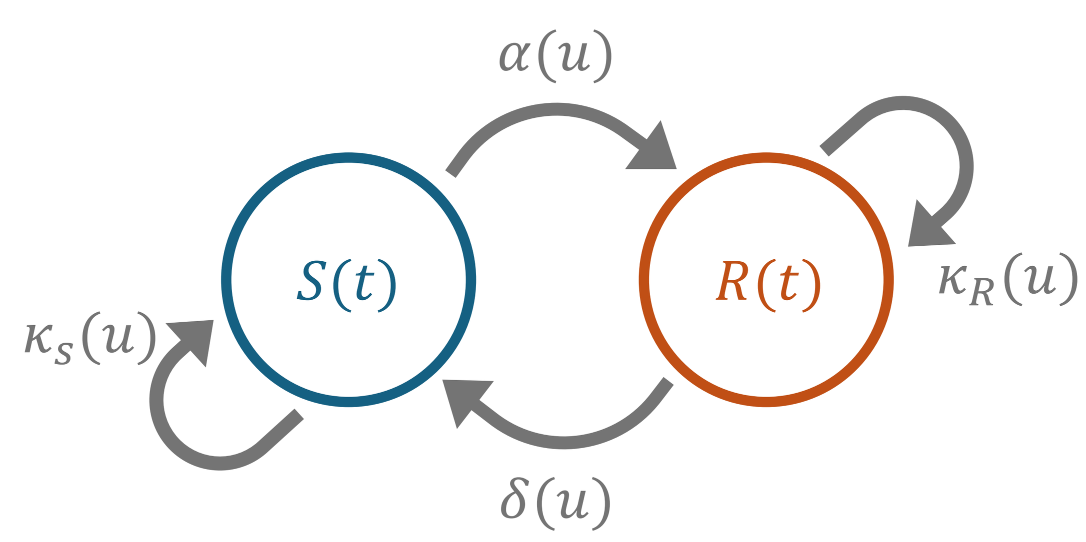
Reinforcement Learning for Optimal Control of Adaptive Cell Populations

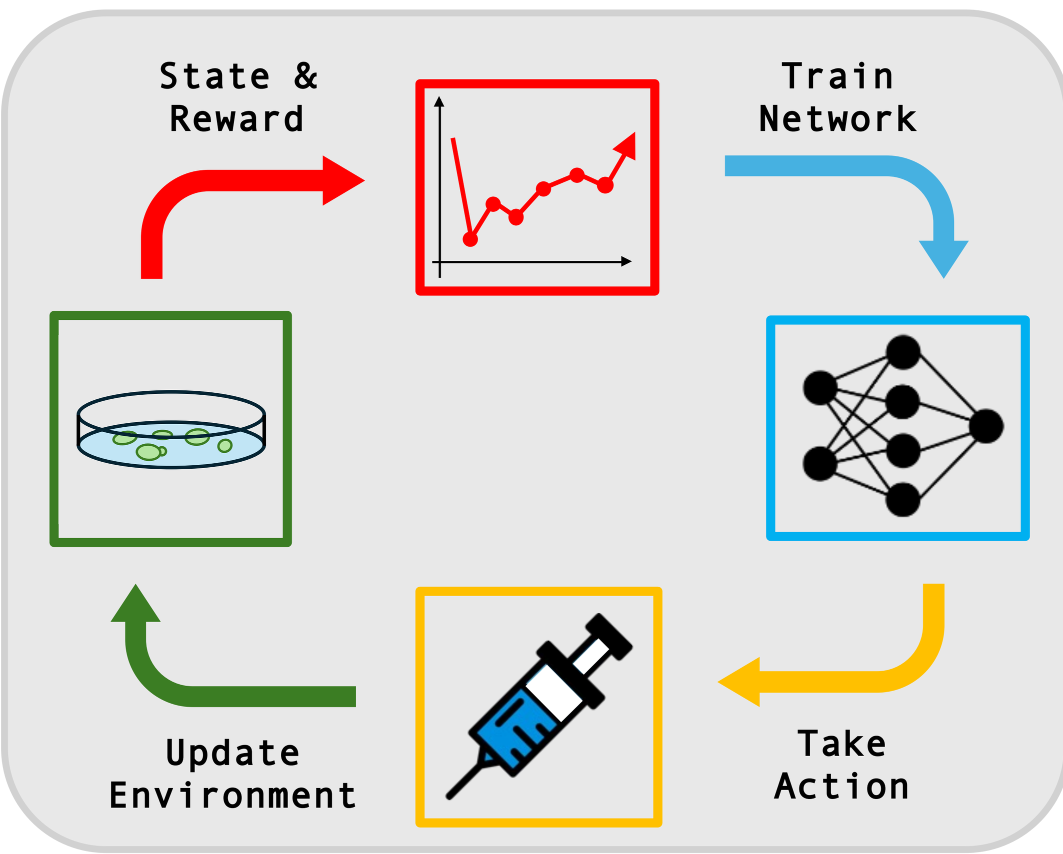
Josiah C. Kratz[1] and Jacob Adamczyk[2]
[1]: Computational Biology Department, Carnegie Mellon University
[2]: Department of Physics, University of Massachusetts Boston

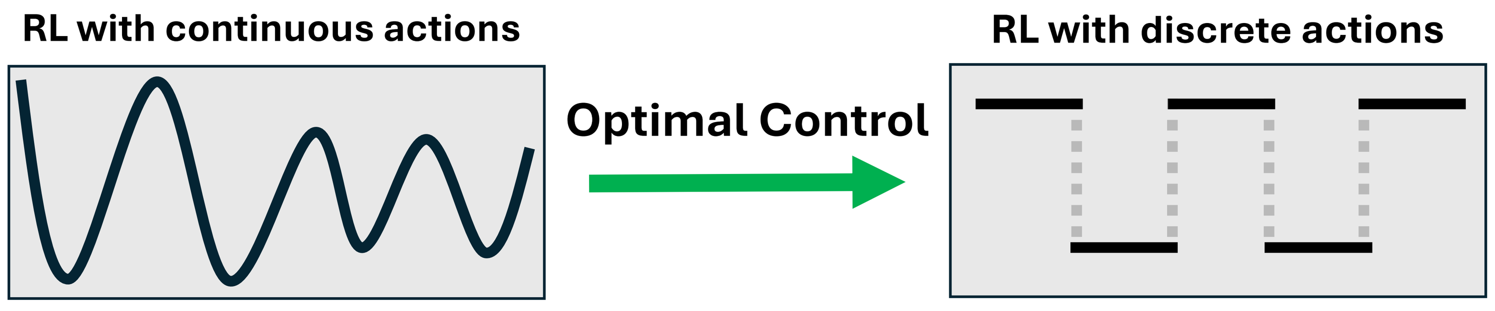
Novel Dynamics Model

- We consider a memory-based dynamics to model evolution of cancerous or bacterial populations;
- Drug-dependent switching rates determine whether the cells are in a susceptible or resistant state;
- When in the susceptible state, the drug kills cells most effectively.



- We pose the problem of reducing total cell population by varying the drug dosage over time;
- We use the framework of optimal control to derive results on bang-bang policies;
- We solve this problem in a custom environment with DQN, showing only clinically accessible data is needed for effective solutions.

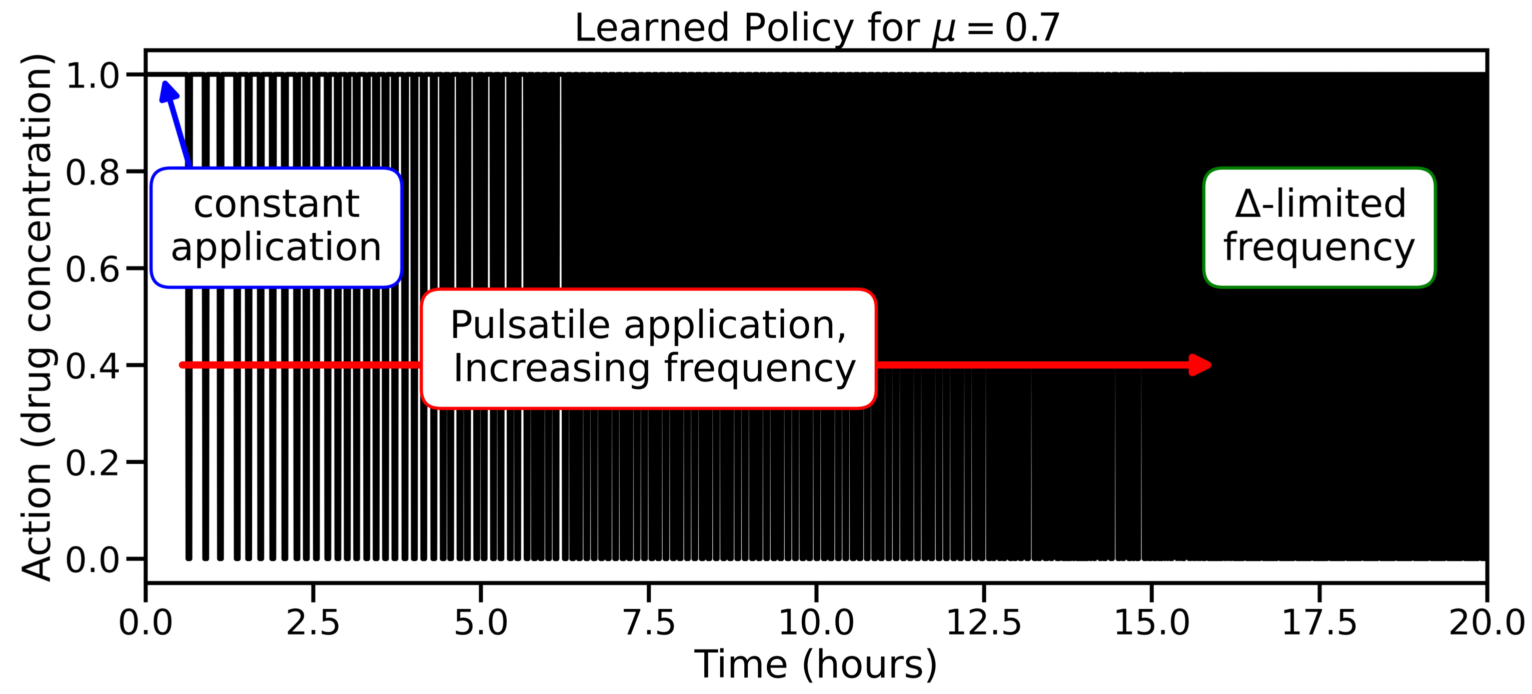




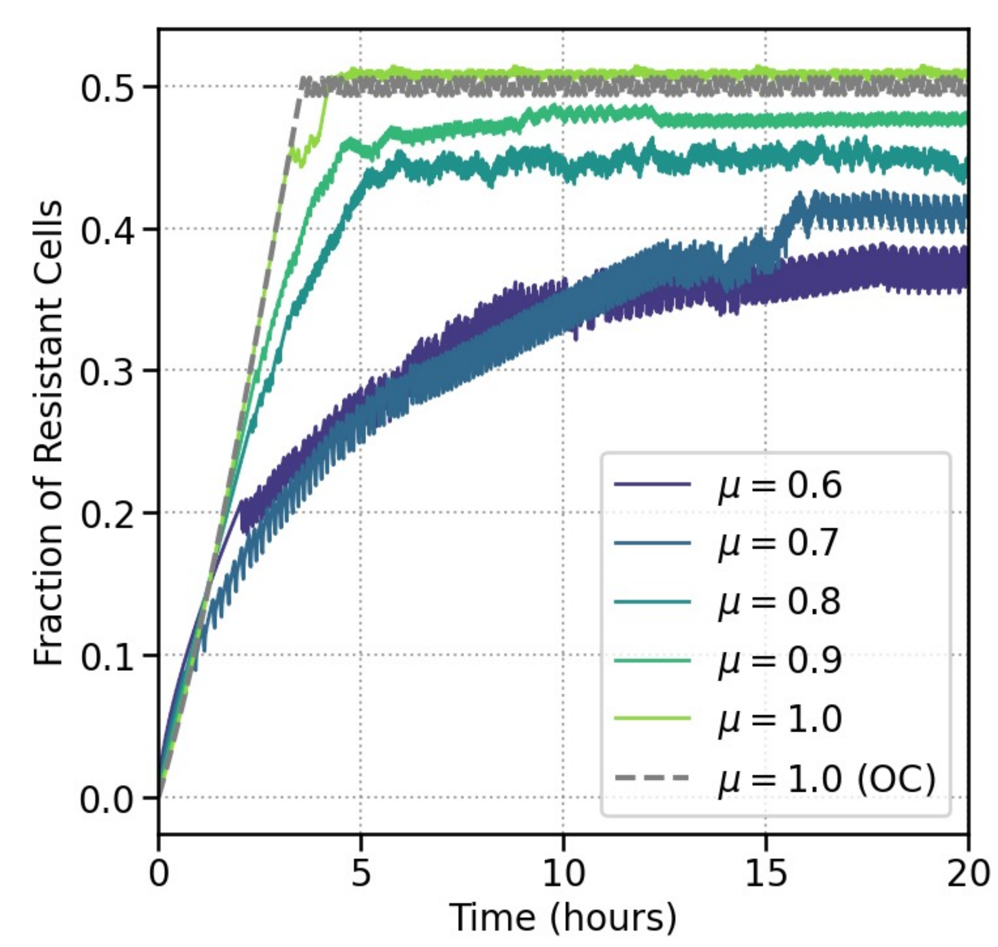
Experimental Results

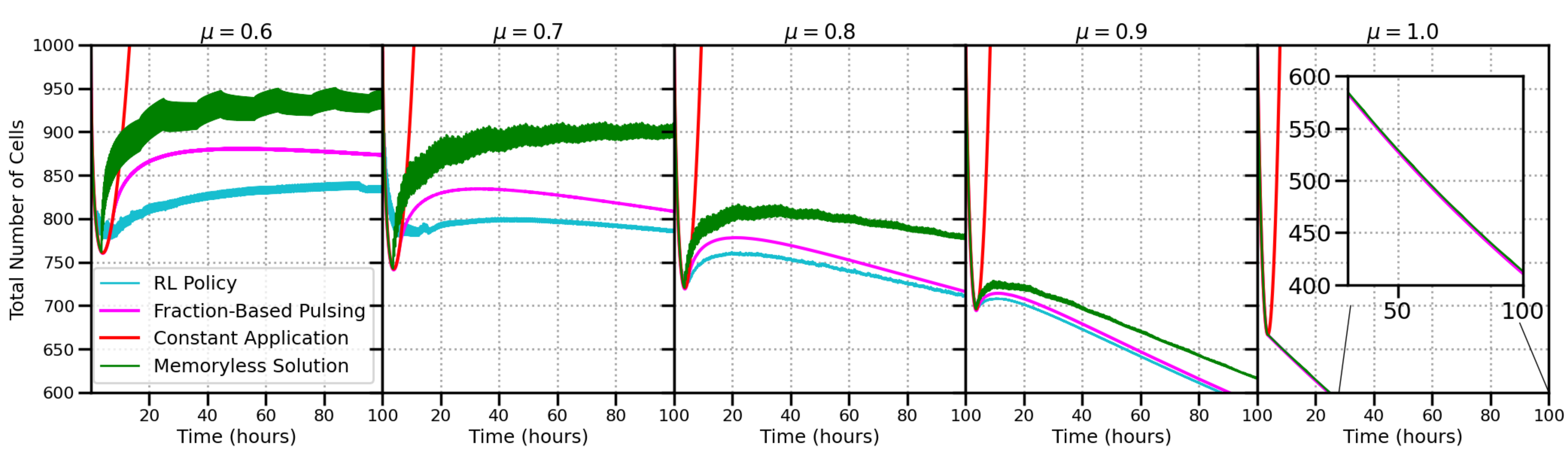
RL agents reliably converge to a dosing strategy with following regimes:

1. Constant application phase: kills the most cells upfront and brings resistant fraction to its limiting value
2. Pulsatile phase with increasing frequency: combats the ongoing cellular adaptation
3. Fastest pulsing frequency: the agent becomes bottlenecked by simulation/administration time.



- Despite not having access to resistant fraction, the RL agent infers the correct threshold values based on total growth rates
- Similar style of policy is found for the memory-based cases (keep the resistant fraction under some threshold value)





- RL recovers the optimal dosing protocol in memoryless case
- RL gives an effective protocol when OC becomes intractable

Conclusions and Outlook

- Model-free RL agents recover analytic solutions for dosing in the memoryless environment
- We find new, interpretable dosing techniques in the non-Markovian setting
- Insights from optimal control theory inform the design of the RL agent
- Compared to prior work, the agent only requires readily accessible data
- Generalization capabilities make the agent robust to changes in model parameters.

In the future, we hope to

1. Expand the deep RL model for the contextual MDP setting
2. Use real-world data to test the sim2real gap
3. Realize such dosing policies in experimental settings

Link to arXiv paper