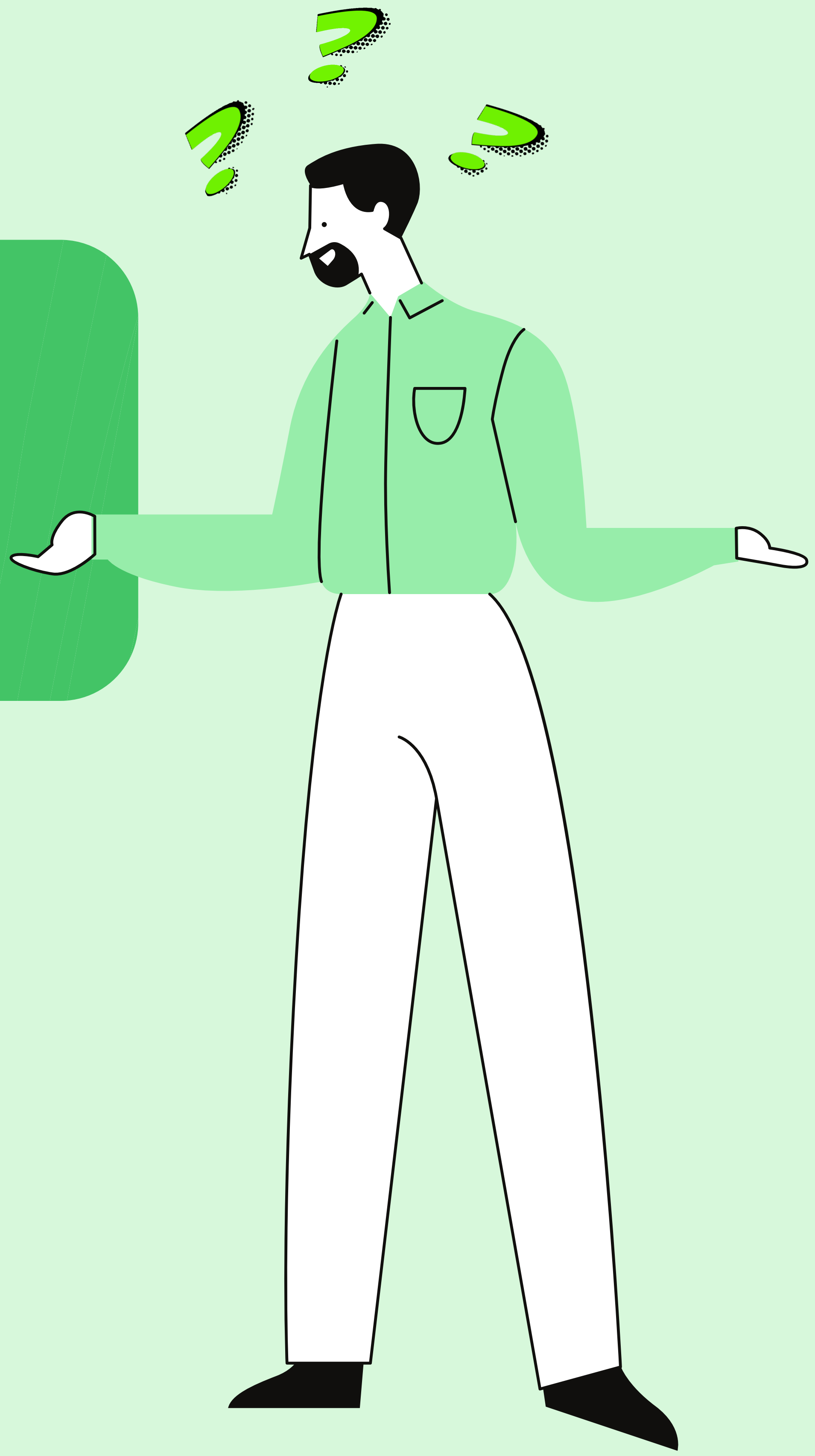
August 12, 2021

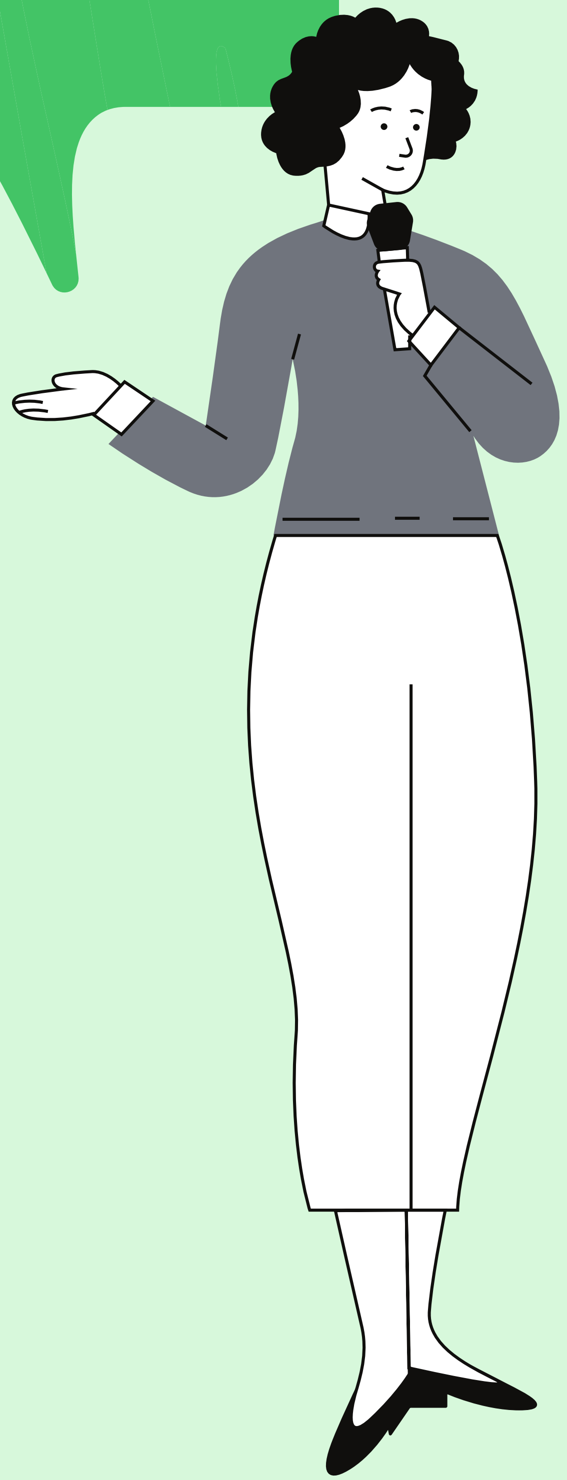
Multi-Agent Systems on Sensor Networks: A Distributed Reinforcement Learning Approach

A further breakdown on:
- Distributed Value Function DRL
- Optimistic DRL



Prepared by: Jennifer Ly

Today's Breakdown



August 11, 2021

1. Distributed Value Function
   - General
   - DVF Algorithm
2. Optimistic DRL
   - General
   - DVF Algorithm
3. Simulation
4. Results
5. Future Work

Distributed Value Function

Let's start with the general idea behind DVF DRL

Generally...

- Nodes communicated by exchanging information about their value functions
- Sum of the individual node value functions

=

- Total of future weighted reward sums from all of the nodes
- Nodes have access to the local rewards and can communicate their value functions with their neighbors

DVF Algorithm

land in, i.e. $V^i(s^i_t)$ for agent i at time t at each iteration.

The update rule at time step t for agent i are given by:

$$\underbrace{Q^i_{t+1}(s^i_t, a^i_t)} = (1-\alpha)Q^i_t(s^i_t, a^i_t) + \alpha\Big(r^i_{t+1}(s^i_{t+1}) + \gamma \sum_{j \in Neigh(i)} \boxed{f^i(j) V^j_t(s^j_{t+1})}\Big) \quad (2)$$

$$V^i_{t+1}(s^i_t) = \max_{a \in A^i} \underbrace{Q^i_{t+1}(s^i_t, a)} \quad (3)$$

discount factor

learning rate

Reward function /per agent

$$r^i(s^i) = G^i(s^i) - C^i$$

Optimistic DRL

Let's start with the general idea behind Optimistic DRL

Generally...

- All agents are given the same reward function
  - Project **all** information about the whole system as a 'single agent'
- Finds optimal policies in deterministic environments
  - Deterministic is very predictable, your next state is completely dependent on your current state
- Not so great in Stochastic environments
  - Continuous variable –> not exactly predictable

OptDRL Algorithm

$$q^i_{t+1}(S_t, a^i_t) = \max\{q^i_t(S_t, a^i_t), (1-\alpha)q^i_t(S_t, a^i_t) + \alpha(r^i_{t+1}(S_{t+1}) + \gamma \max_{a \in A^i} q^i_t(S_{t+1}, a))\} \quad (5)$$

$$\Pi^i_{t+1}(S_t) = a^i_t \text{ iff } \max_{a \in A^i} q^i_t(S_t, a) \neq \max_{a \in A^i} q^i_{t+1}(S_t, a) \quad (6)$$

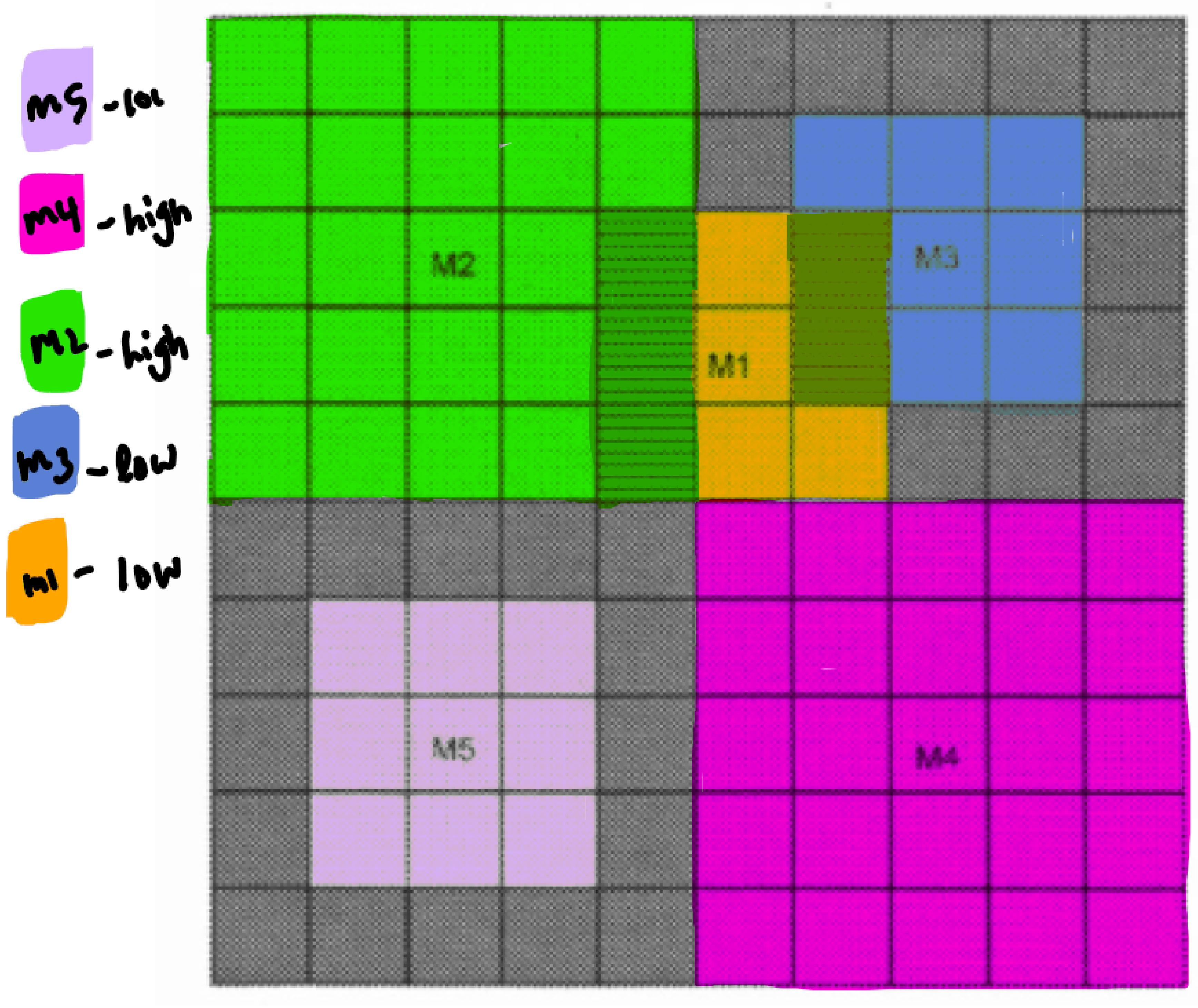
Global State

$$S = \prod_{i=1}^{m} S^i$$

Reward function

$$Glob\_rew(S) = G(S) - C(S)$$

Simulation

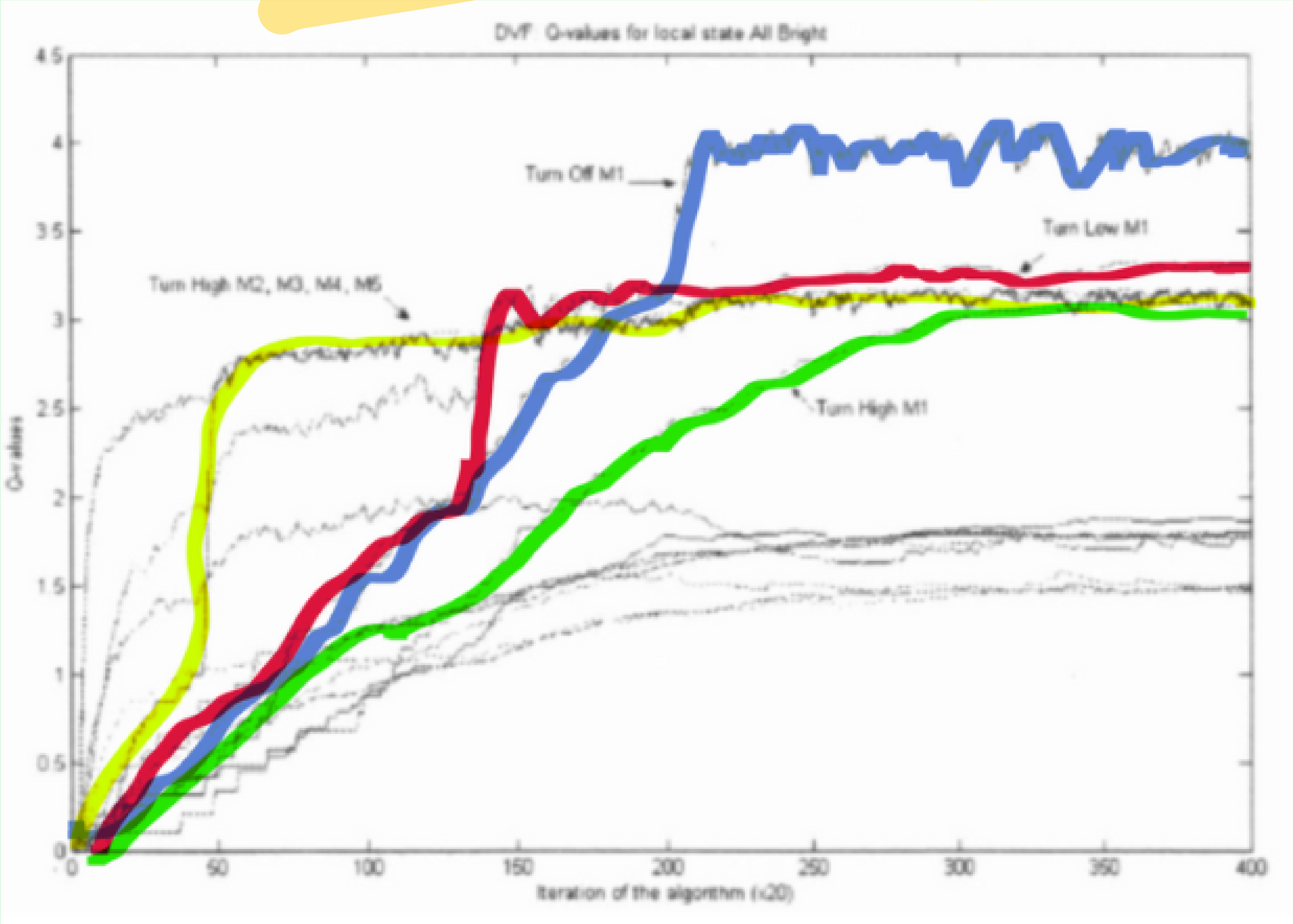


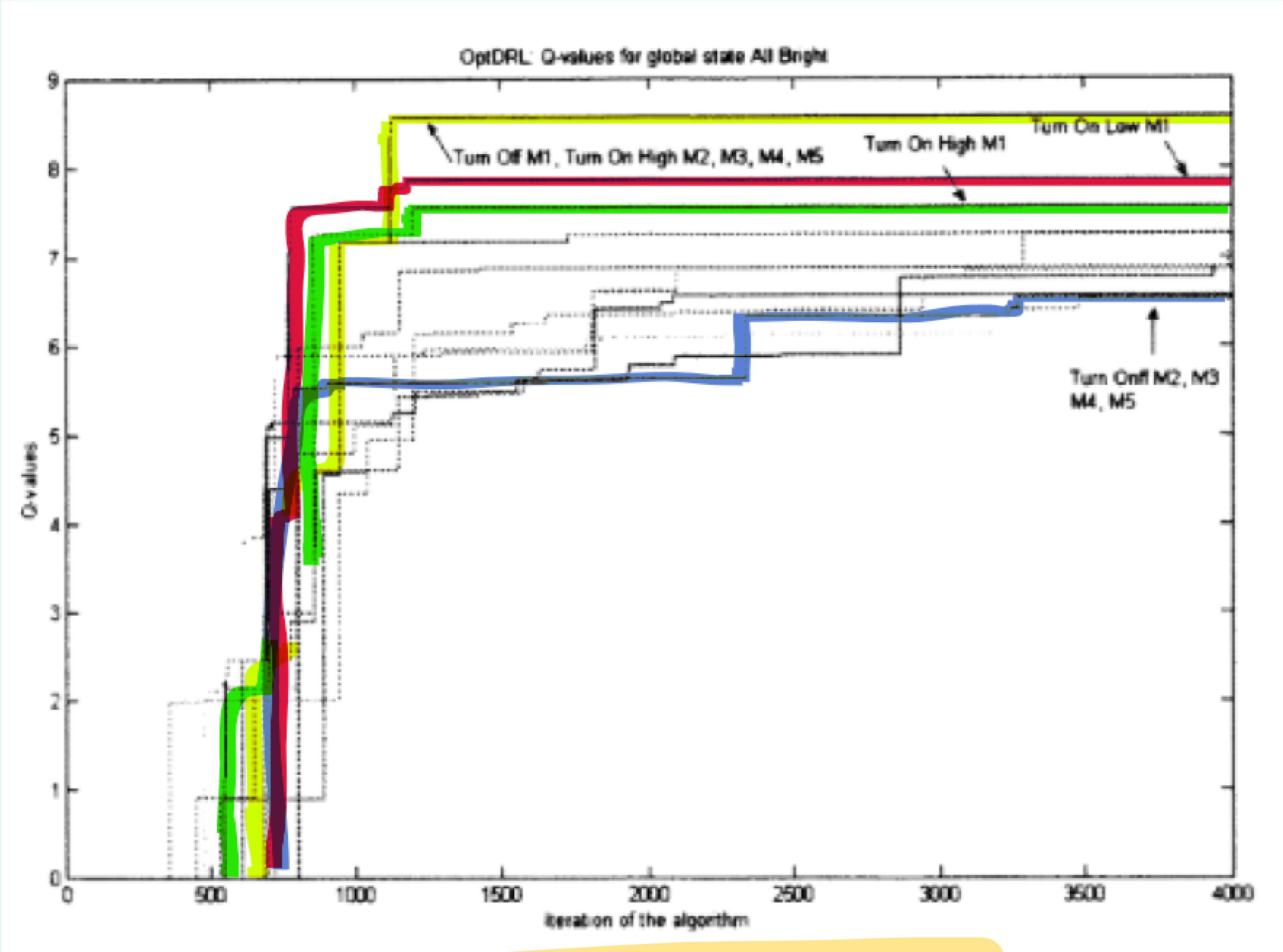
Setting

- 10x10 Grid
- 5 Agents
  - Off – no lights
  - low – 9 squares
  - high – 25 squares
- Goal: light up ALL 100 squares as efficiently as possible

Simulation Results

DVF DRL





OPT DRL

Simulation Results

TABLE 1: ENERGY CONSUMPTION (J) AND APPLICATION-LEVEL PE FORMANCE OF THE MAS WITH 5 MOTES DURING THE FIRST 4000 ITER. TIONS

	Ind Learners	*DVF*	*OptDRL*
Communication	0	1648.4	1681.0
Computation	954.9	971.7	1187.8
Lights LOW	2404	3014	1393
Lights HIGH	12721	12192	12123
Cells Bright	318323	317674	321766
Cumulative Reward	17429	17466	19078

TABLE 2: MEMORY REQUIREMENTS OF THE ALGORITHMS

	Expression	Actual values
IndLearners	$\lvert s^i\rvert \times \lvert A^i\rvert$	$2^{25} \times 3$
DVF	$\lvert s^i\rvert \times \lvert A^i\rvert + \lvert s^i\rvert$	$2^{25} \times 4$
OptDRL	$\lvert S\rvert \times \lvert A^i\rvert + \lvert S\rvert$	$2^{100} \times 4$

Integrating it with Current Project

- The use of multi agent to converge to an optimal policy does seem promising
  - Could use multiple agents to help converge to the reach the lowest communication cost

★ A potential issue is looking for real live coding examples
  - different terms
  - perhaps only present in theoretical studies
  - requires further digging for resources

Works Cited

Chen-Khong Tham and J. -. Renaud, "Multi-Agent Systems on Sensor Networks: A Distributed Reinforcement Learning Approach," 2005 International Conference on Intelligent Sensors, Sensor Networks and Information Processing, 2005, pp. 423-429, doi: 10.1109/ISSNIP.2005.1595616.

SUMMER REU 2021

THANK YOU!

ANY QUESTIONS , COMMENTS, CONCERNS?