ChemGymRL: an interactive framework for reinforcement learning for digital chemistry

Beeler, Chris; Ganapathi Subramanian, Sriram; Sprague, Kyle; Chatti, Nouha; Bellinger, Colin; Shahen, Mitchell; Paquin, Nicholas; Baula, Mark; Dawit, Amanuel; Yang, Zihan; Li, Xinkai; Crowley, Mark; Tamblyn, Isaac

ChemGymRL

An Interactive Framework for Reinforcement Learning for Digital Chemistry

Mark Crowley

Associate Professor
Department of Electrical and Computer Engineering
University of Waterloo

mcrowley@uwaterloo.ca
https://markcrowley.ca/

@compthink
@compthink@mastodon.social

Chris Beeler, Sriram Ganapathi Subramanian, Kyle Sprague, Nouha Chatti, Mitchell Shahen, Nicholas Paquin, Mark Baula, Amanuel Dawit, Zihan Yang, Xinkai Li,
Mark Crowley, Isaac Tamblyn, Colin Bellinger

VECTOR INSTITUTE

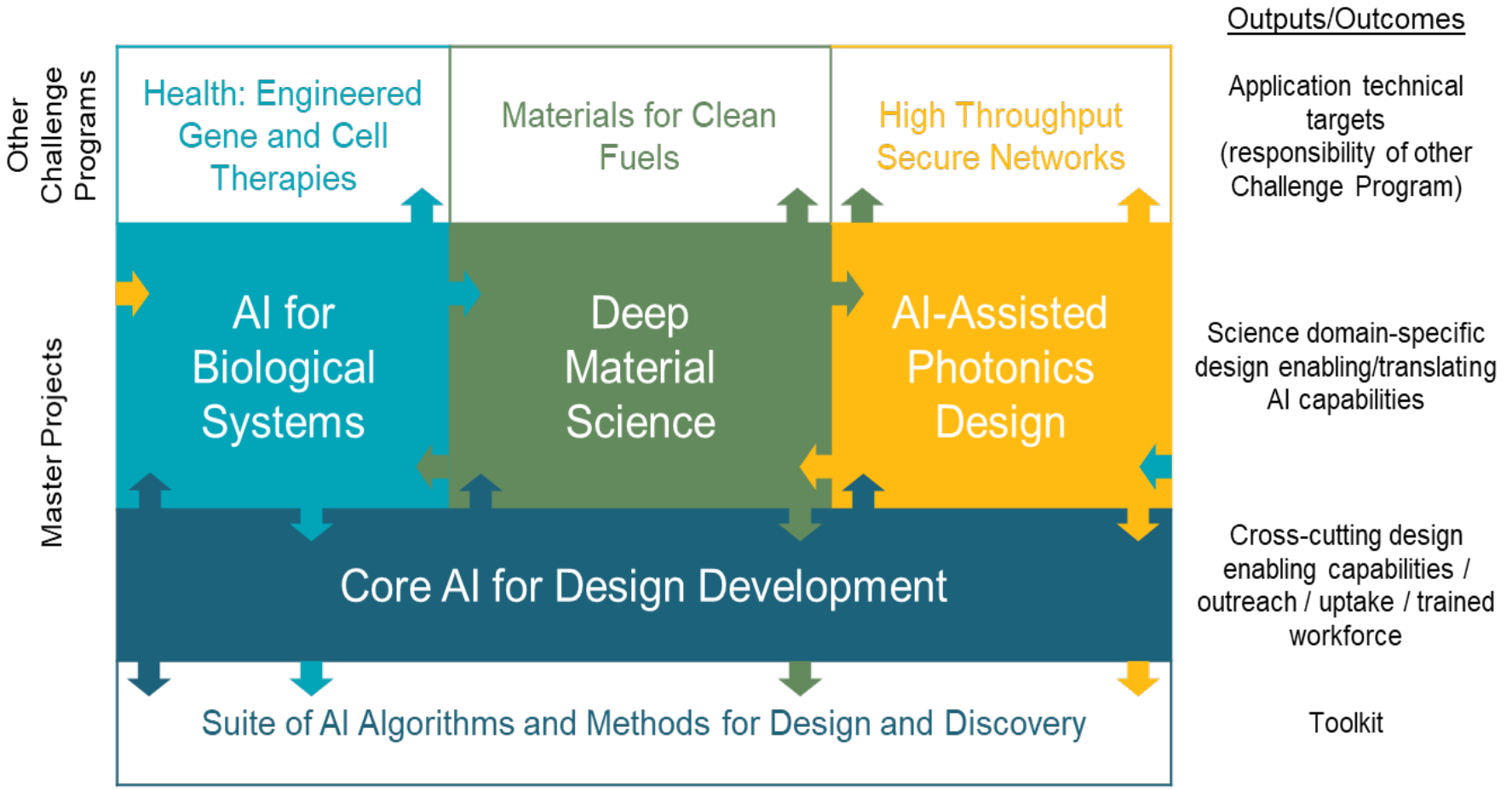
uOttawa

NRC·CNRC

UNIVERSITY OF WATERLOO

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

NRC AI for Design Challenge



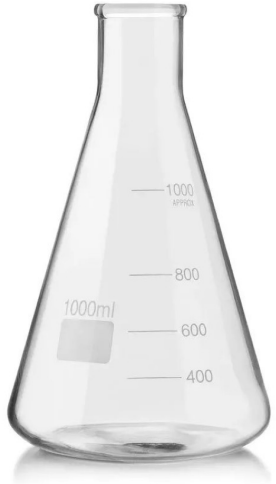
https://nrc.canada.ca/en/research-development/research-collaboration/programs/artificial-intelligence-design-challenge-program

Mark Crowley : ChemGymRL : RL for Digital Chemistry

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

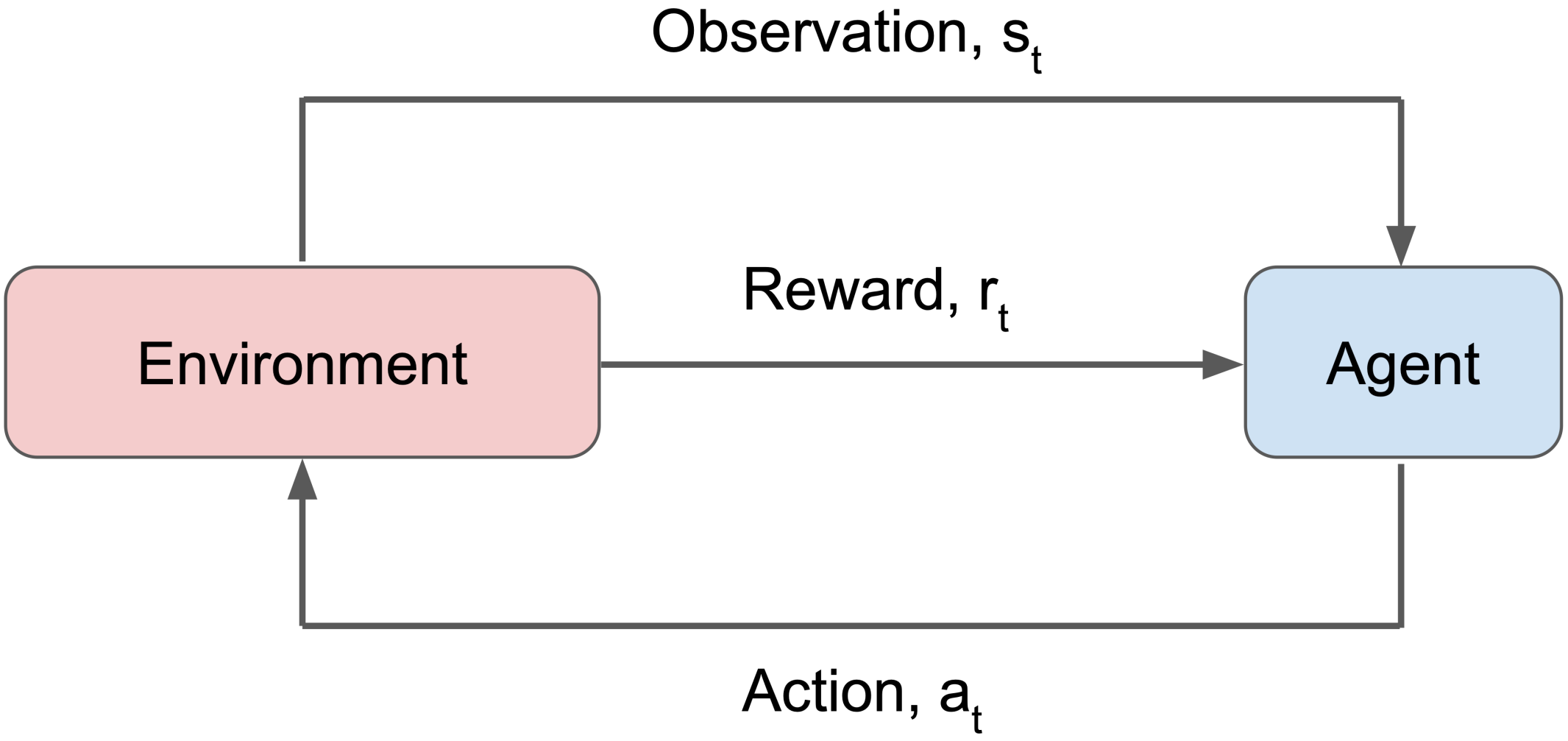
Why RL for Chemistry?

Reinforcement Learning(RL) is class of **Machine Learning** algorithms that learn by taking *actions*, making *observations*, viewing the *results* (or *rewards*), and *updating its model* (or hypothesis is you like)/policy/beliefs.

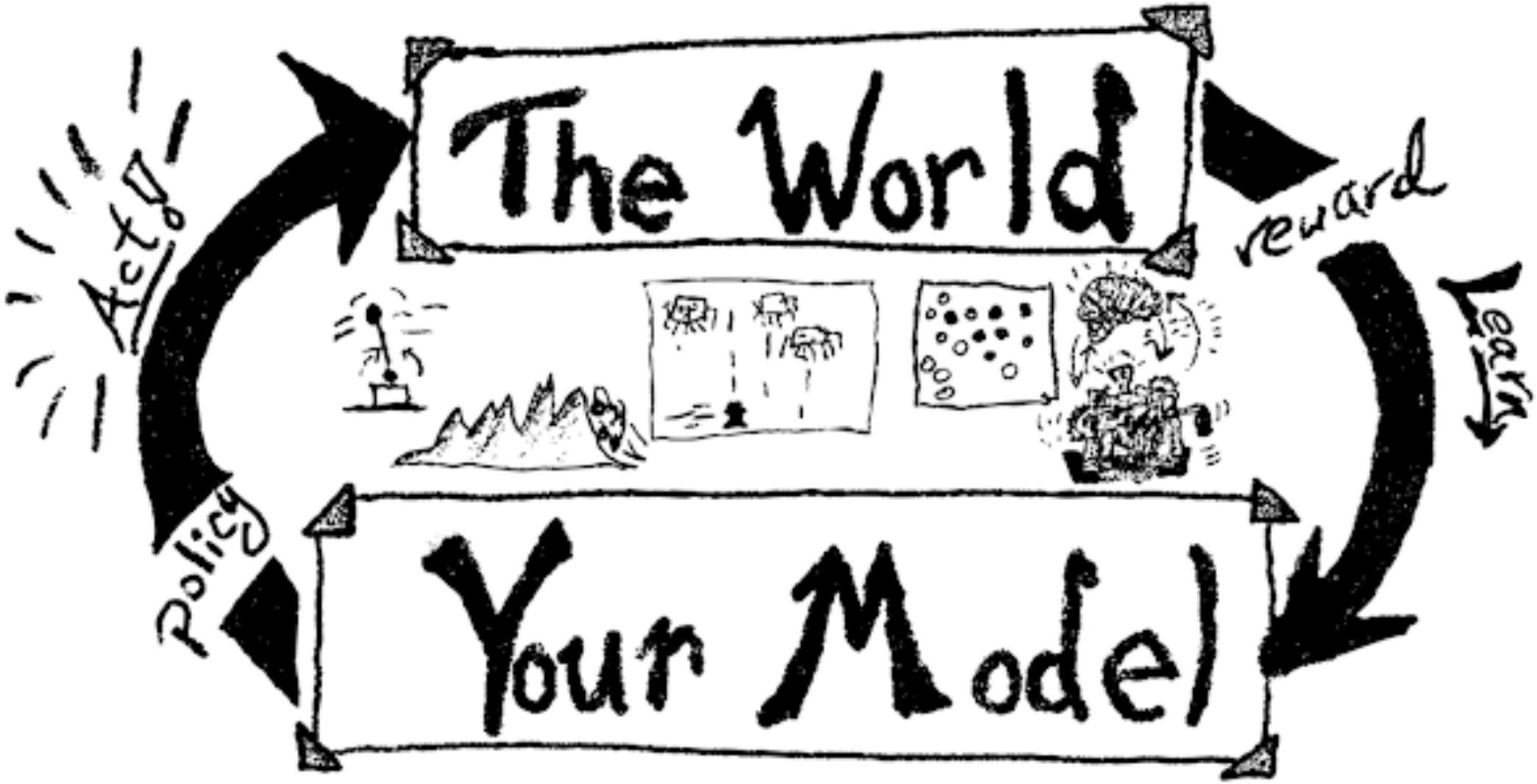


In other words…**RL is a perfect analogy for the experimental scientist!**

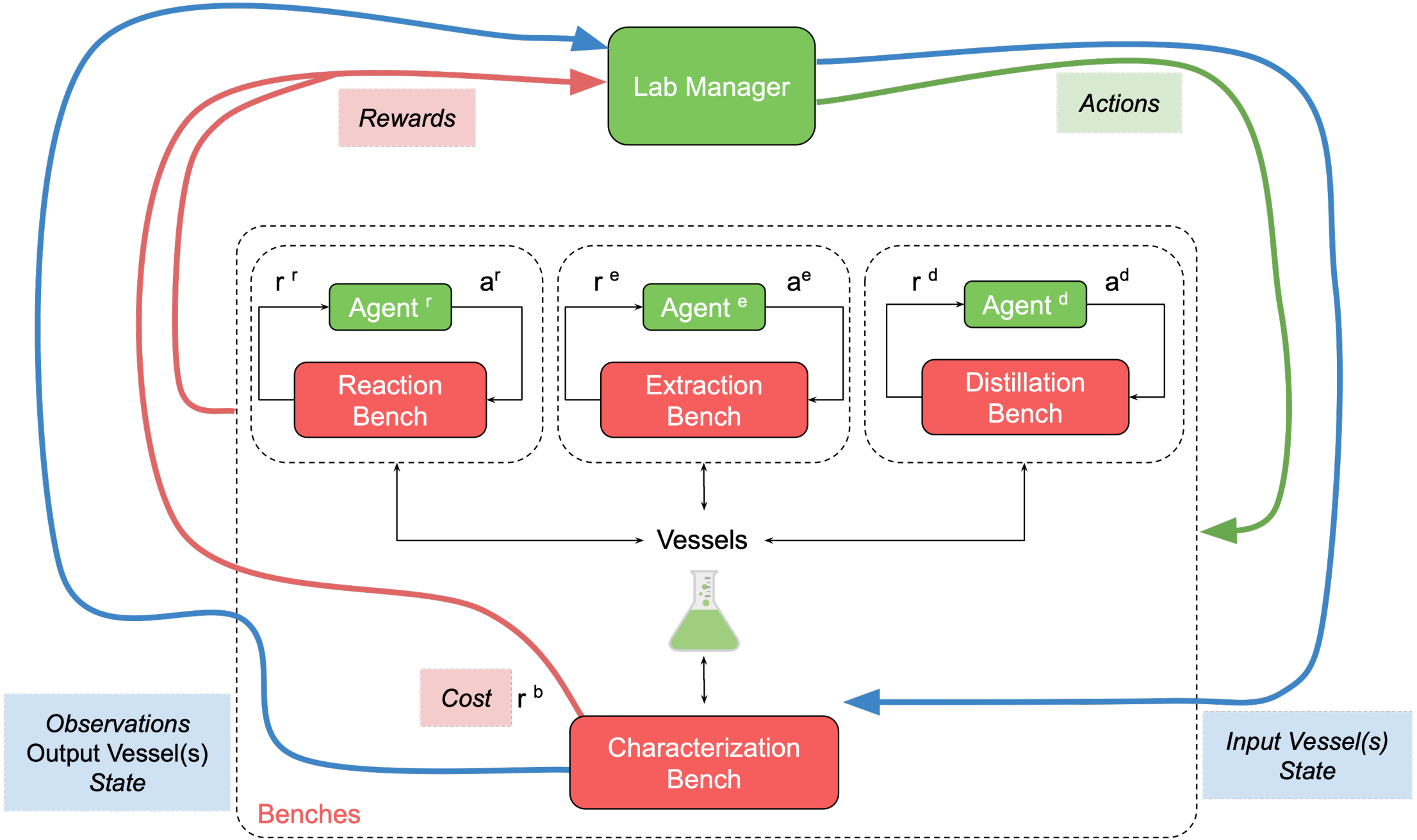
Reinforcement Learning



Reinforcement Learning

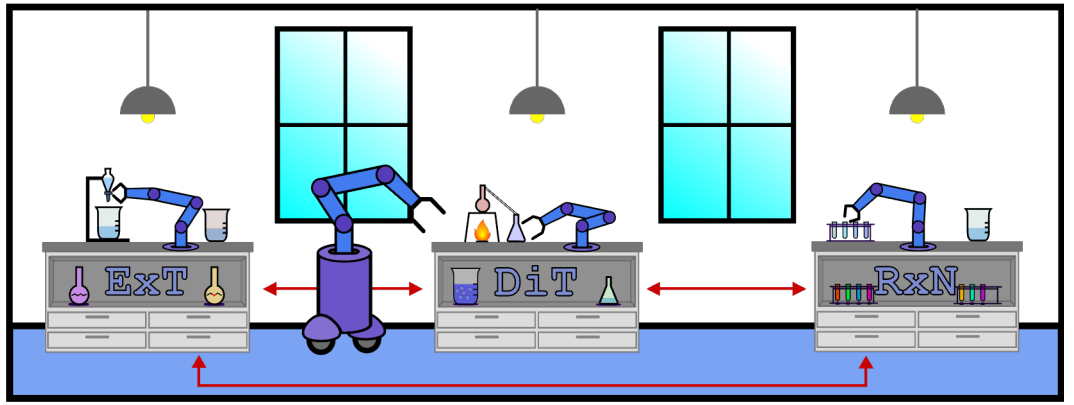


System Design



What is ChemGymRL?

Problem: RL is *very* data intensive, training robotic chemistry agents by taking actions in the real world is *infeasible* and ~~possibly~~ *dangerous*.



Benefits for RL: ML researchers always need new challenges, and Digital Chemistry has challenges (*causality*, *observations* with *impacts* and *costs*) which are rare in RL benchmarks, so offer a rich space to work in.

Our Solution: We introduce a set of highly customizable and *open-source* RL environments, implementing the standard *Gymnasium API*.

THE LABORATORY

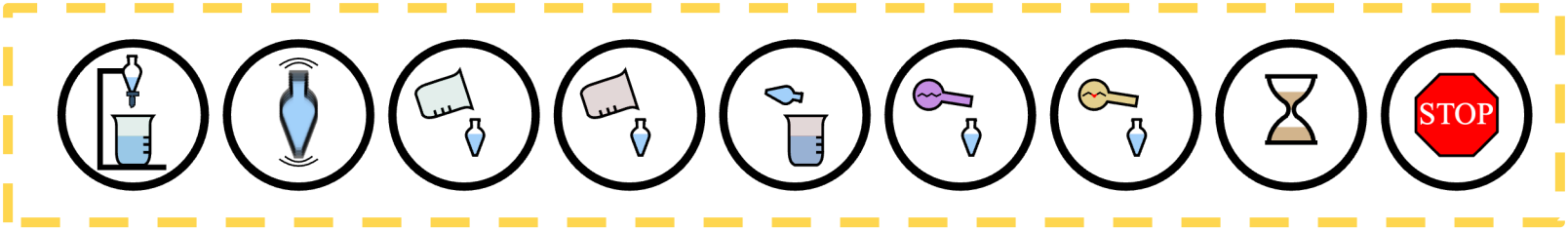
The ***ChemGymRL*** environment can be thought of as a virtual chemistry laboratory consisting of different benches where a variety of tasks can be completed.

- The laboratory is comprised of 3 basic elements:
- ***Vessels*** contain materials, in pure or mixed form, and track their hidden internal state
- **Shelves** are collections of vessels for input/output to benches
- **Benches** are simulations of particular chemistry activities

Each Bench Has…

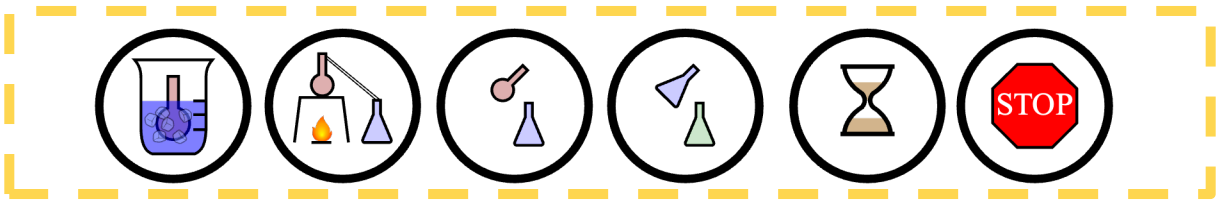
- ***Input:*** target material given as a one-hot vector
- ***State:*** vessels and contained materials
- ***Render*:** Human Rendering and various possible numeric outputs for learning

EXTRACTION BENCH



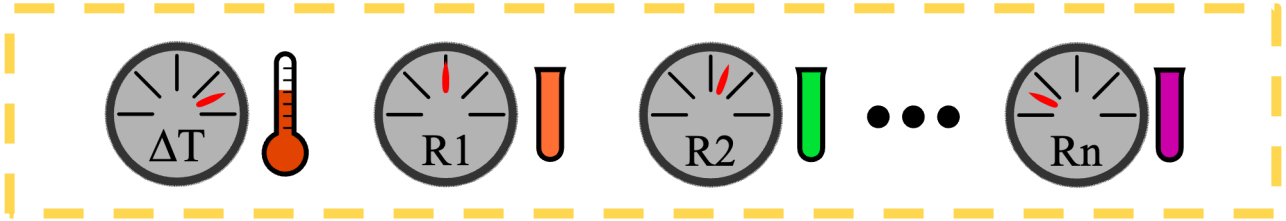
- Extraction is a method to **separate out undesired products** from the outputs of chemical reactions.
- The extraction bench (ExT) aims to *isolate* and *extract* certain dissolved materials from the input vessels.
- **Actions:** *Transferring* materials between different vessels and utilizing specifically selected solvents to separate materials from each other.

DISTILLATION BENCH



- The distillation bench (DiT) aims to **isolate certain materials** from an input vessel containing multiple materials.
- **Actions:** *Transferring* materials between a number of vessels and heating/cooling the vessel to separate materials from each other.
- **Rewards:** *Amount* and *purity* of target in output vessel.

REACTION BENCH



- The reaction bench (RxN) allows the agent to transform available reactants into various products via a chemical reaction.
- **Actions:** The agent has the ability to control:
  - the **temperature of the vessel** and
  - the **amounts of reactants** added.
- **Rewards**: After the 20 steps have elapsed, the agent receives a reward equal to the molar amount of the target material produced.
- The goal of the agent operating this bench is to modify the reaction parameters, in order to increase and/or decrease the yield of certain desired/undesired materials.
- The key to the agent's success in this bench is learning how best to allow certain reactions to occur such that:
- the yield of the desired material is maximized
- While the yield of the undesired material is minimized.

CHARACTERIZATION BENCH

- Not currently operated by an agent. Any *observation* of a vessel made by an agent must pass through this “bench”.

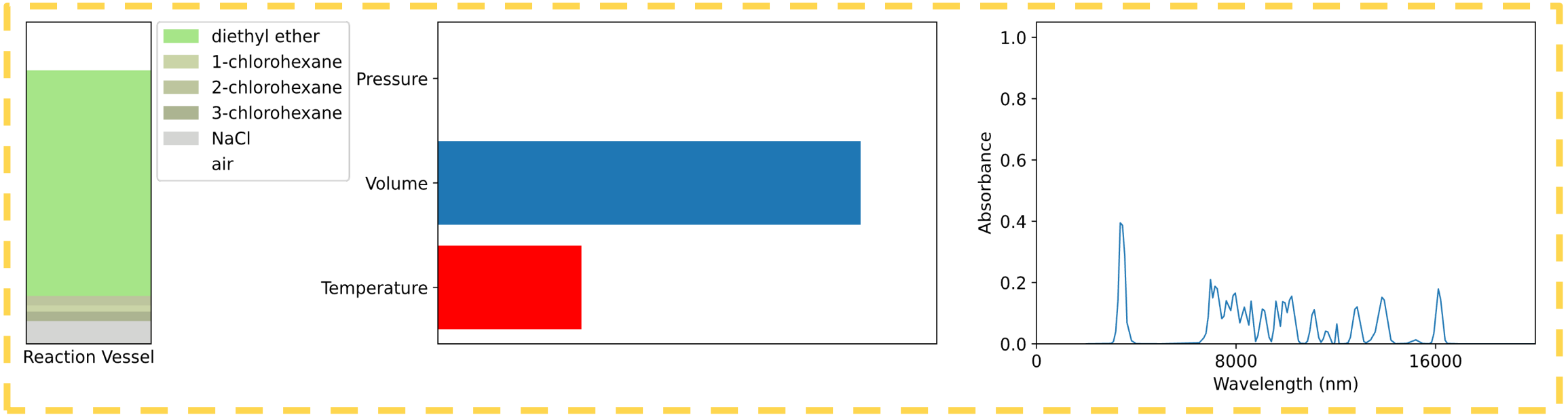
Current Observations:

- **Visual Layers:** same level of information as provided in human operated visualization (ordering of mixture layers)

CHARACTERIZATION BENCH

In future this will allow addition of:

- **Costed returns:** discounted rewards incurred while carrying out the process *minus* the costs incurred by the measurement policy.
- **Partial Observations:** incomplete or different resolutions of observations (eg. Visual vs. weighting vs. spectroscopy)
- **Lab Manager Agent:** this will be a hierarchical, or multi-task agent that operates many benches by sending agents to benches with given inputs and given target outputs

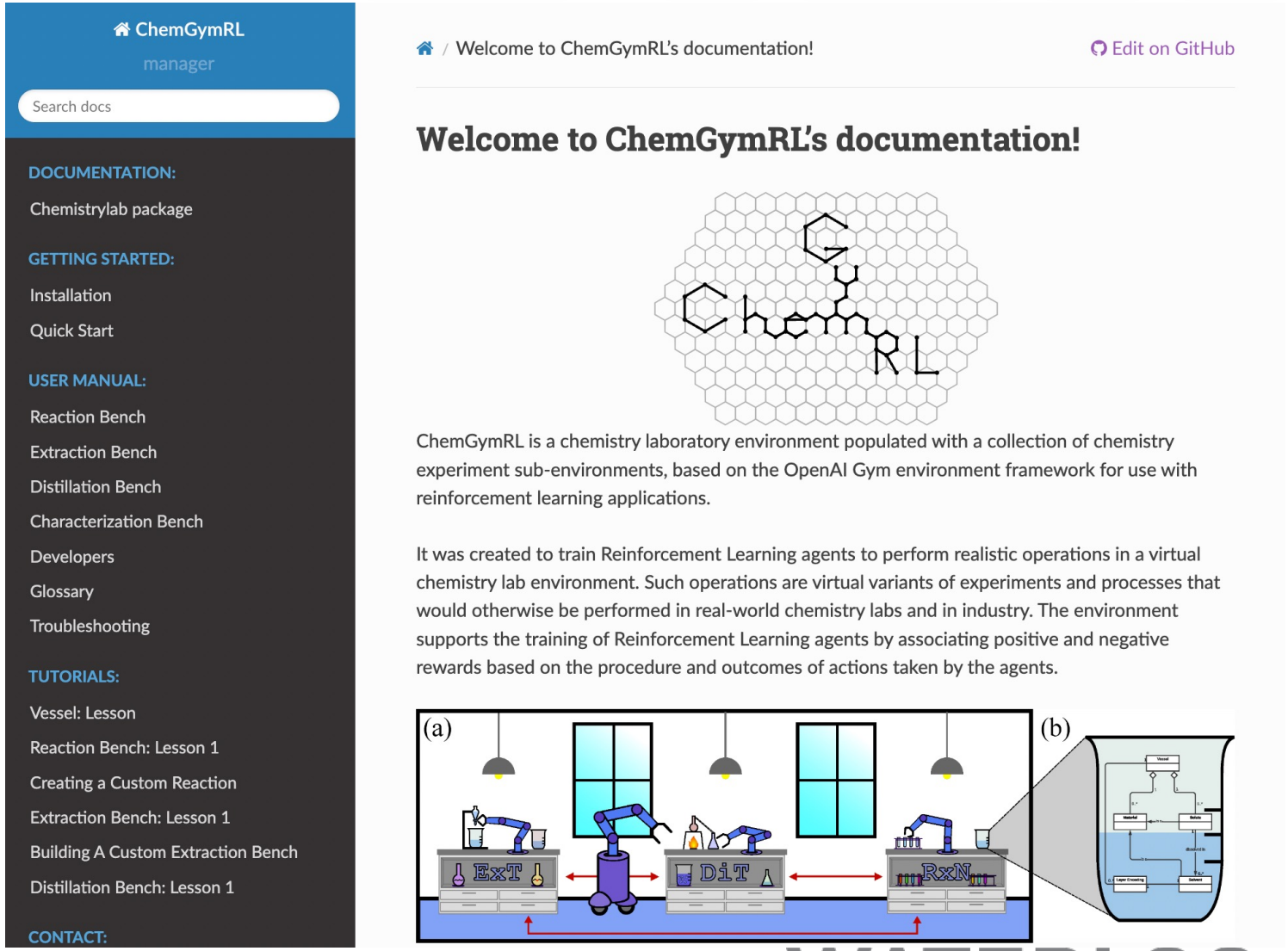


Experiments

- To test and demonstrate the framework, we trained RL agents were trained for 100K time steps across 10 environments in parallel (for a total of 1M time steps).
- **Samples:** 256 time steps of experience (in each environment) to update policies/Value-functions.
- **Replay buffer:** 1M experiences for **DQN**, **SAC**, and **TD3**.
- ***Exploration:*** first 30K steps of DQN used linear schedule from 1.0 down to 0.01, then fixed.
- ***Implementations:*** Stable Baselines 3

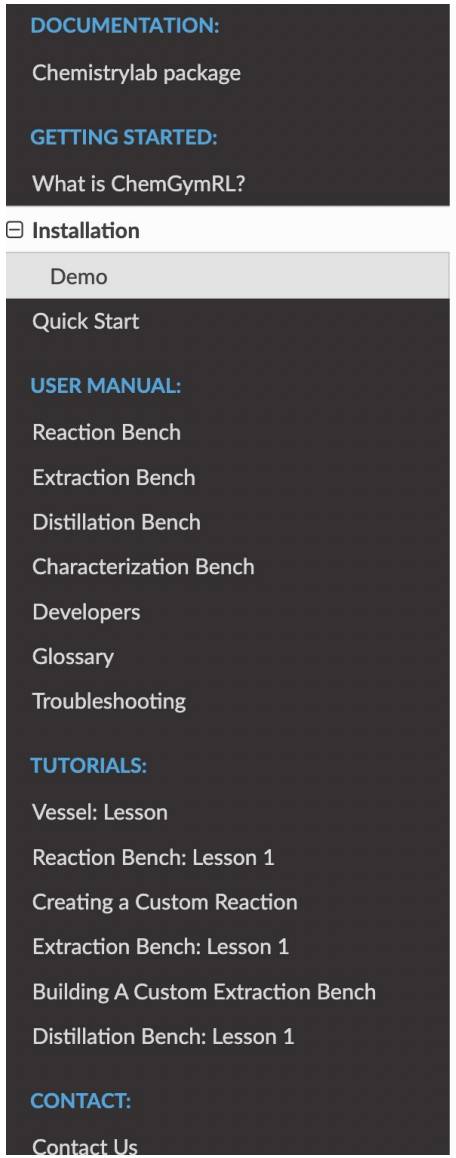
Full Documentation Website

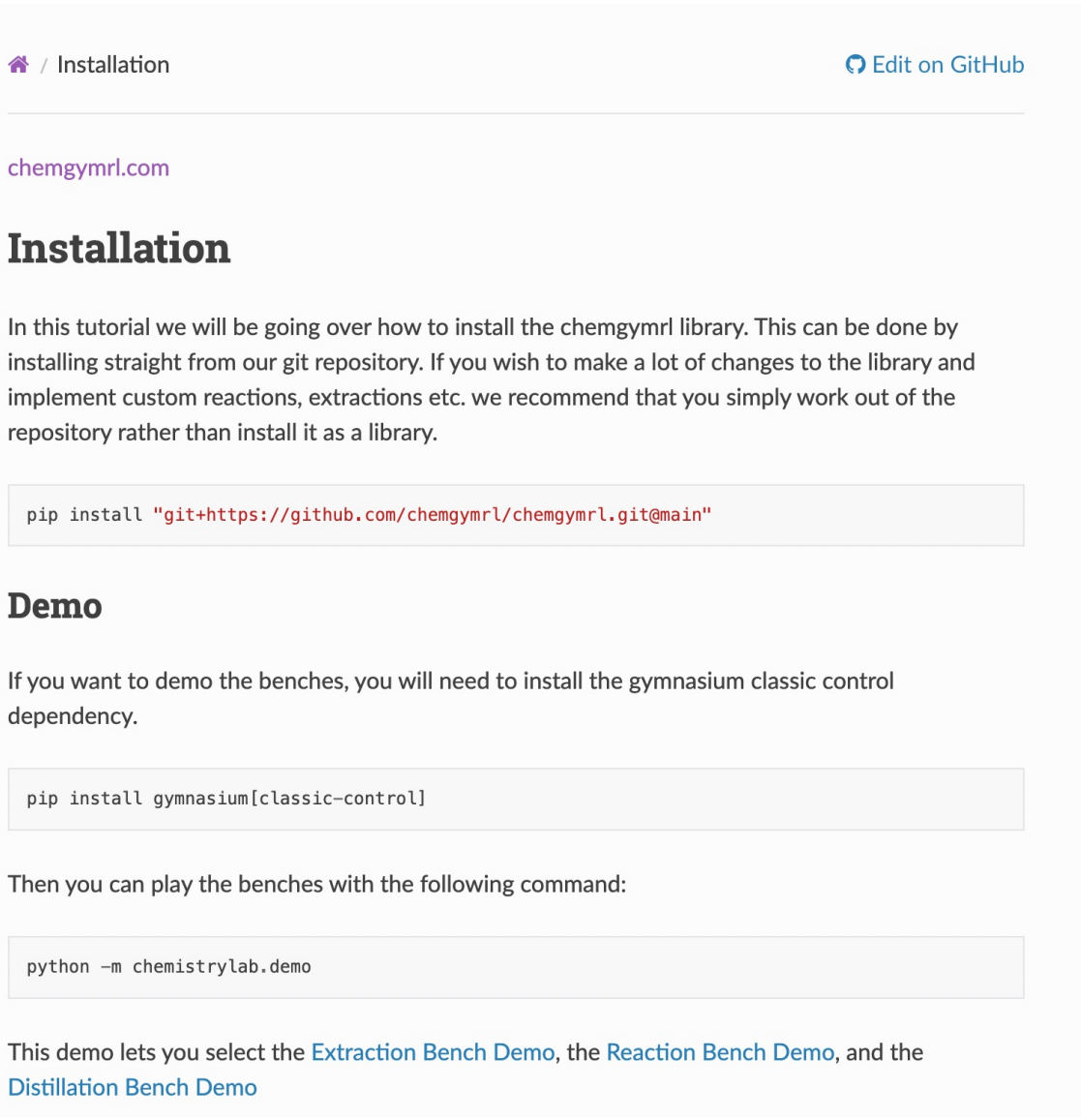
- ReadtheDocs site
- Libraries
- Installation instruction
- Tutorials
  - Running experiments with RL algorithms
  - Extending with new benches, formulations



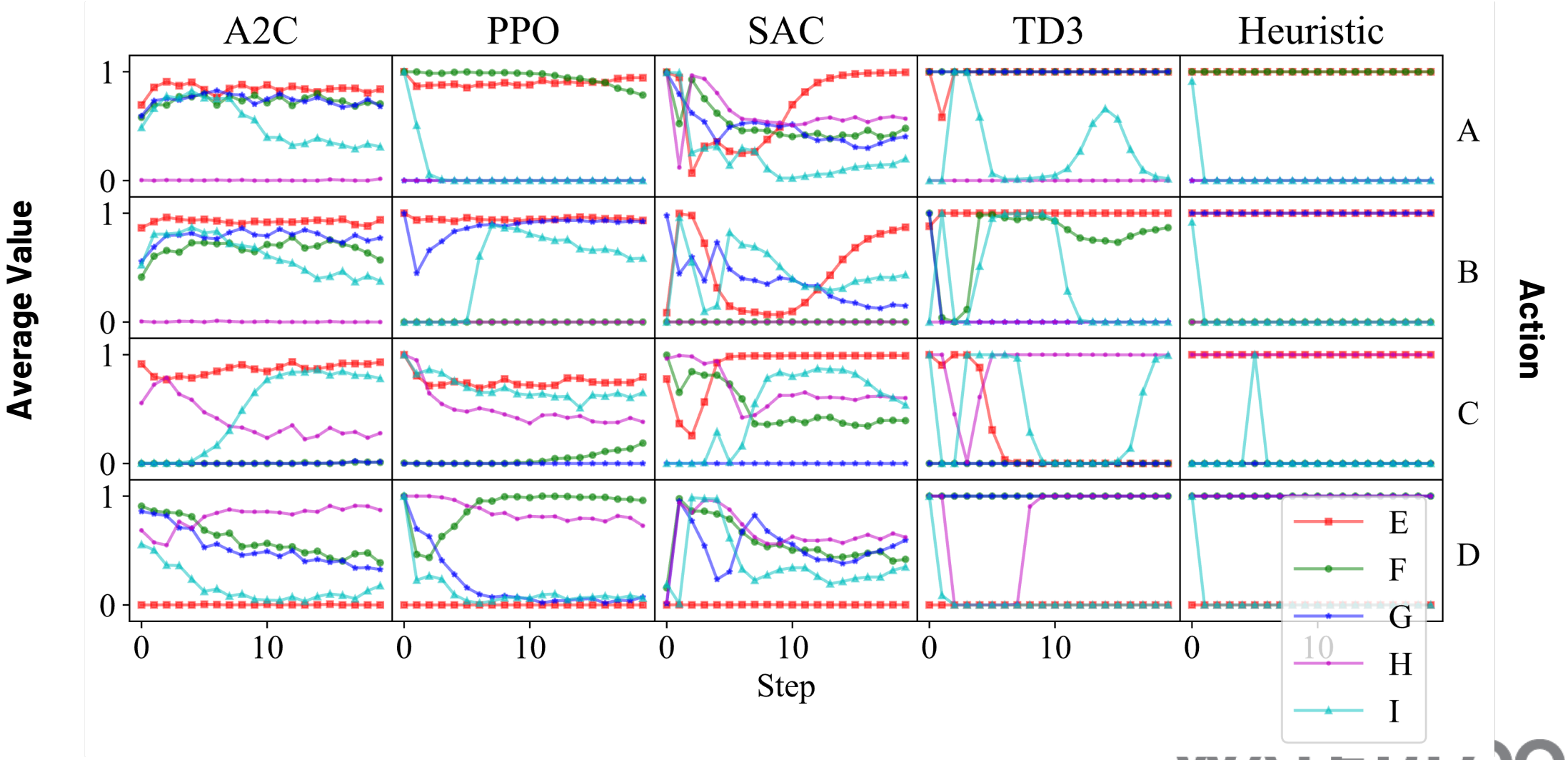
Easy Installation as a Python Library on Github

chemgymrl.com



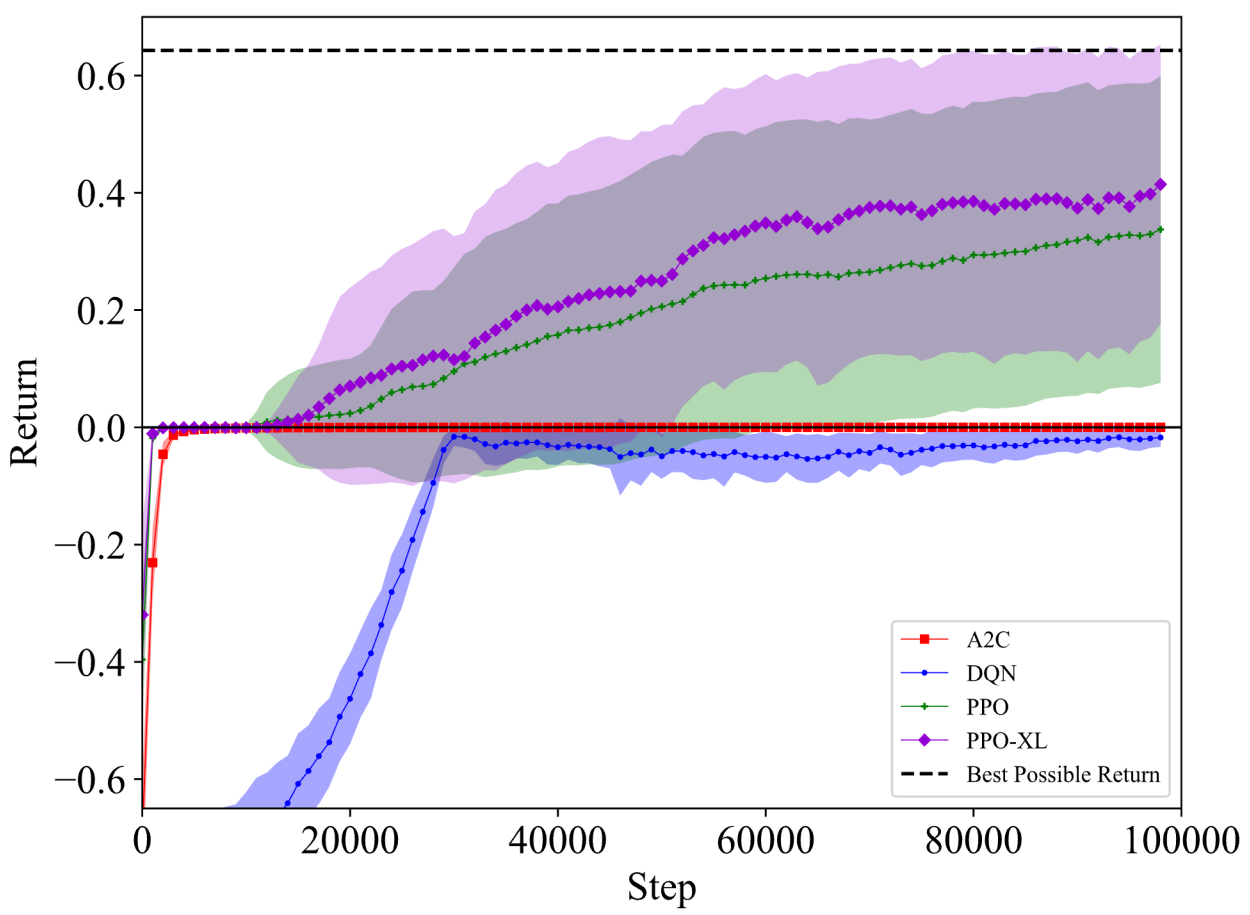


Experiments

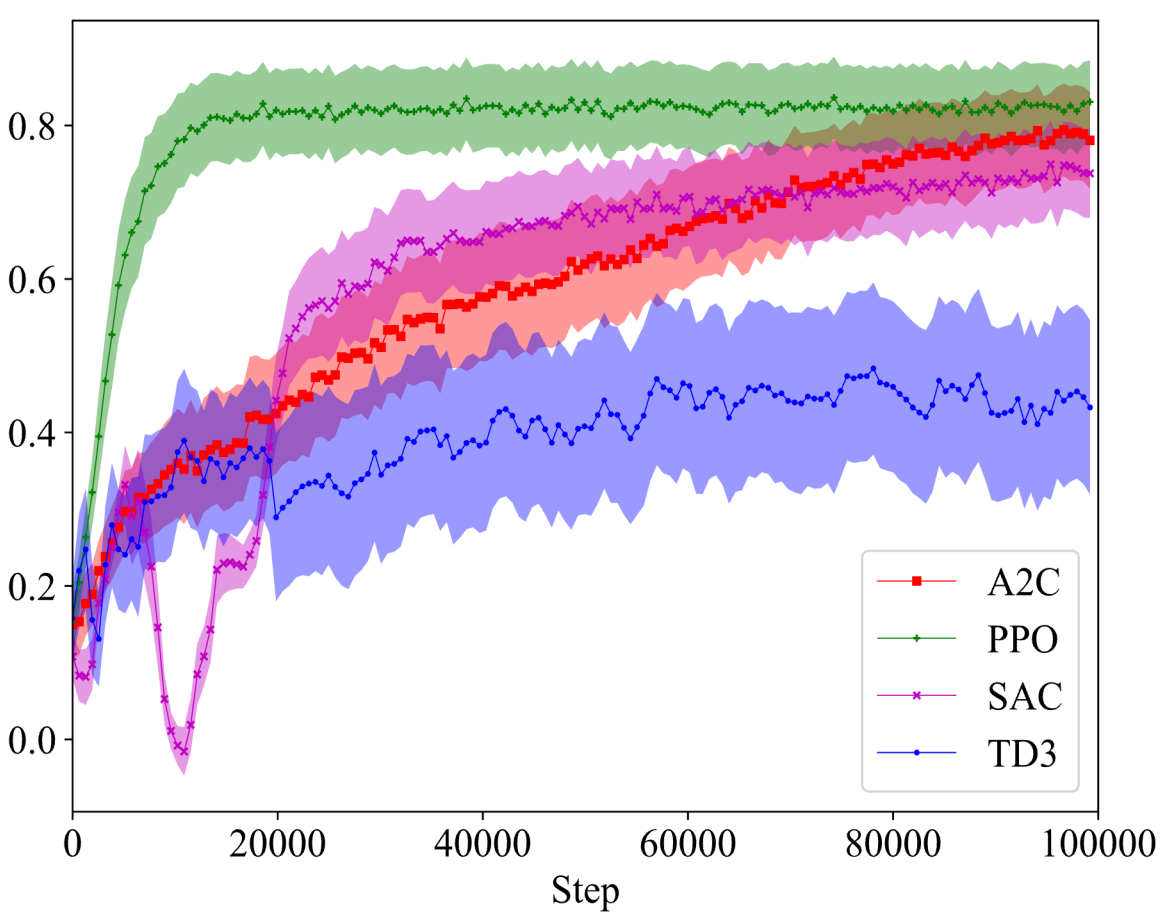


Results

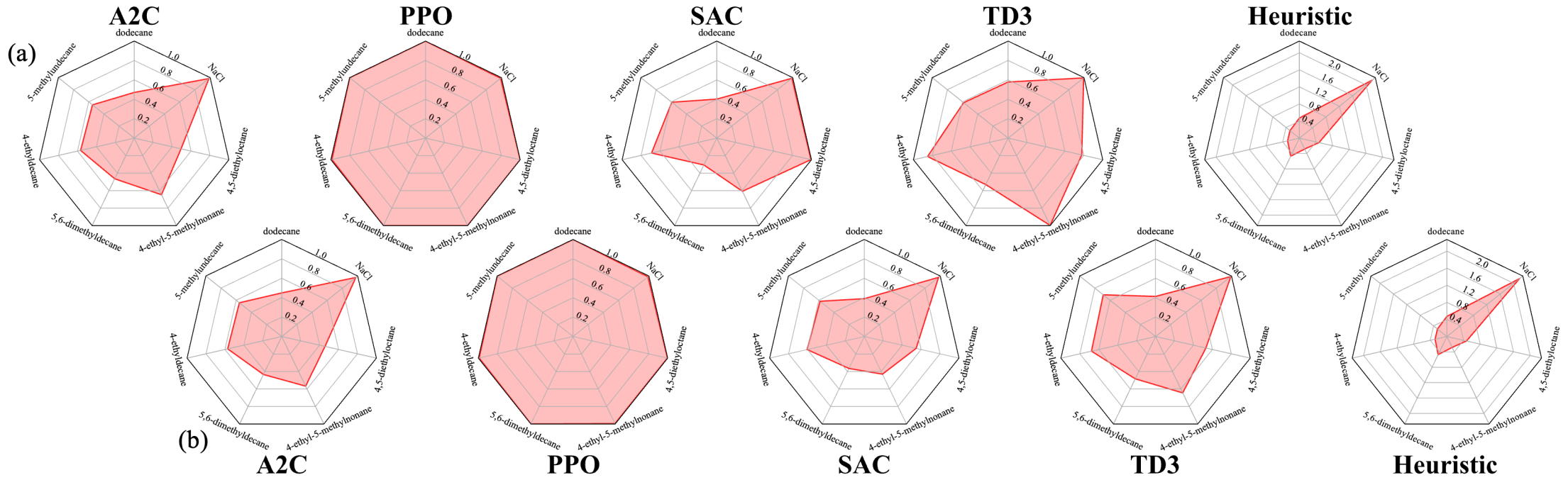
Extraction Bench



Reaction Bench



Results



About Us

ChemGymRL would not be possible without the dedication and years of hard work put in from this group of talented developers. Below, you can find all past and present members involved in making ChemGymRL a reality as well as their contributions to the project.

Contributors

Mark Crowley *- Since time immemorial.*

Principal Investigator

Isaac Tamblyn *- Actually, even before time began. *(¯_(ツ)_/¯ he's a physicist)**
Principal Investigator

Christopher Beeler *- Spring 2018 - Present*
PhD Student - Chemistry Expert

Nicholas Alexander Paquin *- Winter 2021*

Co-op Developer

Mark Baula *- Winter 2021*

Co-op Developer

Amanuel Dawit *- Winter 2021*

Co-op Developer

Sriram Ganapathi Subramanian *- Spring 2018 - Present*
PhD Student - RL Expert

Nouha Chatti *- Spring 2020 - Present*
Master's Student - RL Expert

Mitchell Shahen *- Spring 2020 - Spring 2021*
Co-op and part-time URA Developer

Zihan Yang *- Fall 2019 - Winter 2020*
Co-op and part-time URA Developer

Xinkai Li *- Spring 2019*

Co-op Developer

Kyle Sprague

Human

ChemGymRL

An Interactive Framework for Reinforcement Learning for Digital Chemistry

Mark Crowley

Associate Professor
Department of Electrical and Computer Engineering
University of Waterloo

mcrowley@uwaterloo.ca
https://markcrowley.ca/

@compthink
@compthink@mastodon.social

Chris Beeler, Sriram Ganapathi Subramanian, Kyle Sprague, Nouha Chatti, Mitchell Shahen, Nicholas Paquin, Mark Baula, Amanuel Dawit, Zihan Yang, Xinkai Li,
Mark Crowley, Isaac Tamblyn, Colin Bellinger

VECTOR INSTITUTE

uOttawa

NRC·CNRC

UNIVERSITY OF WATERLOO

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

Try it out yourself!

ChemGymRL

An Interactive Framework for Reinforcement Learning for Digital Chemistry

Join us! - Building Up Through Collaboration

Contribute your own models!
Make new benches, try out existing or new RL algorithms!

chemgymrl.co

UNIVERSITY OF WATERLOO

NRC·CNRC

NSERC CRSNG

uOttawa

VECTOR INSTITUTE

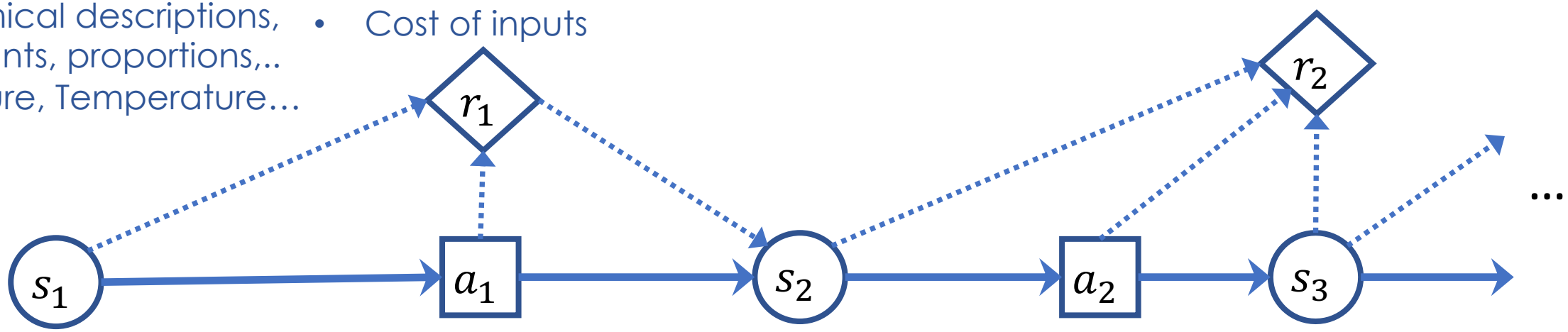
Markov Decision Processes and Reinforcement Learning

Rewards

- Amount of desired material
- Time spent
- Cost of inputs

State of the World

- Chemical descriptions, amounts, proportions,..
- Pressure, Temperature…



Action (Policy)

- **Add/remove** reactant
- **Change** temperature/pressure
- **Choose** which "bench" to use
- **Measure** something about your current samples

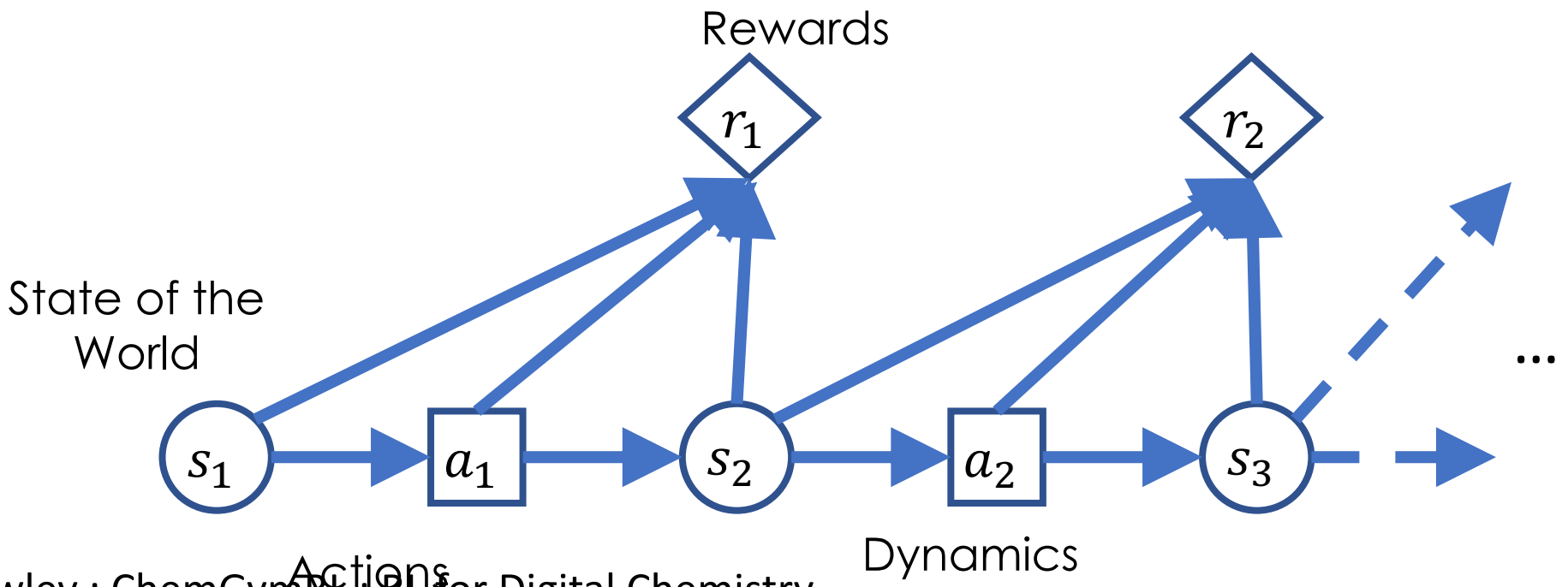
Dynamics (Transition Function)

- *Known via ODEs, but expensive to compute, learn a model instead*
- *Statistical behavior of bench activities:*
  - *extraction amounts for each chemical*
  - *temperature, pressure, etc.*

Markov Decision Process (MDP)

Many different fields of AI arise from how the following questions are answered:

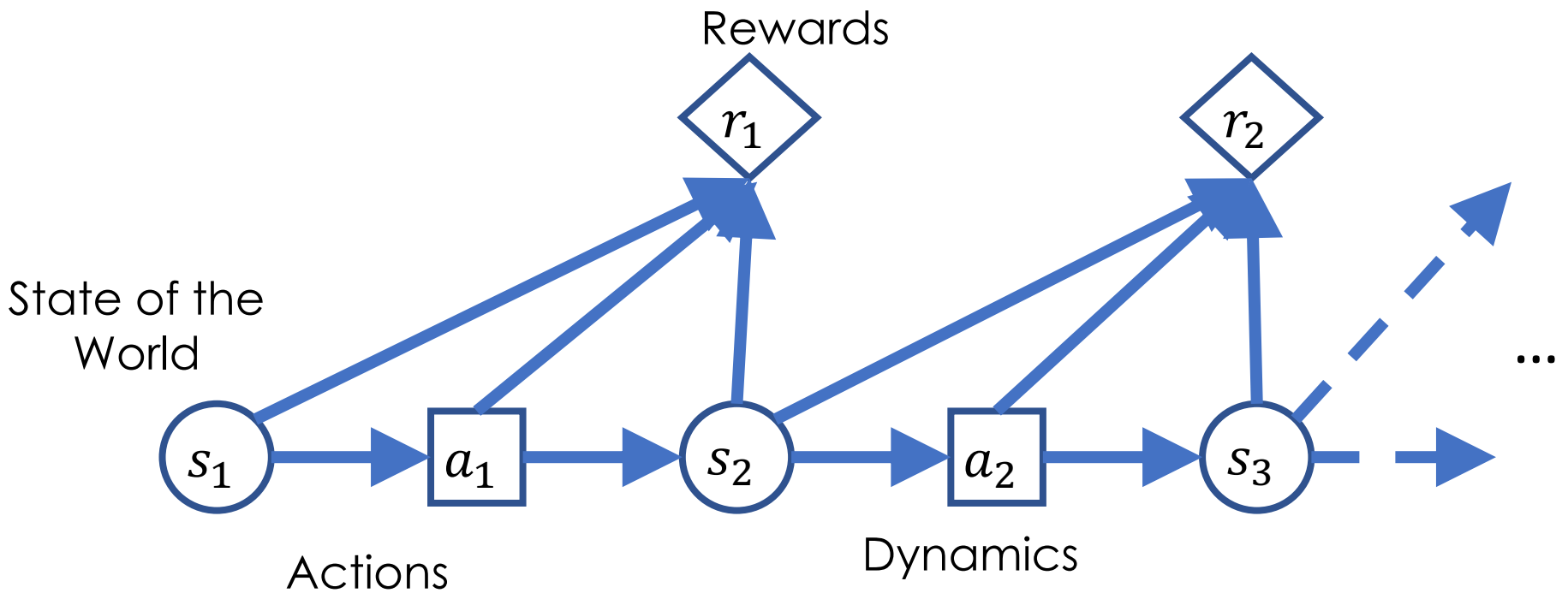
- Which parts of this picture do you know?
- Which can you estimate?
- Which do you *need* to know?



Reinforcement Learning (RL) for Material Design

In this field what we *do* know ahead of time…

- The (simplified) dynamics for basic bench activities
- The immediate costs of each activity, and our distance to the final goal



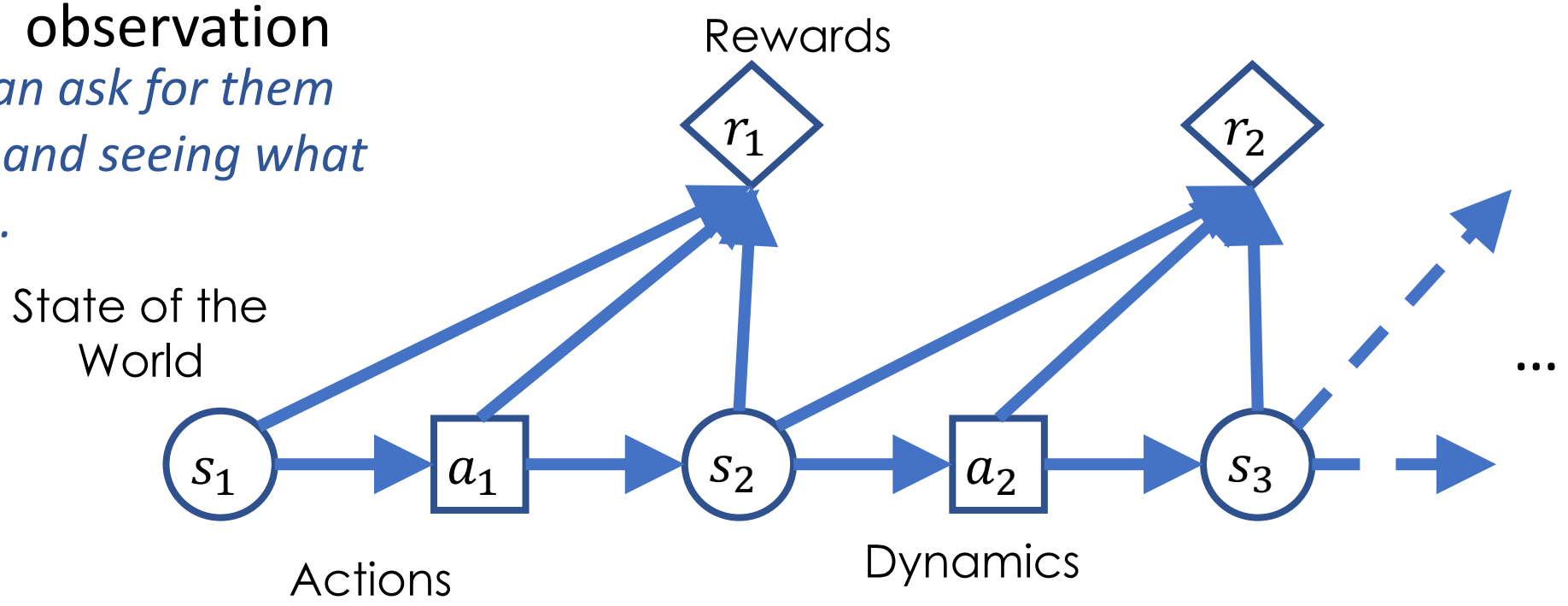
Mark Crowley : ChemGymRL : RL for Digital Chemistry

Reinforcement Learning (RL)

But we *do not* know ahead of time…

- The best (or any) way of stringing together a *series* of transformation activities to achieve a desired material
- The full state of the output of an activity without a destructive observation

But, we can ask for them by acting and seeing what happens…



Reinforcement Learning as a Markov Decision Process

Reinforcement Learning is learning the policy for taking actions for an MDP when you **do not have access** to the full definition of:

- the rewards
- AND/OR the dynamics

Training must be carried out interactively :

1. Commit to action using latest (or some) policy
2. Find out the next state and reward from the world/simulator/environment
3. Improve your policy
4. Repeat until the policy is “good enough” or it stops changing

Reinforcement Learning for Material Design

Reinforcement Learning is learning the policy for taking actions for an MDP when you **do not have access** to the full definition of:

- the rewards
- AND/OR the dynamics

Training must be carried out interactively :

1. Commit to action using latest (or some) policy
2. Find out the next state and reward from the world/simulator/environment
3. Improve your policy
4. Repeat until the policy is “good enough” or it stops changing

The "Physics" of Reinforcement Learning

- RL always comes down to solving a recursive Bellman Equation that **relates the values of states and actions**
- This in many varieties and it **usually solved approximately**
  - MDPs can be solved exactly, but only in small cases with complete knowledge.
- RL algorithms seek to iteratively update a value function, or the policy directly, through experience to make improved decision decisions

Value Iteration

$$V^*(s) = R(s) + max_a \gamma \sum_{s'} P(s'|s,a) V^* s'$$

Policy Gradient

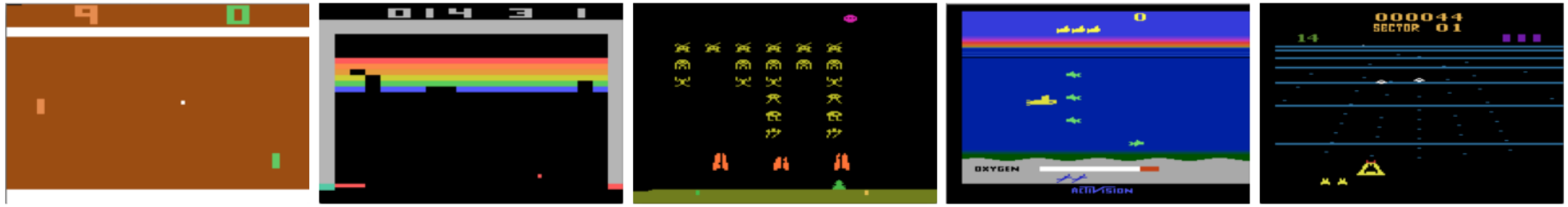
$$\nabla_\theta V^\pi(\mathbf{s}_0) \approx \frac{1}{|K|} \sum_{\mathbf{k} \in K} R(\mathbf{k}) \sum_t \nabla_\theta \log \pi(\mathbf{a}^{\mathbf{k},t} | \mathbf{s}^{\mathbf{k},t}, \theta)$$

Q-learning

$$Q'(s_t, a_t) = (1-\alpha) Q(s_t, a_t) + \alpha (r_t + \gamma \max_a Q(s_{t+1}, a))$$

Mark Crowley : ChemGymRL : RL for Digital Chemistry

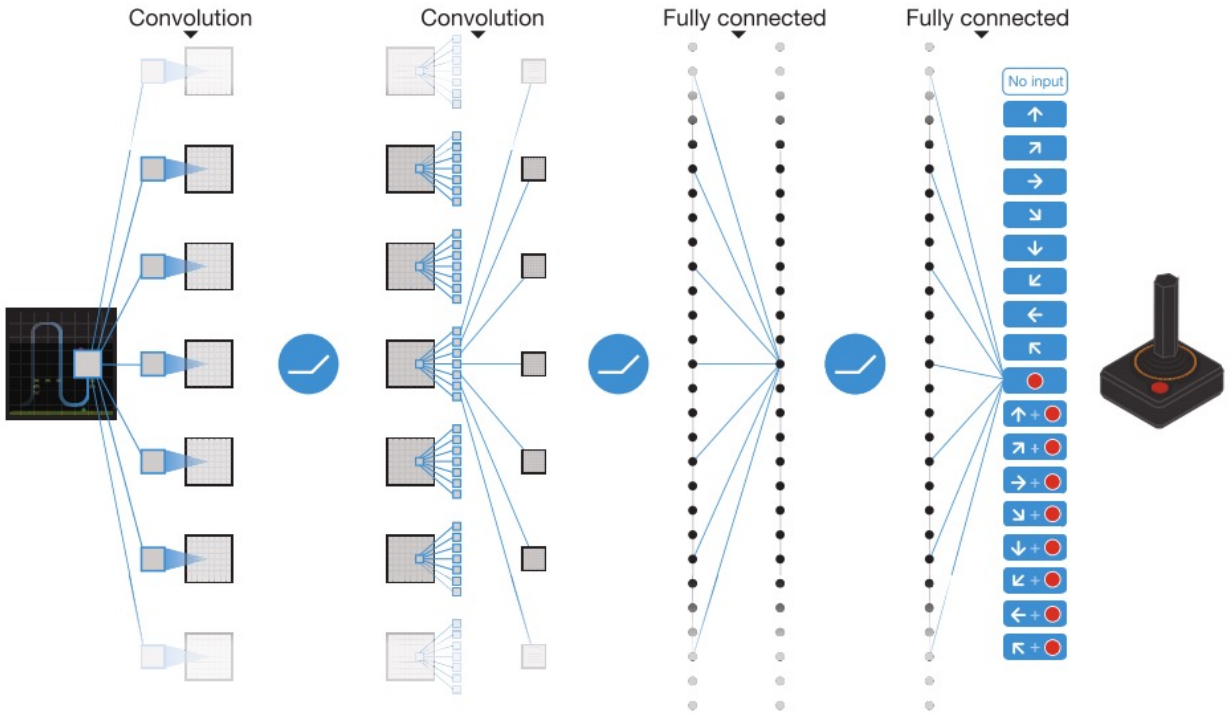
Deep Reinforcement Learning on Atari Games



Flurry of advances since 2014 by Google DeepMind and others applying Deep Learning to RL algorithms.

Many algorithms since then trying to provide a better way to learn the value function with DNNs

- Alpha Go – RL + human training
- Alpha Zero – RL + MCTS search + playing itself (Go, Chess)
- AlphaStar – RL + LSTMS + ? = play StarCraft against human experts

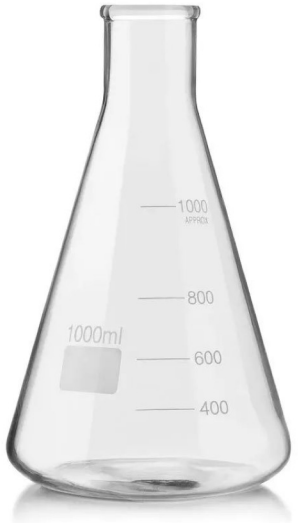


Mark Crowley : ChemGymRL : RL for Digital Chemistry

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

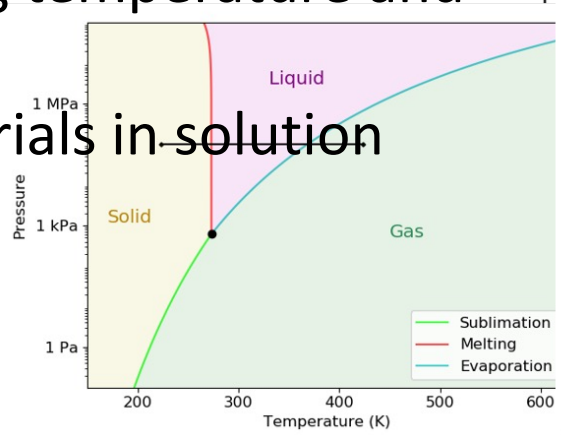
Chemical Laboratory Environment

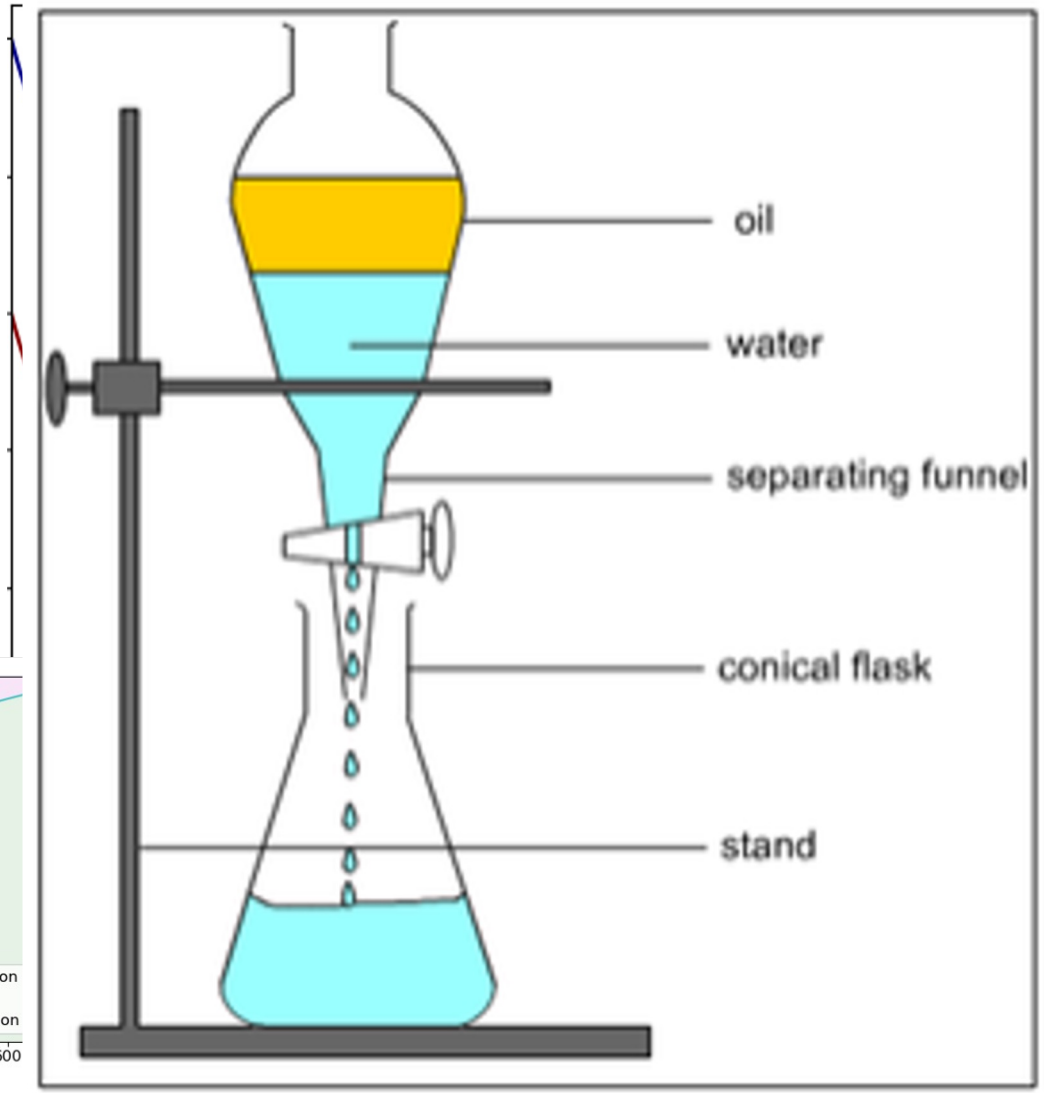
- Environment designed to simulate various aspects of a chemical laboratory.
- Several experimental procedures and observational instruments are available.
- Each experiment is controlled by an individually trained agent.
- A single master agent decides which experiments and measurements are performed and in what order.

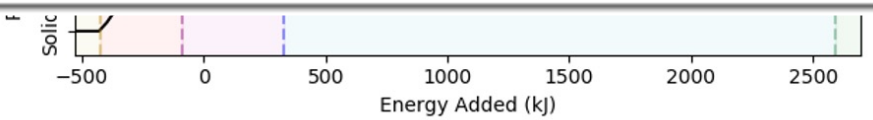


Multi-Stage Exploration

- We approach this by building small, manageable models for component activities in materials design and chemistry
  - (1) Reaction rates of various collections of chemicals at give temperatures and pressures
  - (2) Navigating the phases of matter of a given compound using temperature and pressure changes
  - (3) Extraction of materials in solution via polarizing solvents

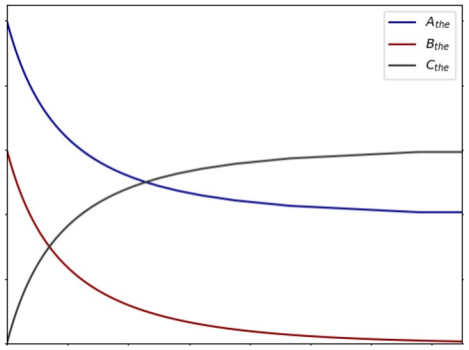


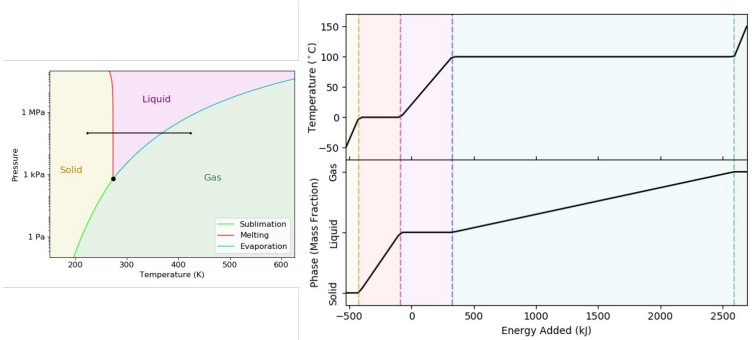


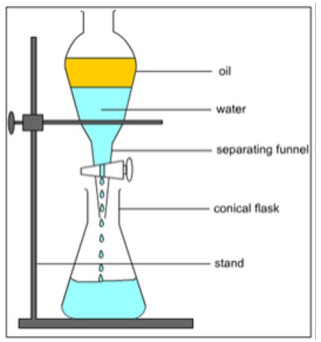


Mark Crowley : ChemGymRL : RL for Digital Chemistry

Multi-Stage Exploration







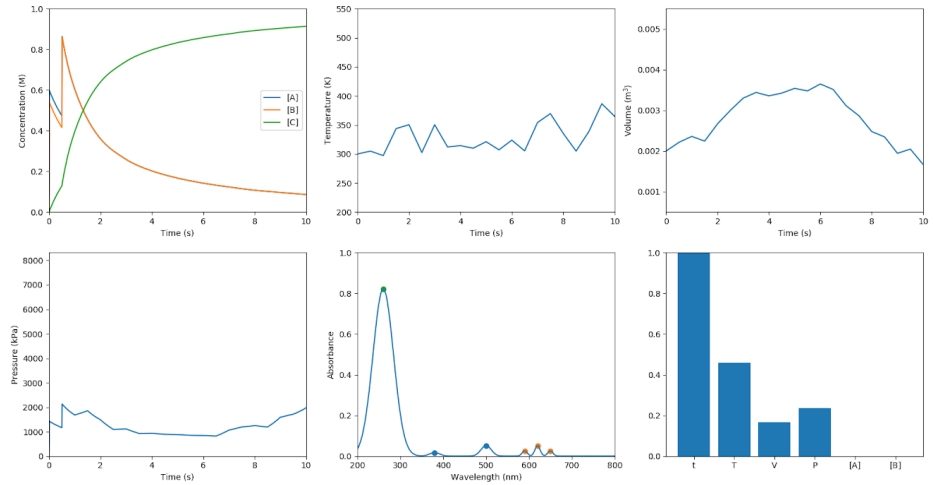
Mark Crowley : ChemGymRL : RL for Digital Chemistry

WATERLOO.AI
WATERLOO ARTIFICIAL INTELLIGENCE INSTITUTE

Chemistry Lab

Process Actions

- (1) ODE-World
- (2) Phase-World
- (3) Extraction-World



Observation Actions

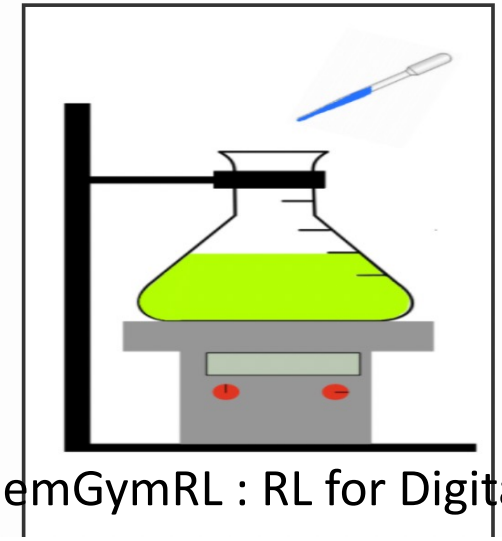
- Mass Spectrometry*
- Nuclear Magnetic Resonance*
- Gas Chromatography*
- High Performance Liquid Chromatography*
- UV-Vis Spectrometry*
- Fluorescence Spectrometry*

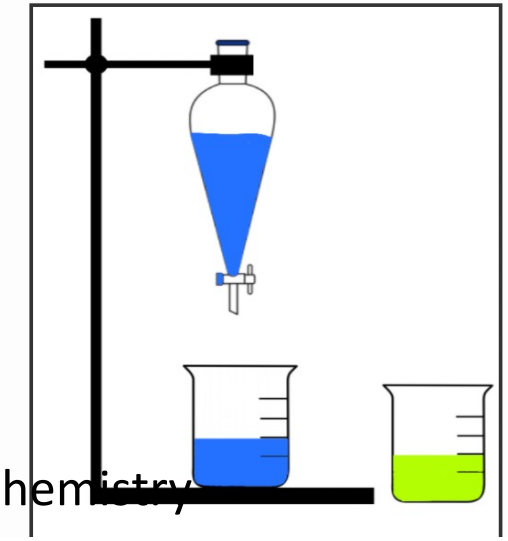
https://ChemGymRL.com

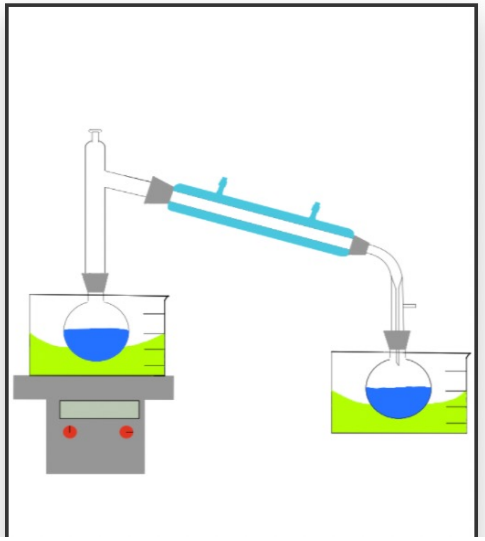
ChemGymRL Home Benches Documentation About Us

ChemGymRL Benches

The environments are made to simulate a chemistry lab with various stations for researchers to carry out experiments, and shown below are the experiments that currently exist within `ChemGymRL`, with more to be added soon. Every bench can be used as a separate experiment or you can also connect multiple benches together using the lab manager. Make sure to check out our documentation here for more information about the Lab Manager or to learn more about the benches and what they have to offer!

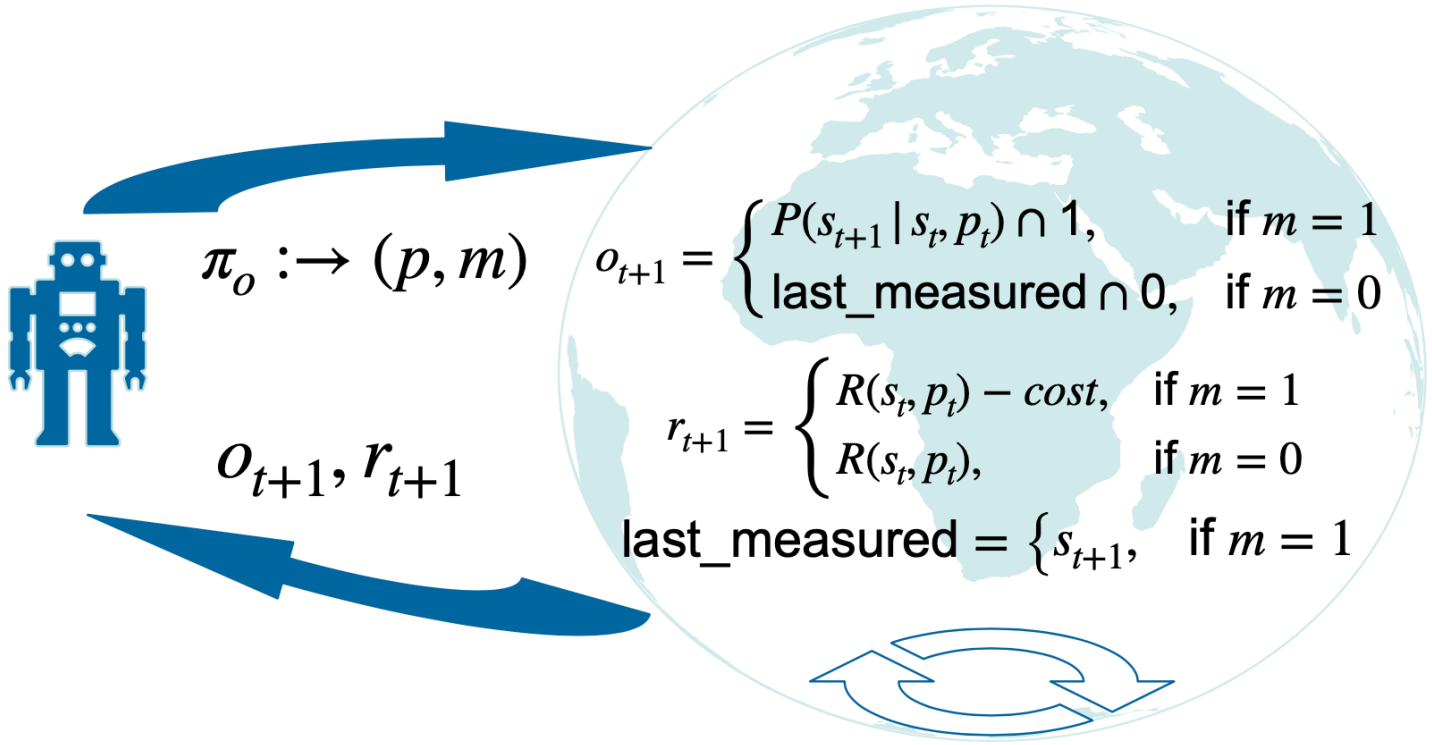






Proposed RL Framework

With Explicit Measurement Actions and Costs



Bellinger, Drozdyuk, Crowley, and Tamblyn.
"Balancing Information with Observation Costs in Deep Reinforcement Learning"
In Canadian Conference on Artificial Intelligence 2022