Joint Power Control, Pilot Assignment, User Association and Flight Control for Massive MIMO Self-Organizing Drones using Reinforcement Learning

Presenter:
Gabe Skidmore

Affiliation:
Miami University

Email:
skidmogm@miamioh.edu

About the Author:

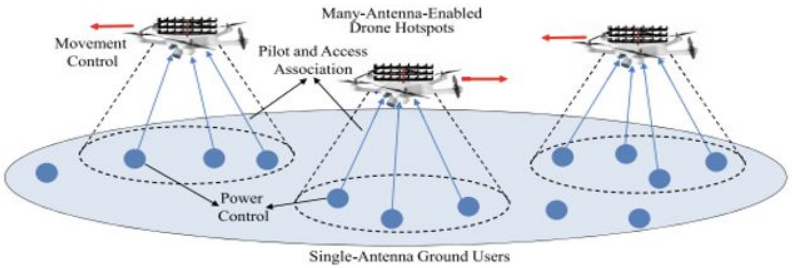
Gabriel Skidmore is currency pursuing his masters degree in the electrical and computer engineering department from Miami University, Oxford, Ohio, USA, and is expected in graduate over the summer of 2022. He is currently a graduate lab assistant for the electrical computer engineering department where he helps grade assignments and assist professors during labs.

Objective

- Establish a Mixed-Integer Nonlinear Programming (MINLP) formulation for the massive Multiple-Input Multiple-Output (MIMO) joint:
      - Transmit power assignment
      - Pilot assignment
      - user association
  - Using a combination of:
      - Convex relaxation
      - Deep reinforcement learning
- Maximize spectral efficiency for ground users using deep reinforcement learning
- Perform a comparison to a convex relaxed global solution

Why It Is Important?

- There is an ever increasing demand for faster wireless communication networks with higher spectral efficiency:
    - Ultra Reliable communication
    - Internet of Things
    - Intelligent Transportation
    - Natural Disasters



- Unmanned Aerial Vehicles (UAVs) is a new alternative approach to provide ground connectivity to multiple users in an area:
    - Low cost
    - Mobile

Why It Is Important?

- Reinforcement learning:
  - Adaptable
    - Can achieve good results when the configuration on the environment changes
  - Bias Resistant
    - Learns from environment instead of labeled data

Introduction

- Order of Introduction:
  - MINLP (Mixed-Integer Nonlinear Program)
  - Massive MIMO (Multiple Input Multiple Output)
  - Reinforcement Learning
  - Deep Reinforcement Learning
    - Deep Q-Learning
  - Limitations of the Literature Review

MINLP (Mixed-Integer Nonlinear Program)

- Optimization Variables:
  - Association matrix
  - Pilot assignment matrix
  - Power control matrix
  - UAV location matrix
- Constraints:
  - Connectivity
  - Power control
  - Pilot assignment
  - Flight control
- Performance: Sum Spectral Efficiency (bits/s/Hz)

$$\text{Given}: \ \mathcal{A}, \ \mathcal{G}, G_{\max}, M, \tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}$$

$$\underset{\boldsymbol{\alpha},\boldsymbol{\mu},\mathbf{p},\mathbf{x},\mathbf{y},\mathbf{z}}{\text{Maximize}}: \ U \triangleq \sum_{g\in\mathcal{G}} C_g(\boldsymbol{\alpha}, \boldsymbol{\mu}, \mathbf{p}, \mathbf{x}, \mathbf{y}, \mathbf{z})$$

$$\begin{aligned} \text{Subject to}: \ & 0 \le p_g \le p_{\max}, \quad \forall g \in \mathcal{G}, \\ & x_{\min} \le x_a \le x_{\max}, \ \forall a \in \mathcal{A}, \\ & y_{\min} \le y_a \le y_{\max}, \ \forall a \in \mathcal{A}, \\ & z_{\min} \le z_a \le z_{\max}, \ \forall a \in \mathcal{A}, \\ & \mathit{Constraints}\ (1),\ (2),\ (3),\ (4),\ (5) \end{aligned}$$

[1] Guan et al.

$$\alpha_{ga} \in \{0,\ 1\}, \quad \forall g \in \mathcal{G},\ a \in \mathcal{A} \tag{1}$$

$$\mu_{gw} \in \{0,\ 1\}, \quad \forall g \in \mathcal{G},\ w \in \mathcal{W} \tag{2}$$

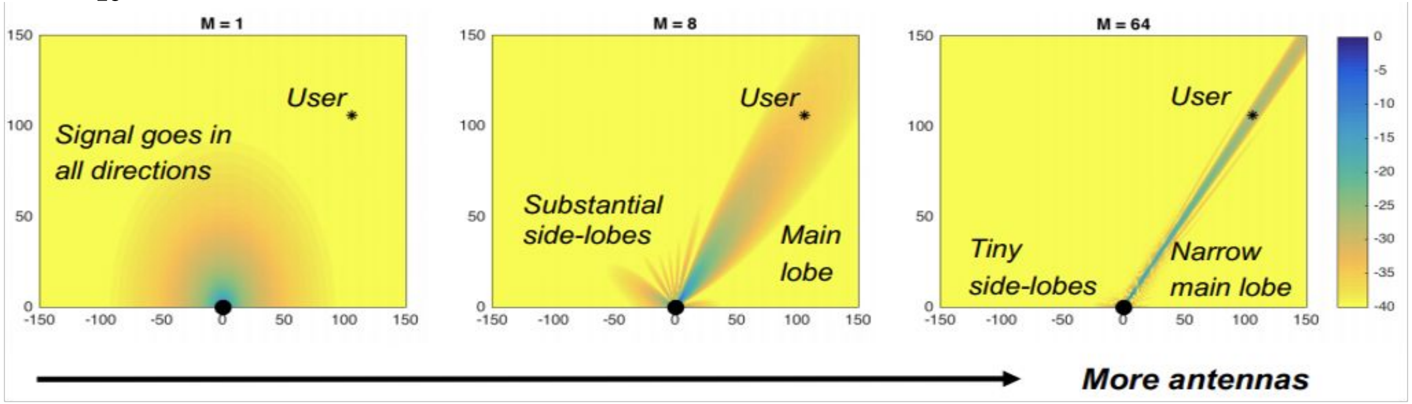
$$\sum_{a\in\mathcal{A}} \alpha_{ga} \le 1, \quad \forall g \in \mathcal{G}, \tag{3}$$

$$\sum_{g\in\mathcal{G}} \alpha_{ga} \le G_{\max}, \quad \forall a \in \mathcal{A}, \tag{4}$$

$$\sum_{w\in\mathcal{W}} \mu_{gw} \le 1, \quad \forall g \in \mathcal{G}, \tag{5}$$

Massive MIMO

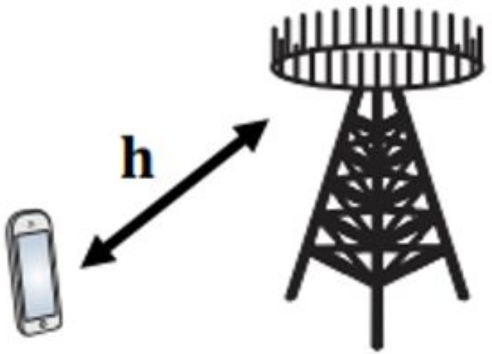
- Uses beamforming with spatial multiplexing to send signals to specific users
- Increases the number of antennas while keeping the power the same:
  - Narrower Beam
  - Main lobe focuses on the user
  - Lower leakage in directions away from the user
  - $10\log_{10}(M)$ dB larger array gain at the user



Massive MIMO

- Network Throughput Formula [bit/s/km^2]:

$$\underbrace{\text{Throughput}}_{\text{bit/s/km}^2} = \underbrace{\text{Cell density}}_{\text{Cell/km}^2} \cdot \underbrace{\text{Available spectrum}}_{\text{Hz}} \cdot \underbrace{\text{Spectral efficiency}}_{\text{bit/s/Hz/Cell}}$$

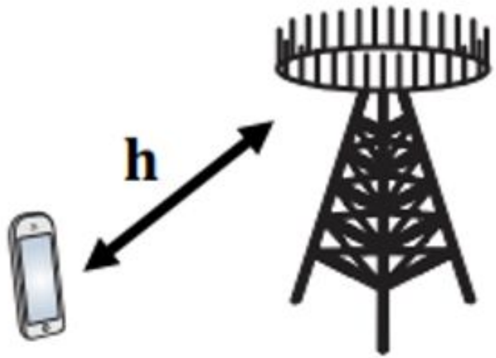


Massive MIMO

- Network Throughput Formula [bit/s/km^2]:

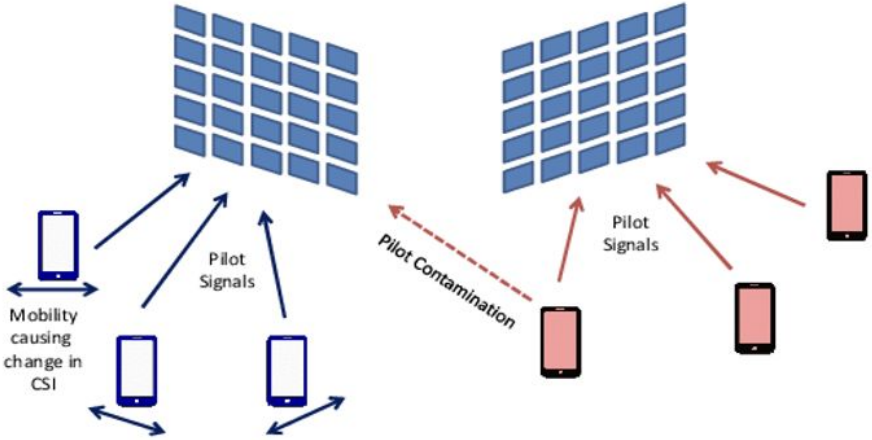
$$\underbrace{\text{Throughput}}_{\text{bit/s/km}^2} = \underbrace{\text{Cell density}}_{\text{Cell/km}^2} \cdot \underbrace{\text{Available spectrum}}_{\text{Hz}} \cdot \underbrace{\text{Spectral efficiency}}_{\text{bit/s/Hz/Cell}}$$

- Main Characteristics
  - Significantly more antennas than users
  - High spectral efficiency
  - Directive signals



Massive MIMO

- Massive MIMO mostly operates using Time Division Duplex (TDD)
- Ground users send out pilot signals to base stations
- A base station estimates a channel based on a pilot signal



Massive MIMO with UAVs

- Given a 2 GHz frequency band, a 100 dual-polarized antenna array only requires 0.75 x 0.75 meters of space.

- Industry:
  - Ericson recently launched the AIR 3268
    - 12 kg
    - 128 radiating elements (32 T/R branches)
    - 23 liters

Limitations in Literature Review

- [1] applied massive MIMO to aerial base stations but they applied a pricing algorithm that achieves 90% of their global optimum
- [2] used a search and sweep algorithm to locate many user clusters
- [3] focused only on the 3D location of UAVs to maximize spectral efficiency
- [4] and [5] focused mostly on the drone-to-base station backhaul connection
- [6] used a Resiliency Aware Deployment (RAD) algorithm to improve the network during transition mode

Limitations in Literature Review

- Reinforcement Learning
  - [7] found the best way to allocate resources in a distributed environment when the channel state information is not known
  - [8] used reinforcement learning to control transmit powers to mitigate interference
  - [17] uses reinforcement learning to analyze radio frequency channels to learn from past occupancy and conditions of the channels.
- Gaps
  - While controlling user transmit powers, pilot assignments and UAV positions has been done using convex relaxation, it only achieves 90% of their optimum solution
  - Other reinforcement learning algorithms for UAV base stations did not use massive MIMO which led them to experience low spectral efficiencies

Q-Learning: Q-Value

- In Q-Learning, an agent updates its q-value when it takes an action in the environment
- Updated in q-table where
  - Columns are states
  - Rows are actions
- Equation for updating q-value

$$q_*^{new}(s,a) = (1 - \alpha)q(s,a) + \alpha\left(R_{t+1} + \gamma max\, q\left(s^{'}, a^{'}\right)\right)$$

- Deep Q-Learning is different because it has a neural network instead of a q-table
- Neural Network gets updated based on the output of the neural network
- Updates q-values with backpropagation

Q-Learning: Parameters

- Number of States Determined by:
  - Number of users
  - Number of pilot sequences available
  - Number of power levels available
  - Size of the grid
- Number of Actions Determined by:
  - 4 movement directions of the agent
  - Number of users
  - Number of pilot sequences available
  - Number of power levels available

Q-Learning: Relaxation

- UAVs can assign multiple pilot sequences to users
- Problem broken down into two sub problems:

1. Users are connected to the UAV that has the highest Single-Input Single-Output (SISO) Signal to Noise Ratio (SNR)
2. Deep Q-Learning controls the UAV movement and power allocation for each pilot sequence

Q-Learning: Reward

- Based on the sum spectral efficiency of users that the UAV has a connection with
- Spectral efficiency is calculated by dividing the capacity by the bandwidth *B*

$$C_g = B \log_2(1 + \gamma_g)$$

- The SINR equation is calculated using a relaxed version of original SINR equation

$$\gamma_{gw}(\tilde{\mathbf{p}}) = \frac{(M - |\mathcal{G}_{a(g)}|)\tau\rho_g\beta_{gg}^2 p_{gw}}{(1 + \tau\mathcal{E})(1 + \sum_{g'\in\mathcal{G}} \sum_{w'\in\mathcal{W}} \mu_{g'g}(\boldsymbol{\mu})p_{g'w'}) + (M - |\mathcal{G}_{a(g)}|) \sum_{g'\in\mathcal{I}_g\setminus g} \sum_{w'\in\mathcal{W}} \rho_{g'}\beta_{g'g}^2 p_{g'w'}}$$

[1] Guan et al.

Q-Learning: Reward

- Lastly, the reward is multiplied by the ratio of the power chosen over the max possible power

$$R_g = C_g \left(\frac{p_{gw}}{p_{max}} \right)$$

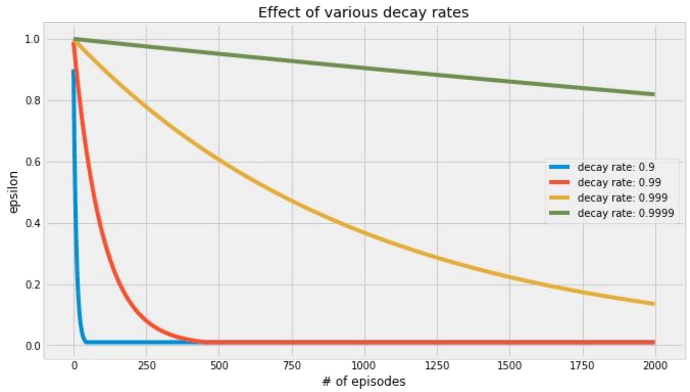
- Ensures the agent finds only one pilot sequence per user that maximizes the capacity

Q-Learning: Epsilon Greedy Strategy

- Epsilon starts off close to one and then decays nonlinearly
- Exploration vs Exploitation:
  - Random number generated
    - Higher than epsilon, agent explores environment
    - Lower than epsilon, agent exploited environment

$$\alpha = \alpha_{end} + \left(\alpha_{start} - \alpha_{end}\right)\left(e^{-n_{step}\lambda}\right)$$

Calculation of epsilon

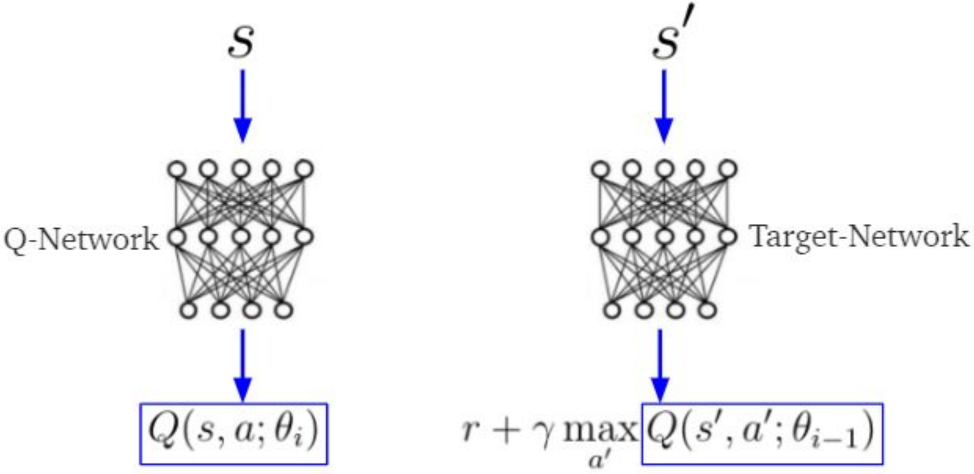


Deep Q Learning: Neural Network

- Input layer:
    - Dimension: Number of states
      (Also tested having the input be the x and y position of the UAVs plus the connection information to each user to reduce input dimension size)
- 2 fully connected hidden layers
    - Dimensions:
        - 200
        - 200
- Output layer
    - Dimension: Number of Actions
- Updates network using the Adam optimizer

Deep Q-learning: Policy and Target Network

- Loss is calculated by taking the Mean Square Error (MSE) between the the q-values calculated in the policy network and the optimal q-values calculated in the target network
- To help avoid instability, the target network is only updated periodically



Deep Q- Learning: Parameters

- Experience Replay: Includes state of environment, action taken, reward given, and next state
- Replay Memory: Array that stores up to N number of experience replays and gets sampled randomly during training
- Batch Size: Number of randomly sampled experience replays from replay memory used for training
- Target Update: Number of episodes before the target network gets updated
- Memory Size: Variable that controls size of replay memory
- Learning Rate: How fast the neural network learns

Assumptions

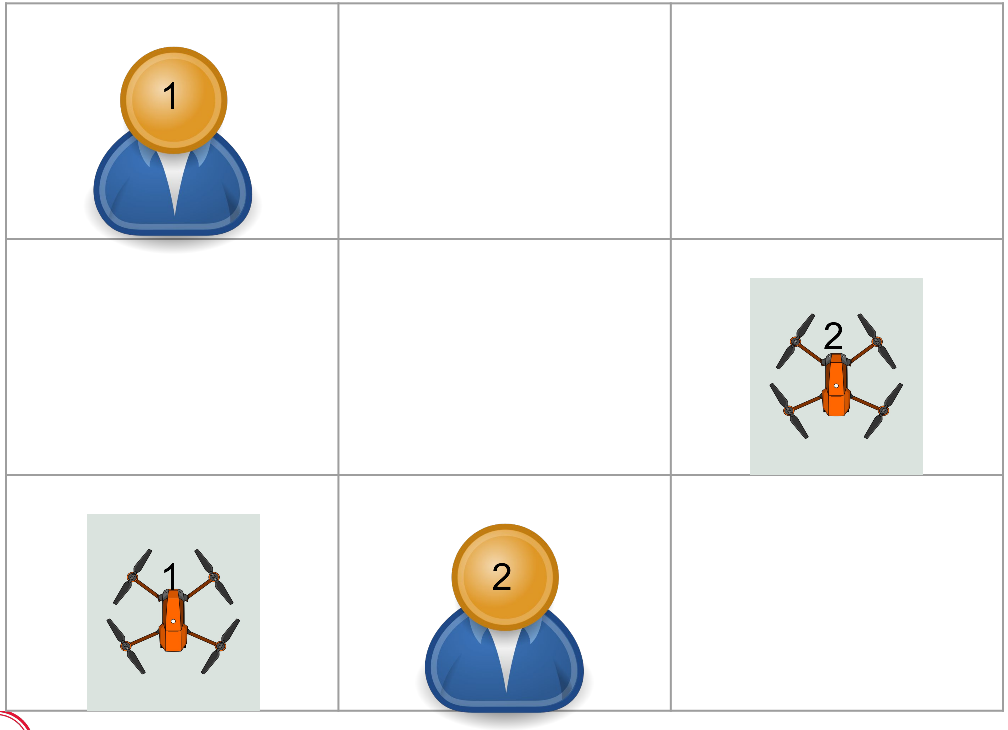
- Users can only be connected to one UAV
- UAVs can be connected to any number of users
- UAVs can take one action at each time step
  - Move up, down, left, right
  - Change a power level for a pilot for one of the users
- All UAVs operate in a 500m x 500m grid divided into equally sized blocks
- UAVs height is a constant 100m above the ground

Assumptions

- Noise power is 10^{-8} mW
- Path loss factor is set to 2
- Signal to Interference Plus Noise Ratio (SINR) for ground user *g* takes into account
  - Channel estimation error
  - Type of linear spatial multiplexing/demultiplexing
  - Power control
  - Noncoherent intercell interference
  - Coherent intercell interference due to pilot contamination

Demonstration

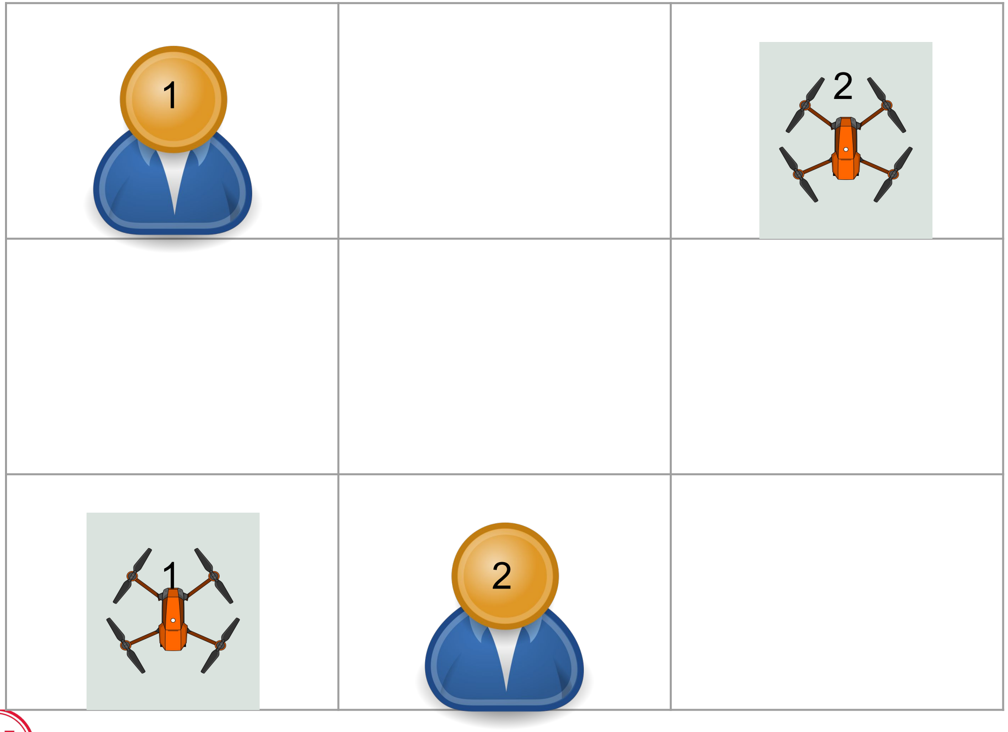
- Power 1 is the power assigned and pilot sequence 1
- Power 2 is the power assigned to pilot sequence 2



	UAV 1	UAV 2
User 1	Power 1: 0 Power 2: 0	
User 2	Power 1: 0 Power 2: 0	

Demonstration

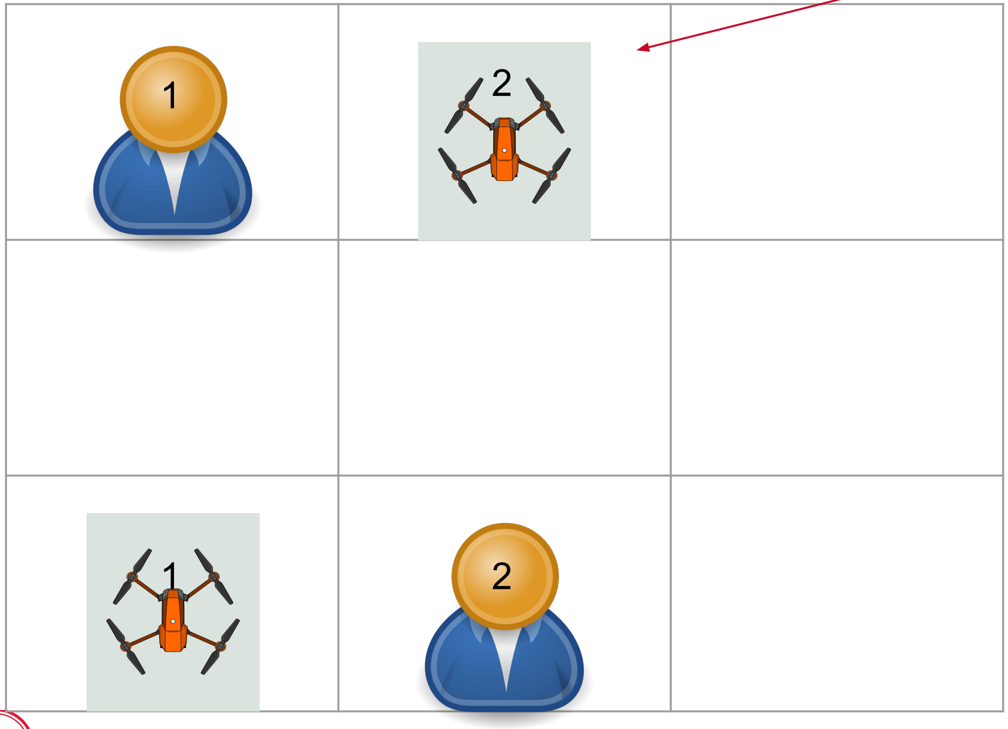
- UAV 1 sets Power 2 for user 1 to 2
- UAV 2 moves up



	UAV 1	UAV2
User 1	Power 1: 0 Power 2: 2	
User 2	Power 1: 0 Power 2: 0	

Demonstration

UAV 2 is now closer to user 1



	UAV 1	UAV 2
User 1		Power 1: 0 Power 2: 0
User 2	Power 1: 0 Power 2: 1	

Performance Results Formulation

- First term is a constant
- Equivalent to minimizing the sum negative logs of the distance between the users and UAVs

$$Minimize: \sum_{g \epsilon G} - log_2(d_{gg}(x, y, z))$$

- Approximated by calculating a 20000 x 20000 grid of all possible x and y positions of the UAVs
- Spacing between grid blocks is 25 mm

Performance Results Formulation

- Assume:
  - All inferences can be ignored
  - SINR of each user is significantly greater than 1
- Problem then becomes maximizing sum capacity of users based on UAVs location
  - Ignore user association
  - Ignore pilot assignment

$$Maximize: \sum_{g \in G} C_g(x, y, z)$$

Performance Results Formulation

- Relation between capacity and SINR

$$C_g(x, y, z) = Blog_2(1 + \gamma_g(x, y, z))$$

$$\gamma_g(x, y, z) >> 1$$

$$C_g(x, y, z)$$

$$\approx Blog_2(\gamma_g(x, y, z))$$

$$\leq Blog_2(M\tau\rho_g p_0 \zeta_{gg}^2 H_{gg}^2(x, y, z))$$

$$= Blog_2\left(\frac{M\tau\rho_g p_0 \zeta_{gg}^2}{d_{gg}^x(x,y,z)}\right)$$

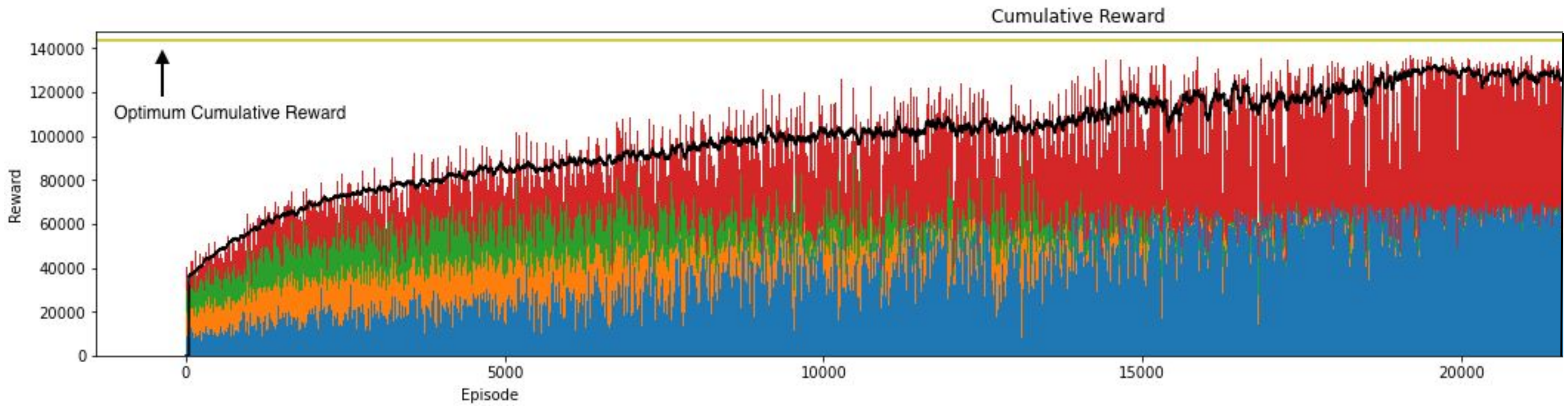
$$= Blog_2(M\tau\rho_g p_0 \zeta_{gg}^2) - xBlog_2(d_{gg}(x, y, z))$$

List of Variables Table

Variable	Sim 1
Batch Size	50
Gamma	0.99
Starting Epsilon	0.9
Ending Epsilon	0.001
Epsilon Decay	$5*10^{-7}$

Variable	Sim 1
Target Update	20
Memory Size	200
Learning Rate	0.001
Number of Episodes	10000
Max Steps Per Episode	4000

Results



Episode

Results

- Average was able to come within 95% the optimum cumulative reward
- With a max step size of 2000, the agents were able to converge at around the 20000 episode
- Agents were able to select a unique pilot sequence for the users even though there is no preference for which user gets what pilot sequence
  - Ex:
    - Pilot 1 for user 1 and pilot 2 for user 2, or
    - Pilot 2 for user 1 and pilot 1 for user 1

Discussion and Future Work

- UAVs were able to get within 95% of the optimum value
- Since spacing between grid blocks is small, difference between global optimum found and true global optimum is small
  - Note: Global optimum cannot be used as the general solution because it assumes the UAVs know the users' locations, which is not the case

- Future Work:
  - Increase number of UAVs
  - Increase number of users
  - Train agents in multiple environments to reduce the possibility of overfitting

References

[1] Z. Guan, N. Cen, T. Melodia and S. M. Pudlewski, "Distributed Joint Power, Association and Flight Control for Massive-MIMO Self-Organizing Flying Drones," in IEEE/ACM Transactions on Networking, vol. 28, no. 4, pp. 1491-1505, Aug. 2020, doi: 10.1109/TNET.2020.2985972.

[2] X. Li, "Deployment of Drone Base Stations for Cellular Communication Without Apriori User Distribution Information," 2018 37th Chinese Control Conference (CCC), 2018, pp. 7274-7281, doi: 10.23919/ChiCC.2018.8482797.

[3] X. Sun, N. Ansari and R. Fierro, "Jointly Optimized 3D Drone Mounted Base Station Deployment and User Association in Drone Assisted Mobile Access Networks," in IEEE Transactions on Vehicular Technology, vol. 69, no. 2, pp. 2195-2203, Feb. 2020, doi: 10.1109/TVT.2019.2961086.

[4] W. Shi et al., "Multiple Drone-Cell Deployment Analyses and Optimization in Drone Assisted Radio Access Networks," in IEEE Access, vol. 6, pp. 12518-12529, 2018, doi: 10.1109/ACCESS.2018.2803788.

[5] J. Kim and J. Kim, "Access Management using Vickrey-Clarke-Groves Auction in Terrestrial-Drone Networks," 2021 International Conference on Information Networking (ICOIN), 2021, pp. 317-320, doi: 10.1109/ICOIN50884.2021.9333869.

[6] A. Akarsu and T. Girici, "Resilient Deployment of Drone Base Stations," 2019 International Symposium on Networks, Computers and Communications (ISNCC), 2019, pp. 1-5, doi: 10.1109/ISNCC.2019.8909193.

[7] J. Jang, H. J. Yang and S. Kim, "Learning-Based Distributed Resource Allocation in Asynchronous Multicell Networks," 2018 International Conference on Information and Communication Technology Convergence (ICTC), 2018, pp. 910-913, doi: 10.1109/ICTC.2018.8539654.

[8] G. Zhao, Y. Li, C. Xu, Z. Han, Y. Xing and S. Yu, "Joint Power Control and Channel Allocation for Interference Mitigation Based on Reinforcement Learning," in IEEE Access, vol. 7, pp. 177254-177265, 2019, doi: 10.1109/ACCESS.2019.2937438.

[9] L. Bondan, M. A. Marotta, L. R. Faganello, J. Rochol and L. Z. Granville, "ChiMaS: A spectrum sensing-based channels classification system for cognitive radio networks," 2016 IEEE Wireless Communications and Networking Conference, 2016, pp. 1-7, doi: 10.1109/WCNC.2016.7564911.