UNIVERSITY OF TORONTO

2023 International Planning Competition

Probabilistic & Reinforcement Learning Track

Ayal Taitler and Scott Sanner

University of Toronto

ICAPS

July 11th, 2023

Motivation & Goals

Sometimes doing just planning/learning is not enough

- Large spaces
- Complex changing topology
- Hybrid spaces

Promote combined planning & learning methods

- Compete with others to develop SOTA RL & planning methods
- Address real world problem the community cares about

Focus

Real-world problems

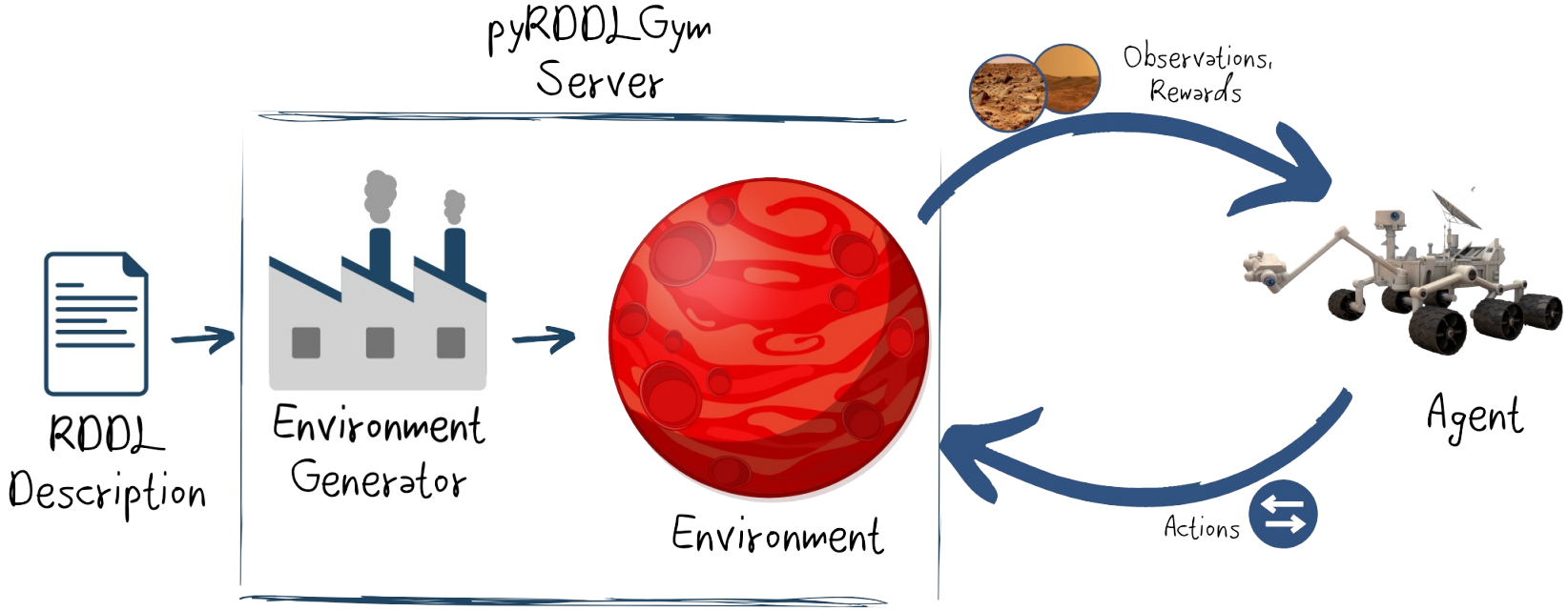
– industrial & environmental

Featuring:

- Noisy
  - Exogenous state noise
  - Stochastic action effect
- Continuous and* discrete state and action spaces
- Strongly structured problems
- Concurrent event

* Not at the same time

Infrastructure



RDDL → compiler → Gym environment

- Standard Gym interface and spaces
- Full access to the underlying model
- Differentiable dynamics*

Setting & Procedure (I)

4 days of competition (online phase)

- **8** problems
- **3** instances per problems in increasing order of difficulty
  - IDs: 1,3,5
- Submission at the end of the 4th day
  - Self contained Docker of the method (prereleased template)

Setting & Procedure (II)

Evaluation

- **8** problems
- **5** instances per problems in increasing order of difficulty
  - IDs: 1,2,3,4,5 – two new unseen instances
- **50** trials per instance (averaged)

- **Time allocation:**
  - **60** “offline” minutes
  - **4** minutes per trial

Setting & Procedure (III)

- **Submission**:
  - Docker is rebuilt on the cluster and exeuted via Singularity
  - pyRDDLGym mapped externally (”local” interaction)
- **Resource allocation**:
  - 32GB RAM
  - 8 CPU cores
- **Scoring:**
  - 0: max(NoOp, Random)
  - 1: Max({methods})

Timeline & Participation

Event	Date
Infrastructure release with sample domains	October, 2022
Call for domains and praticipants	October, 2022
Final domains announcement	March 2, 2023
Competitors registration deadline	March 15, 2023
Dry-run	April 17, 2023
Planner abstract dubmission	May 1, 2023
Contest run	June 5-8, 2023
Final planner abstract dubmission	June 13, 2023
Results announced	July 11, 2023

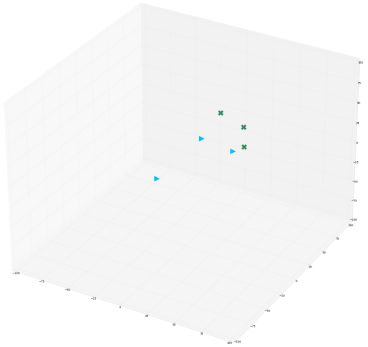
29 Competitors registered

4 Competitors Submitted dockers

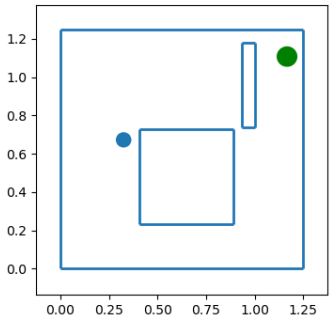
1 Competitor Sucessfully submitted A working method

All The RL folks dropped out 😢

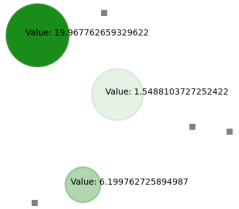
Problems



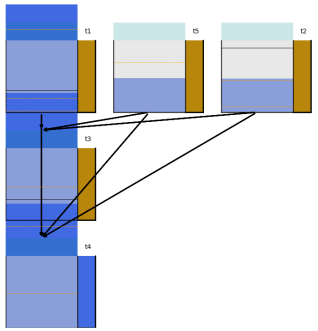
UAV



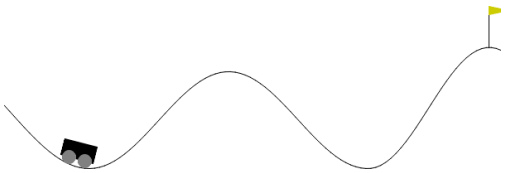
Race Car



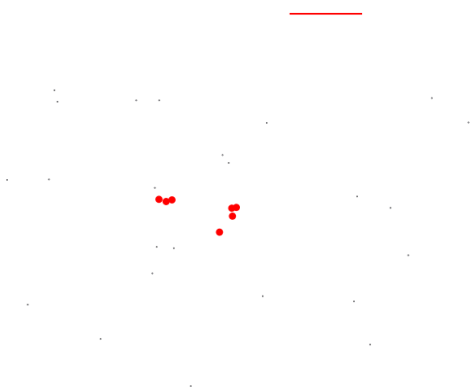
Mars Rover



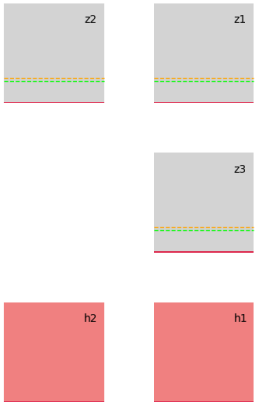
Reservoir Control



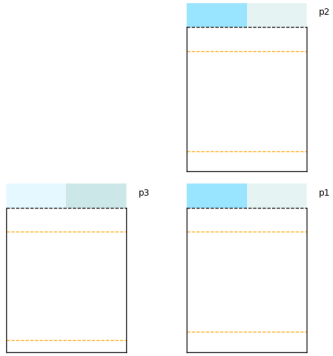
Mountain Car



Recommender System

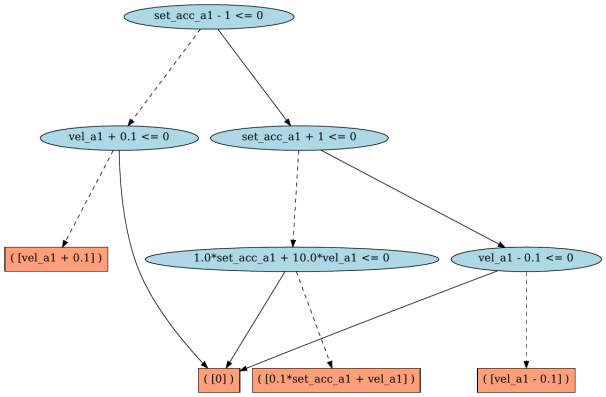


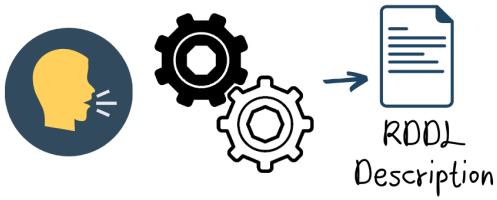
HVAC

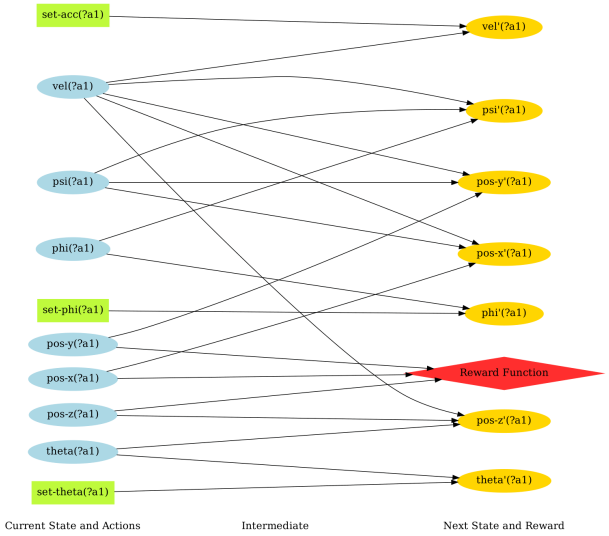


Power Generation

Starter Kit







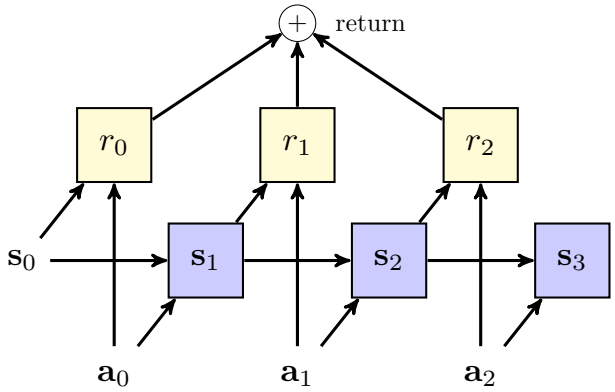
XADD generator

Instance generators

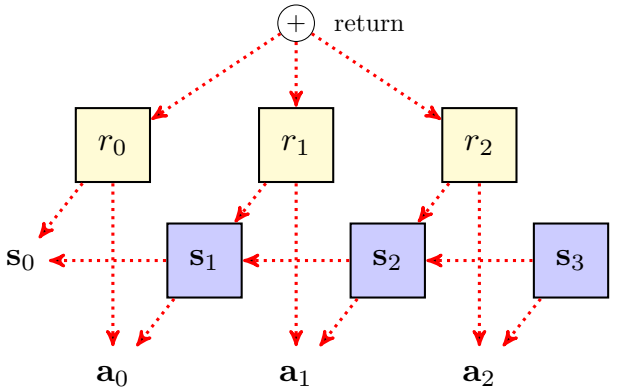
DBN visualizer

JaxPlanner

Simulate: Given plan $a_0, a_1, \ldots,$ simulate states s_t and reward r_t

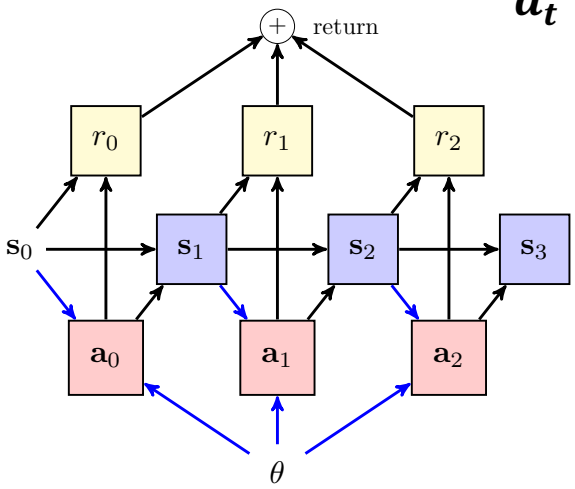


Optimize: Adjust a_t based on the return gradient



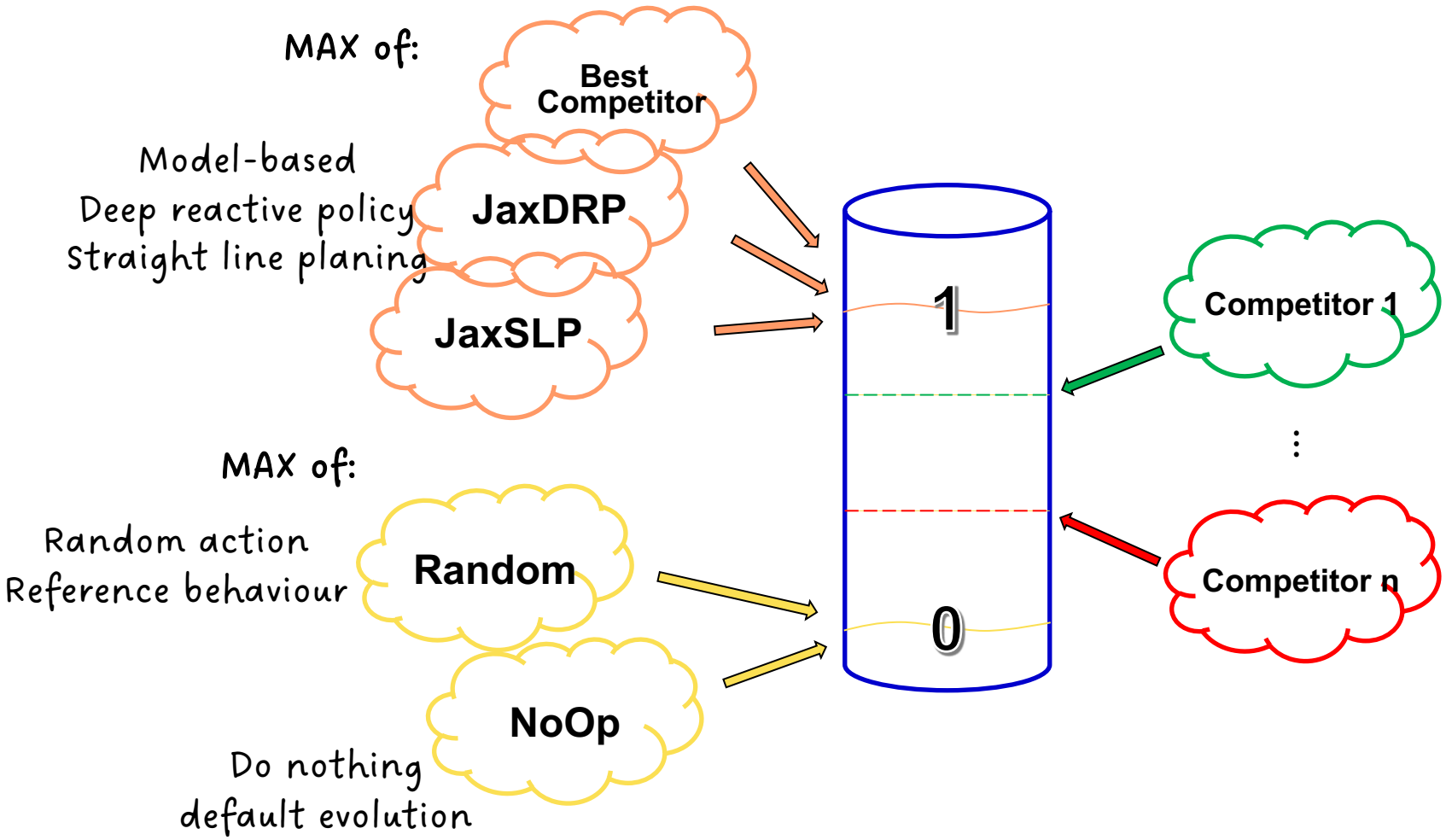
$$a'_t = a_t + \eta \nabla_{a_t} \sum_{\tau} r_\tau$$

Closed-loop plan:
Deep reactive policy



Baselines

Not competitors, Normalizers



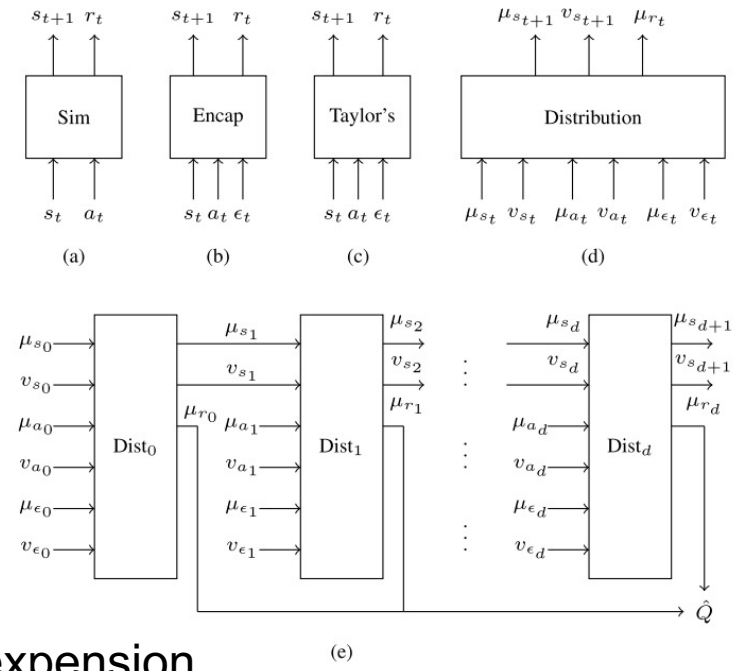
And the winner is…

DiSProd[1]

Palash Chatterjee, Ashutosh Chapagain, Roni Khardon
Indiana University, Bloomington

Handling stochasiticy:

- Encapsulating the stochasiticy with a reparametrization trick
- Dist. are represented with mean and variance,
  → can be propageted



Transition dynamics:

- 2nd order Taylor expansion approximation
- Noise parameters are propageted through the expension
- Stacking one step propagation boxes yields computation graph

[1]Based on the paper: Palash Chatterjee , Ashutosh Chapagain , Weizhe Chen and Roni, "DiSProD: Differentiable Symbolic Propagation of Distributions for Planning", IJCAI, 2023

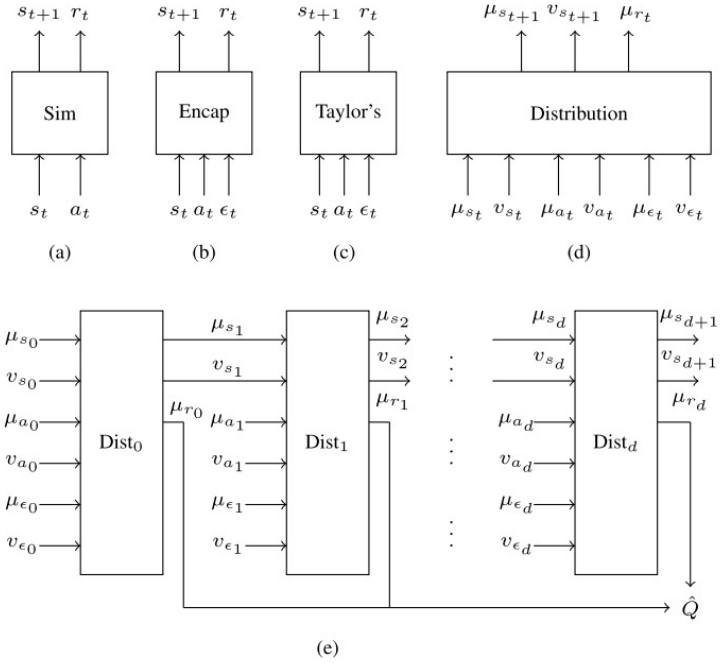
And the winner is…

DiSProd[1]

Palash Chatterjee, Ashutosh Chapagain, Roni Khardon
Indiana University, Bloomington

Search:

- Gradient search finds the parameters of the open loop policy → replanning scheme
- Algorithm integrates multiple gradient searches in “parallel”



The IPPC variant:

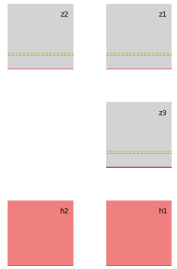
- Running the full Search is too expesive
- Two variants:
  - DiSProd-NV – ignores variance
  - DiDProd-S – samples transitions
- Action space reduction – multi-valued cont. action vector vs of binary action vector

Notations

- **NoOp** – *do nothing (default evolution)*
- **Random** – *random action sampling*
- **JaxSLP** – *open loop backprop planner*
- **JaxDRP** – *deep reactive policy*

Results
HVAC

- *Difficulty*: number of zones, number of heaters, connectivity



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	0.779	**1**	0.932	0.974	0.957
JaxDRP	0.822	0.904	0.847	0.972	0.576
DiSProD	**1**	0.974	**1**	**1**	**1**

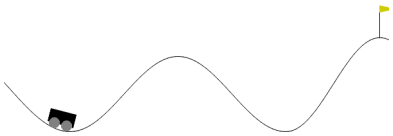
Results
Mars Rover

- *Difficulty*: number of agents vs number of rewards

	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	0	0	0	0	0
JaxDRP	0	**1**	**1**	**1**	**1**
DiSProD	**1**	0	0.528	0.189	0.203

Results
Mountain Car

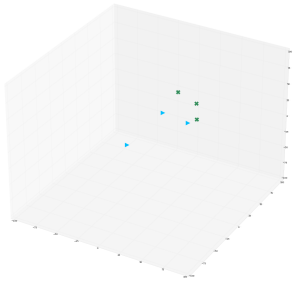
- *Difficulty*: number of hills



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	0.76	**1**	**1**	**1**	**1**
JaxDRP	0.24	**1**	0	0	0
DiSProD	**1**	**1**	0.86	**1**	0

Results
UAV

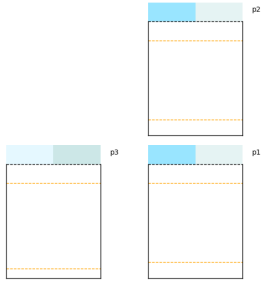
- *Difficulty*: number of UAVs
  (where some might be uncontrollable)



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	0.401	0.529	0.492	0.454	0.450
JaxDRP	0.270	0.272	0.124	0.233	0.216
DiSProD	**1**	**1**	**1**	**1**	**1**

Results
Power Generation

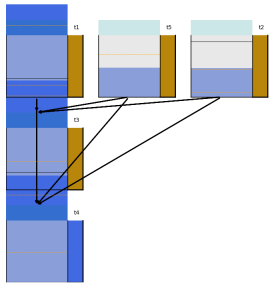
- *Difficulty*: nubmer of generators



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	**1**	0.871	0.970	0.832	0.213
JaxDRP	0.942	**1**	0.871	0.892	**1**
DiSProD	0.975	0.636	**1**	**1**	0.674

Results
Reservoir Control

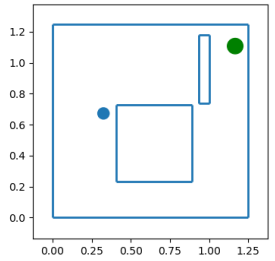
- Difficulty: number of reservoirs



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	**1**	**1**	**1**	0.970	0.956
JaxDRP	0.920	0.923	0.820	**1**	**1**
DiSProD	0.927	0.99	0.808	0.932	0.81

Results
Race Car

- *Difficulty*: number and location of obstacles



	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	**1**	**1**	**1**	**1**	**1**
JaxDRP	0	0.443	0.068	0	0
DiSProD	0	0	0	0	0

Results
Recommender System

- *Difficulty*: number of consumers, number of suppliers

	Instance 1	Instance 2	Instance 3	Instance 4	Instance 5
NoOp	0	0	0	0	0
Random	0	0	0	0	0
JaxSLP	**1**	0.672	0	0	0
JaxDRP	0.929	**1**	0	0	0
DiSProD	0.153	0.054	**1**	**1**	1

Results

Final leaderboard

	HVAC	Mars Rover	Mount. Car	UAV	Power Gen.	Reser. Cont.	Race Car	Rec Sim	Total
NoOp	0	0	0	0	0	0	0	0	0
Random	0	0	0	0	0	0	0	0	0
JaxSLP	1	0	**4***	0	1	**3**	**5**	1	15
JaxDRP	0	**4**	1*	0	**2**	2	0	1	10
DiSProD	**4**	1	3*	**5**	**2**	0	0	**3**	**18**

Results

Generalizability

-DiSPRoD

	HVAC	Mars Rover	Mount. Car	UAV	Power Gen.	Reser. Cont.	Race Car	Rec Sim	Total wins
Seen	3	1	1	3	1	0	0	2	11/24 (~45%)
Unseen	1	0	2	2	1	0	0	1	7/16 (~53%)

Some Conclusions

No "survivors" among the RL competitors

- **Var(Problems)** – robotics, navigation, network control, recsim
- **Var(Properties)** – Continuous/discrete, control, network control, Recsim, goal oriented, oversubscriptive…
- **Concurrent actions**
- **Sizes** – state/action spaces change between instances
- **Non-fluents** – constants/bounds/structure → *state aug.?*