 0 & -1.40 & 0 & -0.50 \end{bmatrix} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} +$$

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_d \\ Y_d \end{bmatrix} \quad (1a)$$

$$\bar{Y} = \begin{bmatrix} 1.75 & 0 & 0 & 0 \\ 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} \quad (1b)$$

where X_d and Y_d are the desired positions of each UAV while x and y are the current positions of each UAV during flight. Also, X_d and Y_d are the inputs to the low-level control of the UAV while x and y are the outputs of this system. The subscript i denotes the i^{th} UAV in the team.

C. Formation Flight

During the formation tactic, the UAVs form a linear formation and advance in a leader-follower manner while maintaining the desired separation distance and matching speed with its neighbors. By subtracting equations (1) for each UAV in the fleet, with the same equations related to its neighbor, the following error dynamics in state-space form is introduced:

$$\begin{bmatrix} \dot{E}_x \\ \dot{E}_y \\ \ddot{E}_x \\ \ddot{E}_y \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1.0 & 0 \\ 0 & 0 & 0 & 1.0 \\ -2.6 & 0 & -1.7 & 0 \\ 0 & -1.4 & 0 & -0.5 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ \dot{E}_x \\ \dot{E}_y \end{bmatrix} +$$

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{d1} \\ X_{d2} \\ Y_{d1} \\ Y_{d2} \end{bmatrix} \quad (2)$$

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} 1.75 & 0 & 0 & 0 \\ 0 & 0.50 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ \dot{E}_x \\ \dot{E}_y \end{bmatrix} \quad (3)$$

where E_x and E_y are the errors in x and y respectively between each two successive UAVs in the formation, and \dot{E}_x and \dot{E}_y are their respective time derivatives. Equation (3) is the output equation that gives us the distance in x and y between the UAVs. LMPC is used to maintain these errors at a desired separating distance between all the

members of the team. The choice of leader UAV is used to form the desired formation during the formation tactic phase.

D. Dynamic encirclement

In section II-B, the system was identified for the Cartesian movement of the Qball-X4 and set up the transformation to the new set of states that allow each UAV to output the radius and angular velocity of each UAV. Since the control of each UAV in the team is done using the radius of encirclement and the angular velocity, a state transformation from Cartesian to Polar coordinates is applied giving a new state vector $\bar{X} = [r \ \phi \ \dot{r} \ \dot{\phi}]^T$. The resultant system is represented in the following form:

$$\dot{\bar{X}} = f(\bar{X}, x_o, y_o, X_d, Y_d) \quad (4)$$

where $f(\cdot)$ is a nonlinear function combining the inputs and states of the system \bar{X} , x_o is the translation in x -direction, and y_o is the translation in y -direction. This translation represents the coordinates of the moving target. Moreover, r represents the radius of encirclement, ϕ the angle between the target and the encircling UAV, and $\dot{\phi}$ is the angular velocity of the UAV. To complete the transformation, the states x, y, \dot{x} and \dot{y} are replaced with their polar equivalents:

$$\begin{aligned} x &= r \cos \phi + x_o \\ y &= r \sin \phi + y_o \\ \dot{x} &= \dot{r} \cos \phi - r \dot{\phi} \sin \phi \\ \dot{y} &= \dot{r} \sin \phi + r \dot{\phi} \cos \phi \end{aligned}$$

Then, the resultant equations characterized by $[\dot{x} \ \dot{y} \ \ddot{x} \ \ddot{y}]^T$ are multiplied using a transformation matrix T :

$$\begin{bmatrix} \dot{r} \\ \dot{\phi} \\ \ddot{r} \\ \ddot{\phi} \end{bmatrix} = T \begin{bmatrix} \dot{x} \\ \dot{y} \\ \ddot{x} \\ \ddot{y} \end{bmatrix} \quad (5)$$

$$T = \begin{bmatrix} \cos \phi & \sin \phi & 0 & 0 \\ -\sin \phi & \cos \phi & 0 & 0 \\ r & r & 0 & 0 \\ \omega \sin \phi & \omega \cos \phi & \cos \phi & \sin \phi \\ \frac{\dot{r} \sin \phi}{r^2} - \frac{\omega \cos \phi}{r} & \frac{-\dot{r} \cos \phi}{r^2} - \frac{\omega \sin \phi}{r} & \frac{-\sin \phi}{r} & \frac{\cos \phi}{r} \end{bmatrix}$$

Now, a nonlinear system with output states r and $\dot{\phi}$ for each UAV in the team is defined. The nonlinear system in (5) will be linearized using an FL technique and the linearized system will be controlled using LMPC to solve the problem of dynamic encirclement.

The main objective of our designed controller is to achieve the required constraints mentioned in the following equations [28]:

$$\mathbf{C1}) \lim_{t \rightarrow \infty} |r_i(t) - R_D| = 0 \quad \forall i \leq N \quad (6)$$

$$\mathbf{C2}) \lim_{t \rightarrow \infty} |\dot{\phi}_i(t) - \dot{\phi}_D| = 0 \quad \forall i \leq N \quad (7)$$

$$\mathbf{C3}) \lim_{t \rightarrow \infty} |\phi_{i+1}(t) - \phi_i(t)| = \frac{2\pi}{N} \quad \forall i \leq N \quad (8)$$

The above equations represent the model constraints that must be achieved by the controller: Condition **C1** states that each UAV in the team maintains a desired distance R_D from the target, while Condition **C2** states that each UAV in the team maintains a desired angular velocity $\dot{\phi}_D$ around the target. Condition **C3** states that each member in the team spreads itself evenly in a circular formation around the target. The LMPC controller respects these constraints when accomplishing dynamic encirclement.

Finally, the goal of our simulation is to guarantee the success of the cooperative UAVs to switch from the formation tactic to the encirclement tactic. According to different constraints, we will smoothly combine the control signals from the LMPC used during the formation tactic and the LMPC used in dynamic encirclement to achieve a stable transition from one tactic to another using a piece-wise approach.

E. Feedback Linearization

In section II, the model is identified for the Cartesian movement of the Qball-X4 and set up the transformation to the new set of states that allows us to output the required radius of encirclement and angular velocity for each UAV in the team. The objective of FL is to linearize (4) and represent it in the standard state-space form allowing us to use the decentralized LMPC during the encirclement phase.

Feedback Linearization (FL) is a common approach used in the control of nonlinear systems, resulting in a standard linear state space system, which allows the use of linear control techniques to implement the system in real time. Different combinations of feedback linearization and various types of controllers were used to compensate the nonlinear dynamics of cooperatives UAVs during their various applications [49, 50, 51].

In this paper, we concentrate on creating a linear map between a set of new inputs (u_1, u_2) and the desired outputs ($r, \phi, \dot{\phi}$) with respect to the original inputs (X_d, Y_d) [52].

By isolating the non-linear dynamics of the system, the original inputs are expressed as follows:

$$\begin{aligned} X_d &= \frac{f_{1n}(u_1, u_2) - f_1(\bar{X}, x_o, y_o)}{g_1(\bar{X}, x_o, y_o, X_d, Y_d)}, \\ Y_d &= \frac{f_{2n}(u_1, u_2) - f_2(\bar{X}, x_o, y_o)}{g_2(\bar{X}, x_o, y_o, X_d, Y_d)} \end{aligned}$$

where $f_1(\cdot)$, $f_2(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ are nonlinear functions in the states and control inputs, while $f_{1n}(u_1, u_2)$ and $f_{2n}(u_1, u_2)$ are linear functions

compensating the nonlinear dynamics of the system. The functions f_{1n} and f_{2n} are chosen to respect the controllability and observability of the closed loop system allowing the use of LMPC.

We may write the resultant linear system as ^[53]

$$\dot{\bar{X}} = g(\bar{X}, x_o, y_o, u_1, u_2) \quad (9)$$

where $g(\cdot)$ is a linear function found through FL substitution. The final linear process model may be represented in state-space form:

$$\dot{X} = \bar{X} + BU; \quad \bar{Y} = C\bar{X} \quad (10)$$

where U is $[u_1 \ u_2]^T$, \bar{X} is $[r \ \phi \ \dot{r} \ \dot{\phi}]^T$, \bar{Y} is the output vector holding r and $\omega = \dot{\phi}$ and matrices A, B, C ensure controllability and observability of the state-space.

This linear system is shown in Fig. 2, where the LMPC produces new control inputs to control the system during encirclement.

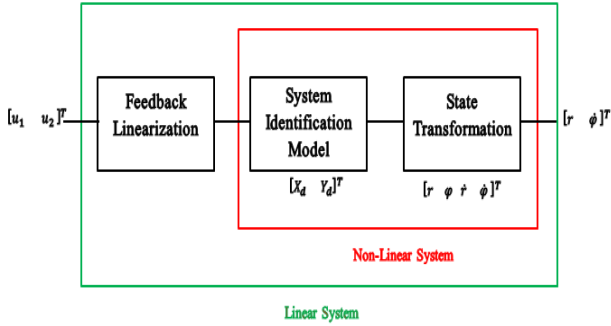


Fig. 2: Summary of system identification, state transformation and FL.

III. CONTROL DESIGN

Now, the MPC strategy will be applied to solve the problem of tactic switching of multiple UAVs. In this section, we outline the design of an LMPC controlling a group of N UAVs to form a line abreast formation tactic and the design of a decentralized LMPC controlling the UAV group encircling a desired target. A smoothing time t_{sm} , used in the piece-wise control signal mixing, is introduced to guarantee the stability of the system at the time of switching from the formation controller to the encirclement controller.

A. Formation flight

An LMPC is used to maintain the desired distance between each UAV and its neighbor during the flight. The first UAV is considered the leader of the fleet while the other UAVs maintain the desired distance in x and y coordinates. The cost function of the LMPC controller is minimized according to the weights of the outputs, and is given as follows:

$$J(\bar{Z}, \Delta u) = \sum_{i=0}^{p-1} \Gamma^T Q \Gamma + \Delta u(k+i|k)^T R \Delta u(k+i|k) \quad (11)$$

The components of the cost function are:

$$\Gamma = \bar{Z}(k+i+1|k) - \bar{D}(k+i+1|k) \quad (12)$$

where p is the prediction horizon and $\bar{Z}(k+i+1|k)$ is the state vector containing the error dynamics as highlighted in (2) and (3). These states are predicted for time $k+i+1$ at time k . Furthermore, $r(k+i+1|k)$ is the reference sampled for time $(k+i+1)$ at time k ; $\Delta u(k+i|k)$ is the manipulated variables rate calculated for time $k+i$ at time k ; Q and R are positive semi-definite matrices that hold the weights for the output variables and the manipulated variables rate respectively. The references \bar{D} represent the desired distances between the UAVs in x and y and their desired rate of change.

B. Dynamic encirclement

The FL method is used to linearize the encirclement model of each UAV in the team, which allows the use of a suitable LMPC in real-time. The optimization problem became convex in nature which allows for faster computational performance compared to its nonlinear counterparts. Moreover, the UAVs accomplish the task of encircling a target autonomously with the user only initiating the system.

In order to accomplish dynamic encirclement using a team of N UAVs, conditions $(C1, C2, C3)$ in section II-D must be met. Each UAV must control its radius of encirclement r , its angular speed w , and its angular separation relative to other members of the team.

Each UAV in the team is cognisant of the leading and lagging UAV in the formation. ϕ_{lead} represents the angular difference between the member being considered and the one in front of it, while ϕ_{lag} represents the angular difference with the one behind it. We also assume that the UAVs' indices are determined by their initial position, such that UAV i leads UAV $i-1$, UAV $i+1$ leads UAV i , etc. in the formation. The optimization cost function of our proposed decentralized LMPC is represented in the quadratic form as follows:

$$J_i^T = J_i^1 + J_i^2 + J_i^3 \quad (13)$$

where J_i^T is the total cost function of i^{th} UAV in the fleet, J_i^1 represents the encirclement objectives $C1$ and $C2$ in section II-D, while J_i^2 and J_i^3 represents the encirclement objective $C3$ for the lagging and leading UAVs, respectively, and are given by:

$$J_i^1 = \sum_{j=1}^M (\alpha(r_i(j\delta; t) - R_D)^2 + \beta((\dot{\phi}_i(j\delta; t) - \dot{\phi}_{Di})^2) \quad (14a)$$

$$J_i^2 = \sum_{j=1}^M (\rho((\phi_i(j\delta; t) - \phi_{lag}(j\delta; t)) - \frac{2\pi}{N})^2) \quad (14b)$$

$$J_i^3 = \sum_{j=1}^M (\zeta((\phi_{lead}(j\delta; t) - \phi_i(j\delta; t)) - \frac{2\pi}{N})^2) \quad (14c)$$

where $r_i(j\delta; t)$ and $\phi_i(j\delta; t)$ are the predicted radius of encirclement from the target and the vehicle's angular

speed j time steps into the future. Also α , β , ρ and ζ are positive constants where α is the radius gain, β is the angular speed gain, ρ is the angular separation gain for lagging UAV while ζ is the angular separation gain for leading UAV, $\phi_{lag}(j\delta; t)$ and $\phi_{lead}(j\delta; t)$ are the angles of the lagging and leading UAVs in the formation respectively. Since only the current position of the leading and lagging UAVs are known, $\phi_{lag}(t)$ and $\phi_{lead}(t)$, the remaining future angles must be estimated by solving the following differential equations:

$$\dot{\phi}_{lead}(t) = \dot{\phi}_D \quad (15)$$

$$\dot{\phi}_{lag}(t) = \dot{\phi}_D \quad (16)$$

One should notice that the conditions described in section II may make the problem infeasible as they might be conflicting (if the angular difference between vehicles is smaller than the required - condition C3 - one vehicle may need to slow down, which may conflict with condition C2). In order to mitigate this potential problem, one of the conditions needs to be relaxed. This is done by using the adaptive strategy derived in [52].

Similarly to [52], we will derive the controller for N vehicles as follows, the desired angular separation $\Delta\phi_D$ is given as:

$$\Delta\phi_D = 2\pi/N \quad (17a)$$

$$\Delta\phi_{ij}(t) = \phi_j(t) - \phi_i(t) \in [0, 2\pi] \quad (17b)$$

where ϕ_i represents the angle of the i^{th} UAV, and ϕ_j is the angle of the j^{th} UAV. We are assuming that the angular speed for each vehicle in the formation is varying towards the desired angular speed leading to the convergence of the error between the current and desired angle of separation for N UAVs. We calculate the error in the formation as follows:

$$e_i(t) = \Delta\phi_{i(i+1)}(t) - \Delta\phi_D \quad (18a)$$

$$\dot{e}_i(t) = \dot{\phi}_{i+1}(t) - \dot{\phi}_i(t) \quad (18b)$$

and the chosen Lyapunov candidate function is

$$V(t) = \frac{1}{2}e_1^2(t) + \frac{1}{2}e_2^2(t) + \dots + \frac{1}{2}e_N^2(t) \quad (19a)$$

$$\begin{aligned} \dot{V}(t) &= e_1(t)\dot{e}_1(t) + e_2(t)\dot{e}_2(t) + \dots + e_N(t)\dot{e}_N(t) \\ &= e_1(t)(\dot{\phi}_2(t) - \dot{\phi}_1(t)) + e_2(t)(\dot{\phi}_3(t) - \dot{\phi}_2(t)) \\ &\quad + \dots + e_N(t)(\dot{\phi}_1(t) - \dot{\phi}_N(t)) \end{aligned} \quad (19b)$$

which is Lyapunov stable. The desired angular speed is calculated for each consecutive three vehicles in the formation is given as follows:

$$\dot{\phi}_{Di}(t) = \frac{3\dot{\phi}_D + \gamma(\Delta\phi_{lead}(t) - \Delta\phi_{lag}(t))}{3} \quad i \in [1, N] \quad (20)$$

where $\Delta\phi_{lead}(t)$ and $\Delta\phi_{lag}(t)$ are the angular difference between the UAV with its leading and lagging UAVs, respectively and $\gamma \in [0, 1]$ to be a positive gain constant that controls the rate of convergence of the angle of separation.

IV. DECISION ALGORITHM

The switching algorithm proposed in this paper is based on a fuzzy logic controller that is derived based on online training using a Fuzzy Q-Learning (FQL) algorithm. In this algorithm, a fuzzy logic controller (FLC) is learned from experience and is then ported to real UAVs. The cooperative UAVs make their decision to switch from the formation tactic to the dynamic encirclement depending on the output of this decentralized (FLC) as this type of controllers (FLC) has the ability to manage ill-defined processes that are characterized with poor knowledge of their underlying dynamics [54].

The fuzzy controller is composed of four main elements [54]:

- **Fuzzy inputs:** A fuzzy description of the inputs to the system, in our case *target threat*, *target importance* and *separating distance*.
- **Fuzzy rule base:** It is a set of (If/Then rules) containing the experts description about how to achieve the desired control goal.
- **Fuzzifier:** It is the transformation of the non-fuzzy measured data into suitable linguistic values.
- **Inference engine:** It is the heart of the FLC, and it has the ability to simulate the human decision making to achieve the desired control goal.
- **Defuzzifier:** It is the transformation of the fuzzy control action into non-fuzzy control decision.

The decision to switch is learned from simulations of missions using three UAVs. This speeds up the learning as the Q-tables can be shared by the different vehicles. The three inputs are represented as follows.

First, the degree of threat of the target $\varepsilon \in [0, 1]$ is simulated. Certain targets are more dangerous than others due to natural reasons (wind, turbulence, etc) or man caused reasons (weapons, surveillance). In the simulation, the robots evaluate the degree of threat based on unsuccessful missions, but in the experiment the degree of threat is assumed to be known beforehand, for example, flying close to a structure may lead the UAV into an area of turbulence and this risk is known to the robot.

The second input is the importance of the target $\eta \in [0, 1]$. This is entered based on the type of the target by a specialist and is mission dependent. For example, let us assume that the mission is a rescue mission and the target is a person. If a person is found, its importance is high. However, if the found target is a vehicle and the missing person is known to be on foot, the target could be seen as less important.

Finally, the last input is the separating distances between the UAVs and the target d which varies from $[0,30]$. This parameter is related to the sensing radius of the vehicle. Therefore, vehicles have an incentive, whenever possible, to be at close proximity to each other.

The output $\Delta S \in \{0,1\}$ is used to determine which tactic the UAV will use, either the formation tactic or the dynamic encirclement tactic. This can also be thought of as the action of the controller. Fig. 3 represents the fuzzy logic controller scheme used by the UAV.

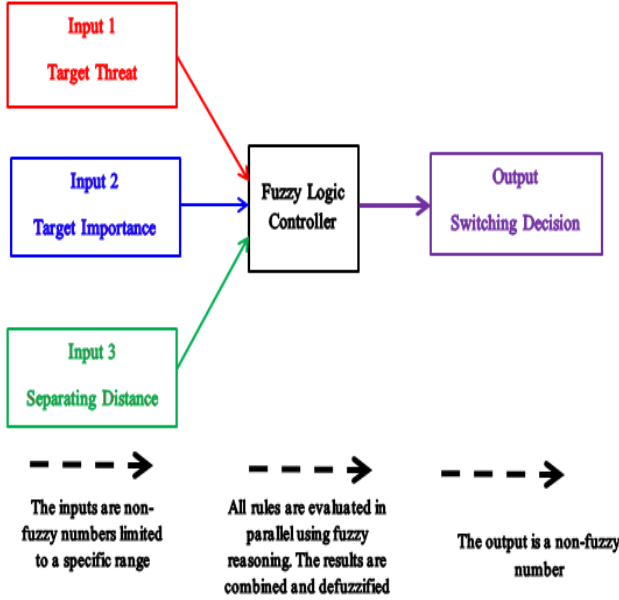


Fig. 3: Fuzzy logic controller scheme by the UAV to make the decision of switching.

A. The fuzzy logic controller

Fig. 4 represents the membership functions for the inputs of the controller.

The fuzzy sets for the threat level are the following:

- SS: $0 \leq \varepsilon \leq 0.35$ (small)
- MM: $0.35 \leq \varepsilon \leq 0.7$ (medium)
- LL: $0.7 \leq \varepsilon \leq 1$ (large)

For the importance of the target, the fuzzy sets are represented in the same way, ie:

- SS: $0 \leq \eta \leq 0.35$ (small)
- MM: $0.35 \leq \eta \leq 0.7$ (medium)
- LL: $0.7 \leq \eta \leq 1$ (large)

Finally, for the distance between UAVs, the fuzzy sets are:

- VS: d is near 0 (very small)
- SS: $0 \leq d \leq 10$ (small)
- MS: $10 \leq d \leq 15$ (medium small)
- MM: $15 \leq d \leq 20$ (medium)
- ML: $20 \leq d \leq 25$ (medium large)
- LL: $25 \leq d \leq 30$ (large)

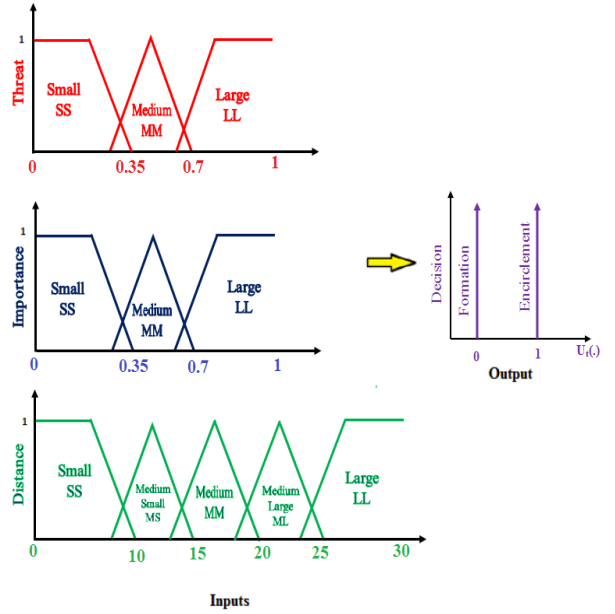


Fig. 4: The fuzzy membership functions for features inputs beside the fuzzy membership functions for feature output variable ΔS .

This generates $3 \times 3 \times 5 = 45$ rules for each UAV to decide in a decentralized manner which task to execute. The learning algorithm works by assigning an action to each of the rules based on the states of each one of the three vehicles. The learning algorithm will be explained in the next section.

B. Fuzzy Q-Learning (FQL)

Using the product inference for fuzzy implication, t-norm, singleton fuzzifier and center average defuzzifier [56][41], the output of the fuzzy system U_t for the current state \bar{x}_t becomes at time t ,

$$U_t(\bar{x}_t) = \text{round}(\sum_{l=1}^N \phi_t^l a_t^l) \quad (21)$$

$$\phi^l = \frac{\prod_{i=1}^n \mu_i^l(x_i)}{\sum_{l=1}^N (\prod_{k=1}^n \mu_k^l(x_k))} \quad (22)$$

where μ is the membership function, x_i is the i th input; such that $\bar{x} = [x_1, x_2, x_3] = [\varepsilon, \eta, d]$, n is the number of inputs ($n = 3$) and N is the number of rules ($N = 45$). The term a_t^l is the constant describing the centre of the fuzzy

set for each rule, which in our case is the action selected at time t for rule l based on the Q-table (will be described afterwards) ^[57]. Also, action a belongs to an action set A , which contains the possible tactics (formation = 0, encirclement = 1). The action set therefore is $A = \{0,1\}$, but notice that the method described in here would work for any number of actions. For the tactics switching as presented, if $\sum_{l=1}^N \phi_t^l a_t^l < 0.5$, the tactic chosen will be *formation* ($U_t(\bar{x}_t) = 0$) and if $\sum_{l=1}^N \phi_t^l a_t^l \geq 0.5$ the tactic chosen will be *encirclement* ($U_t(\bar{x}_t) = 1$).

After taking any action, the action is evaluated. Based on our evaluation a value is added to the Q-table. The Q-table contains a value for each action in the action set for each of the rules. For each rule, the action with a higher value is more favourable; however, a random selection factor was added to create exploration and exploitation. The values for each of the actions are updated and adapted based on the learning algorithm. The value for the current state is represented by capital Q , and the Q-table is represented by lower-case q . One can compute Q for the current state as,

$$Q(\bar{x}_t) = \sum_{l=1}^N \phi_t^l q_t(l, a_t^l) \quad (23)$$

The maximum possible value is represented as Q^* and computed as,

$$Q^*(\bar{x}_t) = \sum_{l=1}^N \phi_t^l \max_{a^l \in A} q_t(l, a_t^l) \quad (24)$$

Lastly, one computes the future temporal difference as ^[57],

$$E_{t+1} = r_{t+1} + \gamma Q^*(\bar{x}_{t+1}) - Q(\bar{x}_t) \quad (25)$$

where γ is the forgetting factor and it focuses on the expected future rewards and r_{t+1} is the reward received by the agent after doing an action. The key factor in deriving an efficient tactic switching approach is in designing a reward function that captures (quantifies) the success or failure of a mission. The design of the reward function will be discussed in Section IV-C.

After calculating the temporal difference, the learning agent is ready to adapt its Q-table for each of the fuzzy rules $l \in \text{fuzzy rules}$. The Q-table is adapted according to:

$$q_{t+1}(l, a) = q_t(l, a) + \alpha E_{t+1} \phi_t^l \quad (26)$$

where α is the learning rate.

Exploration and exploitation is done by selecting a random action from action set A with probability ϵ (exploration rate). We call the forgetting factor, learning rate, and exploration rate as Learning Factors. The FQL algorithm is shown in Algorithm 1. This FQL was implemented in ^[58].

Algorithm 1 FQL algorithm

- 1: $q(l, a) \leftarrow 0, \forall a \in A$ and $\forall l \in \text{rules}$
 - 2: **for** each time step t **do**
 - 3: For each rule, choose action a_t^l based on:
 - 4: $\left(\begin{array}{l} a_t^l = \underset{a \in A}{\text{argmax}} q^l(l, a), \text{ with probability } (1 - \epsilon) \\ \text{random action from } A, \text{ with probability } \epsilon \end{array} \right.$
 - 5: Calculate ϕ^l for each rule based on (22)
 - 6: Estimate current state value $Q(\bar{x}_t)$ using (23)
 - 7: Calculate system output for current state $U_t(\bar{x}_t)$ using (21)
 - 8: Take action U_t
 - 9: Observe new state
 - 10: Obtain reward (either using (27) or (28))
 - 11: Calculate maximum possible future value $Q^*(\bar{x}_{t+1})$ based on (24)
 - 12: Calculate temporal difference E_{t+1} using (25)
 - 13: Adapt the Q-table using (26)
 - 14: **end for**
-

C. Reward Design

The reward function is calculated depending on how well the UAV executes its mission and, therefore, is divided in two possibilities: a formation mission or an encirclement mission. At the beginning of each training epoch, all UAVs are set to the formation mission. Moreover, when the UAVs receive their rewards, they all update the same Q-table. In this way, training is sped up and experience is shared among all UAVs. Notice that exploration or exploitation is decided at the UAV level and does not impact learning. Also, the rewards are normalized and $r + t + 1 \in [0,1]$. In case the chosen tactic is formation, the reward is

$$r_{t+1}^f = \max\left(1 - \frac{\max_{j \in \mathcal{N}} (\|d_{sep} - d_{ij}\|)}{d_{sep}}, 0\right) \quad (27)$$

where d_{sep} is the desired distance of separation between the i^{th} UAV and their neighbors (in our case only two) in set \mathcal{N} . If the chosen tactic is encirclement, the reward is

$$r_{t+1}^e = \max\left(1 - \frac{\|r_e - r_{i\|}}{r_e} - \max_{j \in \mathcal{N}} (\|\phi_i - \phi_j\|) \frac{N_e}{2\pi}, 0\right) \quad (28)$$

where r_e is the desired encirclement radius, N_e is the number of vehicles encircling the target and \mathcal{N} is the set of neighboring UAVs (in our case, mostly two, one leading and the other lagging). The function in (28) measures if the UAV is able to encircle the target and if the encirclement with other vehicles is stable.

If the UAV is damaged (because it got too close to a dangerous target, for example), it receives an automatic reward of 0 for the remainder of the learning epoch. This is done to discourage the UAV of getting in dangerous situations.

The reward in either (27) or (28) is used by Algorithm 1 for updating the Q-table. As discussed in

section V, Algorithm 1 is used to train the UAVs until a satisfactory behavior is achieved.

V. SIMULATION RESULTS

The control strategy discussed in sections II and III in addition to the decision algorithm introduced in section IV was successfully implemented in simulation on a multiple cooperative UAV team consisting of three vehicles.

The objective of our simulations is to show that the designed LMPC policy is fit for the formation and dynamic encirclement tactics. Also, the success of each UAV to decide independently to switch from one tactic to another using the proposed fuzzy logic controller. Our requirements during the simulation are divided into two sets.

In the formation phase, the requirements are a desired separation distance of 1.7 m between the neighboring UAVs and a desired velocity 0.2 m/sec for the whole formation. In the encirclement phase, our requirements are a radius of encirclement r of 1m, angular velocity $\dot{\phi}$ of 0.15 rad/sec and an angular separation of $\frac{2\pi}{N_e}$ rad, where N_e is the number of UAVs that decided to switch to the dynamic encirclement tactic.

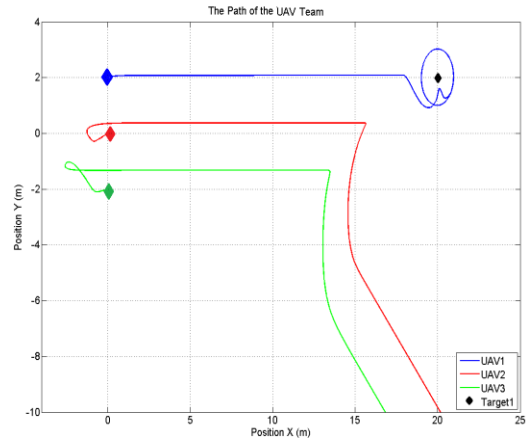
During the simulation, the UAV team flies in a desired formation until their sensing radii indicate the presence of a target ($r_s = 30m$). The UAV which meets the target first, checks its radius of communication ($r_c = 50m$) and sends a package of information including the location of the target, the importance of the target η and the estimation of the degree of threat of the target ε to the other members of the team in its zone of communication.

The Q-table is initialized with zeros and one thousand missions are executed with the UAVs starting in random positions (although in formation, i.e., with same orientation and equally spaced) until learning takes place. For each training epoch, at each time instant t each UAV decides to used either the formation tactic ($U_i(\cdot)=0$) or the encirclement tactic ($U_i(\cdot)=1$).

Fig. 5 shows a sample simulation after learning took place. The UAV team for this scenario consists of three cooperative UAVs with initial positions at (0,2),(0,0) and (0,-2). During the formation phase, the cooperative UAVs successfully form the desired configuration converging to the desired requirements of separation distance of 1.7 m between each other and desired velocity 0.2 m/sec. While no target has been sensed, the UAVs continue with the original tactic (formation). When UAV1 senses the presence of a target located at (20,2), it sends a package of information to the other UAVs within its communication radius. In this example, UAV1 takes the decision to switch to the encirclement tactic while the other two members in the team take their decisions to

continue in the formation tactic according to the output of their fuzzy logic controller derived by Algorithm 1.

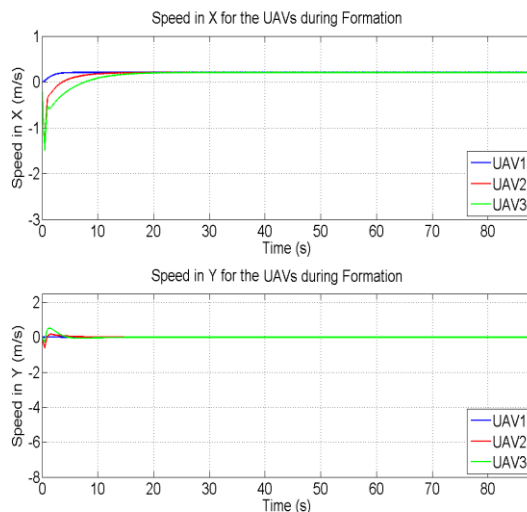
The simulation ran for 240 sec with prediction horizon $M=8$, $r_{max}=5m$, $r_{min}=0m$, and $\dot{\phi}_{min} = -\dot{\phi}_{max} = -0.5 \text{ rad/sec}$. Fig. 5 also shows that the UAVs converge to the desired requirements during the formation phase and



UAV1 converges to the desired encirclement requirements.

Fig. 5: UAV team members switching from formation flight to dynamic encirclement tactic while maintaining the required separating distance during the formation phase and the desired radius and angular velocity during the encirclement phase. UAV 1, 2 and 3 are represented by the blue, red and green diamonds, respectively, while the target is the black diamond.

The speed of each UAV in the fleet during the formation phase is shown in Fig. 6, where all the cooperative UAVs match their flockmates' speed of 0.2m/sec. Moreover, Fig. 7 shows the radius of encirclement and angular velocity for UAV1 during the encirclement phase. The cost output of the MPC controller



for UAV1 during the encirclement phase is presented in Fig. 8 .

Fig.6: The speed of the UAV team in during the formation phase.

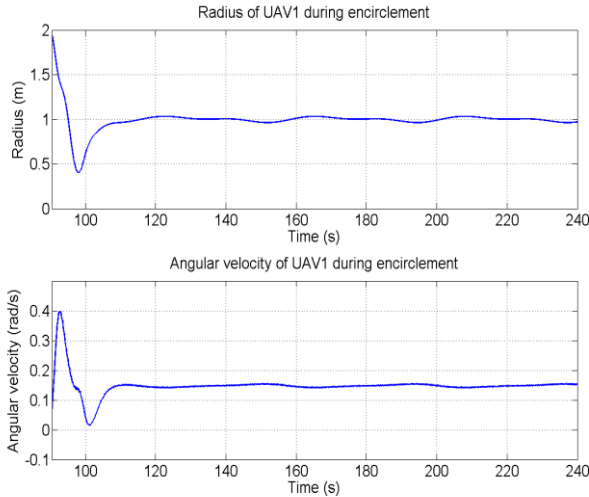


Fig.7: The radius of encirclement and angular velocity for UAV1 encircling a stationary target at (20,2).

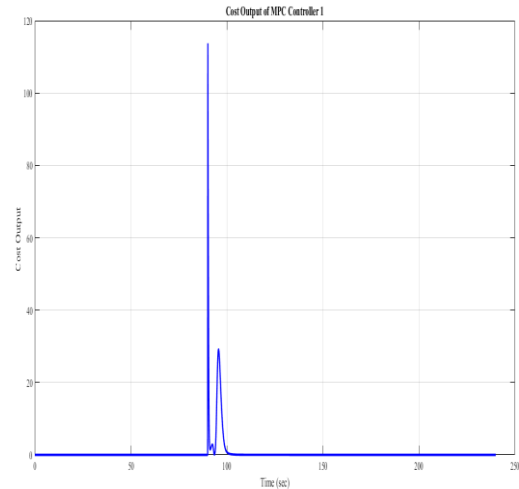
VI. EXPERIMENTAL RESULTS

In order to show that the UAVs may accomplish tactic switching in real-time, the control strategies and the fuzzy Q-learning tactic switching are implemented on actual Qball-X4 quadrotors. In this section, we will describe our experimental apparatus, particularly the quadrotor vehicle used and our laboratory setup including the method of sensing the position of the quadrotors during flight.

The main element of our system is the Quanser Qball-X4, shown in Fig.9. It is an innovative rotary wing vehicle platform suitable for a wide variety of UAV research applications. The Qball-X4 is a quadrotor helicopter design, propelled by four motors fitted with 10-inch propellers. The entire quadrotor is enclosed within a protective carbon fiber cage that ensures safe operation as well as opens the possibilities for a variety of novel applications. The electronic speed controllers receive commands from the high Qball-X4’s data acquisition board (HiQ DAQ) which operates together with the target computer (Gumstix) to control the vehicle by reading on-board sensors and outputting the required rotor commands [59].

The laboratory setup for the experimental work includes three quadrotors, a ground station and an Opti-track system of 24 cameras responsible for measuring the position of the quadrotors during flight and covering an area of $5 \times 3 \times 2.8$ m. All the components are communicating through an internal WiFi network as shown in Fig. 10.

Due to the small area of our laboratory arena, the initial positions of the three UAVs and the target are different from these of the simulation case. Also, our requirements during the real-time implementation are a divided into two sets. In the formation phase, our requirements are a desired **Fig. 8:** The Cost Output for the MPC controller for UAV1 during



the encirclement phase.

separation distance of 1.7 m between the N UAVs and a desired velocity 0.2 m/sec. In the encirclement phase, our requirements are a radius of encirclement of 1m, angular velocity of 0.15 rad and an angular separation of $2\pi/N_c$ rad.



Fig. 9: The quadrotor “Quanser Qball-X4” [59].

The experiment ran with prediction horizon $M=8$, $r_{max}=5m$, $r_{min}=0m$, and $\dot{\phi}_{min} = -\dot{\phi}_{max} = -0.5rad/sec$. The three quadrotors hover in their initial positions for 20 sec, then start to move in the desired formation in the -Y direction respecting the formation tactic requirements. After reaching the edge of the experimental arena, they

change their direction of flight to the +Y direction in order to cover more area and thus show the convergence over a longer period. UAV1 senses the presence of a target located at the origin, sending a package of information to the other members of the fleet within its radius of communication ($r_c=5m$). UAV1 makes the decision to encircle while the other two members decide to hover away to clear the area for encirclement. UAV1 converges to the desired encirclement requirements.

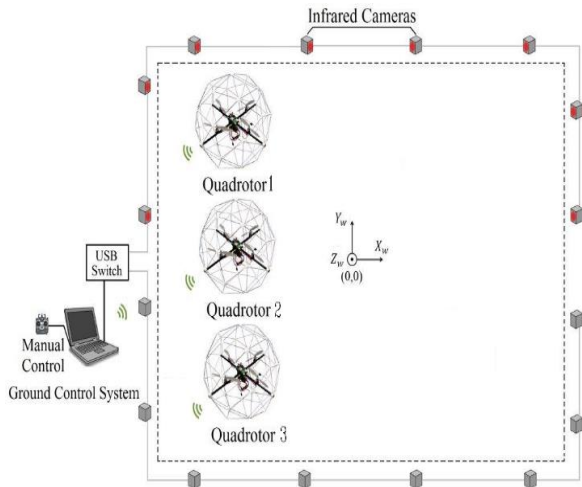


Fig. 10: The laboratory setup.

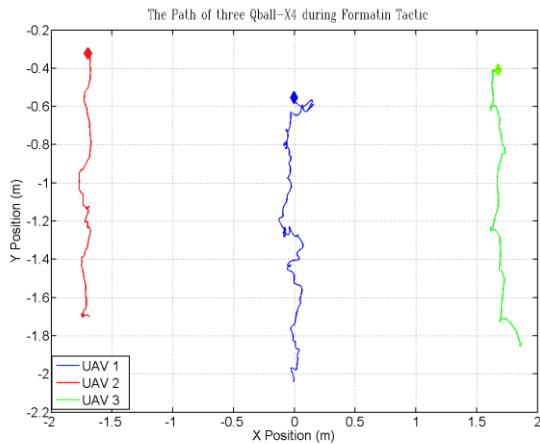


Fig.11: The team of three Qball-X4 flying in a desired formation in the $-Y$ direction.

Fig. 11 represents the path of the three Qball-X4 flying direction during the formation phase, where the UAVs converge to the desired separation distance 1.7 m between each other. Fig. 12 shows the separation distance between the leading UAV and its neighbors in the X direction, while Fig. 13 shows separating distance between the leading UAV and its neighbors in Y direction. The radius of encirclement and the angular speed for UAV1 are

shown in Fig. 14, where all the results converge to the desired radius of 1m and desired angular speed of 0.15rad/sec. Snapshots of our experimental flight tests are shown in Fig. 15.

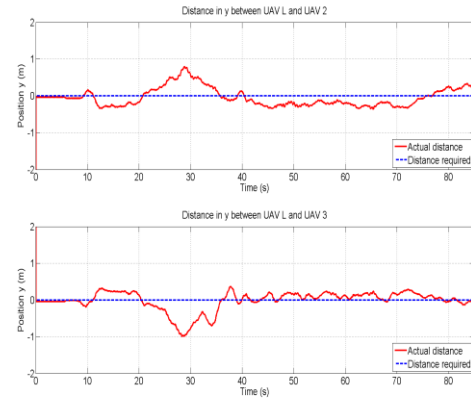


Fig. 12: The separating distance between the leading UAV and the other members in X-direction

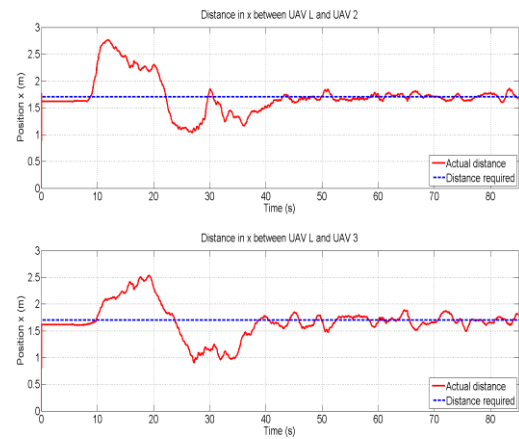


Fig. 13: The separating distance between the leading UAV and the other members in Y-direction

VII. DISCUSSION OF RESULTS

By comparing our simulation results represented in section V with the experimental results represented in section VI, we can notice that the designed controllers, LMPC and the combination of FL and LMPC, were able to solve the problem of tactic switching for the cooperative team successfully. Also, the fuzzy logic controller presented in section IV succeeded in generating the switching decision for each UAV in a decentralized manner. All vehicles converge to the desired requirements during the formation phase while UAV1 converges to the desired radius of encirclement and angular speed during the encirclement phase. The simulation results are more stable compared with the experimental results due to the

wind disturbance effects produced by the rotors of each quadrotor flying at such proximity in the enclosed lab environment. However, the system is still robust to the disturbances and converges to the desired behavior as it may be seen in Figs. 12- 14.

VIII. CONCLUSION

In this paper, we show that the proposed LMPC policy can be successfully applied to solve the problem of formation flight combined with dynamic encirclement of a stationary target using a group of cooperative UAVs in simulation and real-time implementation. A fuzzy logic controller was used to successfully control the switching decision for each member in the fleet. Our simulations and experimental results show that our control policy succeeded to control the UAV system by converging to the desired requirements for formation and encirclement. This control policy is characterized by robustness, scalability, and the real-time implementation.

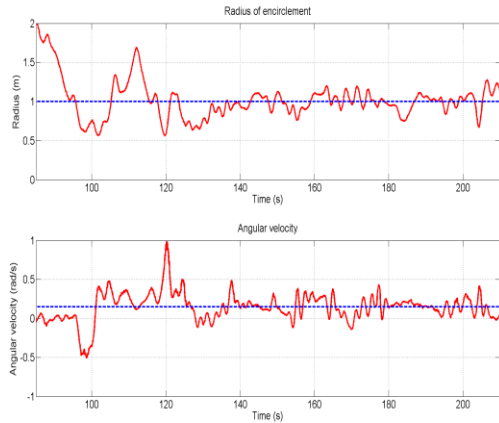


Fig. 14: The radius of encirclement and the angular velocity for UAV1 encircling a stationary target at the origin.

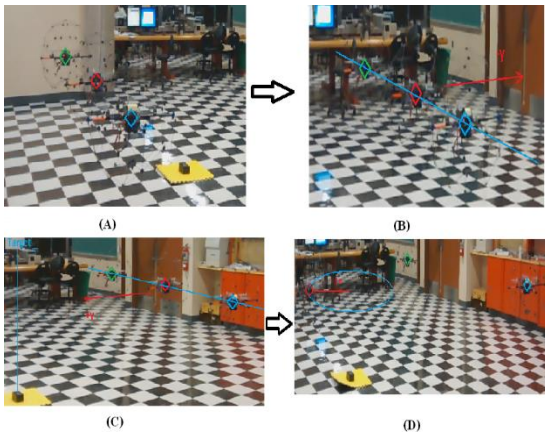


Fig. 15: (A) Three UAVs hover for 20sec. (B) Three UAVs in line abreast formation moving towards the (-Y) direction. (C) Target appears and UAV1 (red diamond) senses the target. (D) UAV1 decides to encircle while UAV2 (green diamond) and UAV3 (blue diamond) clear the area.

In the future, we will apply the designed controllers derived in this paper combined with a learning algorithm to solve the tactic switching problem for multiple UAVs aiming to reduce the imperfections between the linearized system and the actual model. We see this as a necessary step to improve the stability, robustness, convergence and safety of autonomous applications.

IX. REFERENCES

- [1] E. Bone and C. Bolkom, "Unmanned aerial vehicles: Background and issues for congress." DTIC Document,2003.
- [2] G. Cai, K. Lum, B. Chen, and T. Lee, "A brief overview on miniature fixed-wing unmanned aerial vehicles," in 8th IEEE International Conference on Control and Automation (ICCA), 2010, pp. 285–290.
- [3] A. I. Alshbatat and Q. Alsafasfeh, "Cooperative decision making using a collection of autonomous quadrotor unmanned aerial vehicles interconnected by a wireless communication network," Global Journal on Technology, vol. 1, 2012.
- [4] A.J.Marasco, "Control of Cooperative and Collaborative Team Tactics in Autonomous Unmanned Aerial Vehicles Using Decentralized Model Predictive Control," Master's thesis, Royal Military College - Canada, 2012.
- [5] A. Ryan, M. Zennaro, A. H., R. S., and J. Hedrick, "An overview of emerging results in cooperative UAV control," in 43rd IEEE Conference on Decision and Control (CDC),vol.1.IEEE, 2004, pp.602–607.
- [6] K. Do and J. Pan, "Nonlinear formation control of unicycle-type mobile robots," Robotics and Autonomous Systems, vol. 55, no. 3, pp. 191–204, 2007.
- [7] D. Scharf, F. Hadaegh, and S. Ploen, "A survey of spacecraft formation flying guidance and control. part ii: control," in American Control Conference, 2004. Proceedings of the 2004, vol. 4. IEEE, 2004, pp. 2976–2985.
- [8] Y. Ding, C. Wei, and S. Bao, "Decentralized formation control for multiple UAVs based on leader-following consensus with time-varying delays," in Chinese Automation Congress (CAC), 2013. IEEE, 2013, pp. 426–431.
- [9] W. Ren and R. Beard, "A decentralized scheme for spacecraft formation flying via the virtual structure approach," in Proceedings of the 2003 American Control Conference,vol.2.IEEE, 2003, pp.1746–1751.
- [10] J. Lawton, R. Beard, and B. Young, "A decentralized approach to formation maneuvers," IEEE Transactions on Robotics and Automation, vol. 19, no. 6, pp. 933–941, 2003.
- [11] X. Dong, Y. Zhou, Z. Ren, and Y. Zhong, "Time-varying formation control for unmanned aerial vehicles with switching interaction topologies," Control Engineering Practice, vol. 46, pp. 26–36, 2016.
- [12] S. Bhattacharya and T. Basar, "Game-theoretic analysis of an aerial jamming attack on a UAV communication network," in American Control Conference (ACC), 2010. IEEE, 2010, pp. 818–823.
- [13] H. Rezaee, F. Abdollahi, and M. B. Menhaj, "Model free fuzzy leader-follower formation control of fixed wing UAVs," in 2013 13th Iranian Conference on Fuzzy Systems (IFSC). IEEE, 2013, pp. 1–5.
- [14] D. Luo, T. Zhou, and S. Wu, "Obstacle avoidance and formation regrouping strategy and control for UAV formation flight," in 10th IEEE International Conference on Control and Automation (ICCA), 2013. IEEE, 2013, pp. 1921–1926.
- [15] A. S. Brandao, J. Barbosa, V. Mendoza, M. SarcinelliFilho, and R. Carelli, "A multi-layer control scheme for a centralized UAV formation," in 2014 International Conference on Unmanned Aircraft Systems (ICUAS),.
- [16] G. Regula and B. Lantos, "Formation control of a large group of UAVs with safe path planning," in 21st Mediterranean Conference on Control & Automation (MED). IEEE, 2013, pp. 987–993.
- [17] Y. Kuriki and T. Namerikawa, "Consensus-based cooperative formation control with collision avoidance for a multi-UAV system," in American Control Conference (ACC), 2014. IEEE, 2014, pp. 2077–2082.
- [18] P. Yao, H. Wang, and Z. Su, "Real-time path planning of unmanned aerial vehicle for target tracking and obstacle avoidance in complex



- dynamic environment,” *Aerospace Science and Technology*, vol. 47, pp. 269–279, 2015.
- [19] X. Chen and C. Zhang, “The method of multi unmanned aerial vehicle cooperative tracking in formation based on the theory of consensus,” in *5th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2. IEEE, 2013, pp. 148–151.
- [20] R. Sharma, M. Kothari, C. Taylor, and I. Postlethwaite, “Cooperative target-capturing with inaccurate target information,” in *American Control Conference (ACC)*. IEEE, 2010, pp. 5520–5525.
- [21] A. Sudsang and J. Ponce, “On grasping and manipulating polygonal objects with disc-shaped robots in the plane,” in *98 IEEE International Conference on Robotics and Automation*, vol. 3. IEEE, 98, pp. 2740–2746.
- [22] S. Hara, T. Kim, Y. Hori, and K. Tae, “Distributed formation control for target-enclosing operations based on a cyclic pursuit strategy,” in *Proceedings of the 17th IFAC World Congress*, 2008, pp. 6602–6607.
- [23] Y. Lan, G. Yan, and Z. Lin, “Distributed control of cooperative target enclosing based on reachability and invariance analysis,” *Systems & Control Letters*, vol. 59, no. 7, pp. 381–389, 2010.
- [24] H. Yamaguchi, “A cooperative hunting behavior by mobile-robot troops,” *The International Journal of Robotics Research*, vol. 18, no. 9, pp. 931–940, 1999.
- [25] H. Kawakami and T. Namerikawa, “Cooperative target capturing strategy for multi-vehicle systems with dynamic network topology,” in *American Control Conference (ACC)*. IEEE, 2009, pp. 635–640.
- [26] T. Kim, S. Hara, and Y. Hori, “Cooperative control of multi-agent dynamical systems in target-enclosing operations using cyclic pursuit strategy,” *International Journal of Control*, vol. 83, no. 10, pp. 2040–2052, 2010.
- [27] M. Sadraey, “Multi-vehicle circular formation flight in an unknown time-varying flow-field,” in *2013 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2013, pp. 859–868.
- [28] A. J. Marasco, S. N. Givigi, C. A. Rabbath, and A. Beaulieu, “Dynamic encirclement of a moving target using decentralized nonlinear model predictive control,” in *American Control Conference (ACC)*. IEEE, 2013, pp. 3960–3966.
- [29] A. T. Hafez, A. J. Marasco, S. N. Givigi, M. Iskandarani, S. Yousefi, and C. A. Rabbath, “Solving multi-uav dynamic encirclement via model predictive control,” *IEEE Transactions on Control Systems Technology*, vol. 23, no. 6, pp. 2251–2265, 2015.
- [30] A. T. HAFEZ, “Design and implementation of modern control algorithms for unmanned aerial vehicles,” Ph.D. dissertation, Queens University, 2014.
- [31] E. F. Camacho and C. Bordons, *Model Predictive Control*. London: Springer-Verlag, 2007.
- [32] R. Negenborn, B. De Schutter, M. Wiering, and H. Hellendoorn, “Learning based model predictive control for markov decision processes,” in *Proceedings of the 16th IFAC world congress*, 2005, pp. 1–7.
- [33] X. Wang, V. Yadav, and S. Balakrishnan, “Cooperative UAV formation flying with obstacle/collision avoidance,” *IEEE Transactions on Control Systems Technology*, vol. 15, no. 4, pp. 672–679, 2007.
- [34] J. Lavaei, A. Momeni, and A. G. Aghdam, “A model predictive decentralized control scheme with reduced communication requirement for spacecraft formation,” *IEEE Transactions on Control Systems Technology*, vol. 16, no. 2, pp. 268–278, 2008.
- [35] Z. Chao, S.-L. Zhou, L. Ming, and W.-G. Zhang, “UAV formation flight based on nonlinear model predictive control,” *Mathematical Problems in Engineering*, vol. 2012, 2012.
- [36] H.-S. Shin and M.-J. Thak, “Nonlinear model predictive control for multiple UAVs formation using passive sensing,” *International Journal of Aeronautical and Space Science*, vol. 12, no. 1, pp. 16–23, 2011.
- [37] A. Bemporad and C. Rocchi, “Decentralized linear time varying model predictive control of a formation of unmanned aerial vehicles,” in *50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011. IEEE, 2011, pp. 7488–7493.
- [38] F. Allgöwer and A. Zheng, *Nonlinear model predictive control*. Birkhäuser, 2000, vol. 26.
- [39] H. Shin, M. Thak, and H. Kim, “Nonlinear model predictive control for multiple UAVs formation using passive sensing,” *International Journal of Aeronautical and Space Sciences*, vol. 12, pp. 16–23, 2011.
- [40] M. Iskandarani, A. Hafez, S. Givigi, A. Beaulieu, and A. Rabbath, “Using multiple quadrotor aircraft and linear model predictive control for the encirclement of a target,” in *IEEE International System Conference (SYSCON)*, 2013, pp. 1–7.
- [41] A. J. Marasco, S. N. Givigi, and C. A. Rabbath, “Model predictive control for the dynamic encirclement of a target,” in *American Control Conference (ACC)*. IEEE, 2012, pp. 2004–2009.
- [42] A. Aswani, P. Bouffard, and C. Tomlin, “Extensions of learning-based model predictive control for real-time application to a quadrotor helicopter,” in *American Control Conference (ACC)*, 2012. IEEE, 2012, pp. 4661–4666.
- [43] A. T. Hafez and S. Givigi, “Formation reconfiguration of cooperative UAVs via learning based model predictive control in an obstacle-loaded environment,” in *Annual IEEE Systems Conference (SysCon)*. IEEE, 2016, pp. 1–8.
- [44] A. T. Hafez, S. N. Givigi, K. A. Ghamry, and S. Yousefi, “Multiple cooperative uavs target tracking using learning based model predictive control,” in *International Conference on Unmanned Aircraft Systems (ICUAS)*, 2015. IEEE, 2015, pp. 1017–1024.
- [45] P. Marinos, “Fuzzy logic and its application to switching systems,” *IEEE Transactions on Computers*, vol. C-18, no. 4, pp. 343–348, April 1969.
- [46] Y.-C. Chiou and L. Lan, “Genetic fuzzy logic controllers,” in *Intelligent Transportation Systems*, 2002. *Proceedings on The IEEE 5th International Conference*, 2002, pp. 200–205.
- [47] H. Raslan, H. Schwartz, and S. Givigi, “A learning invader for the guarding a territory game,” in *2016 Annual IEEE Systems Conference (SysCon)*. IEEE, 2016, pp. 1–8.
- [48] S. N. Givigi, H. M. Schwartz, and X. Lu, “A reinforcement learning adaptive fuzzy controller for differential games,” *Journal of Intelligent and Robotic Systems*, vol. 59, no. 1, pp. 3–30, 2010.
- [49] D. Lee, H. Jin Kim, and S. Sastry, “Feedback linearization vs. adaptive sliding mode control for a quadrotor helicopter,” *International Journal of Control, Automation and Systems*, vol. 7, no. 3, pp. 419–428, 2009.
- [50] A. Mokhtari and A. Benallegue, “Dynamic feedback controller of Euler angles and wind parameters estimation for a quadrotor unmanned aerial vehicle,” in *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 3. IEEE, 2004, pp. 2359–2366.
- [51] Z. Fang, Z. Zhi, L. Jun, and W. Jian, “Feedback linearization and continuous sliding mode control for a quadrotor UAV,” in *27th Chinese Control Conference*, 2008. CCC. IEEE, 2008, pp. 349–353.
- [52] A. Hafez, M. Iskandarani, A. Marasco, S. Givigi, Y. Shahram, and C. Rabbath, “Solving multi-UAV dynamic encirclement via model predictive control,” *IEEE Transactions on Control Systems Technology*, vol. 23, no. 6, pp. 2251–2265, 2015.
- [53] A. Hafez, M. Iskandarani, S. Givigi, S. Yousefi, C. A. Rabbath, and A. Beaulieu, “Using linear model predictive control via feedback linearization for dynamic encirclement,” in *American Control Conference (ACC)*, 2014. IEEE, 2014, pp. 3868–3873.
- [54] K. M. Passino, S. Yurkovich, and M. Reinfrank, *Fuzzy control*. Citeseer, 1998, vol. 42.
- [55] L. Wang, *A Course in Fuzzy Systems and Control*. Prentice Hall PTR, 1997.
- [56] H. T. Nguyen and E. Walker, *A first course in fuzzy logic*. Boca Raton, FL: Chapman and Hall, 2006. [Online]. Available: www.summon.com
- [57] H. Schwartz, *Multi-Agent Machine Learning: A Reinforcement Approach*. Wiley, 2014.
- [58] M. J. Er and L. San, “Automatic generation of fuzzy inference systems using incremental-topological-preserving-map-based fuzzy q-learning,” in *IEEE International Conference on Fuzzy Systems*, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). June 2008, pp. 467–474.
- [59] Quanser, “Qball-x4 user manual,” Technical Report 829.