Explainable and Differential Reinforcement Learning
for Multi-objective optimization
in Particle Accelerators

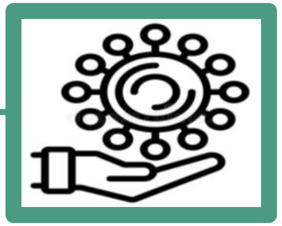
Kishansingh Rajput
Data Science Department, Jefferson Lab

This research was supported by the U.S. Department of Energy, through the “Data-Driven Decision Control for Complex Systems (DnC2S)” project under the Advanced Scientific Computing Research program and the ”Next Generation Compact SRF Accelerators for Industrial Applications” project under the Nuclear Physics program. Contributions from Auralee Edelen and Ryan Roussel were supported by the Department of Energy, Laboratory Directed Research and Development program at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515.

OUTLINE

01

Background and Problem Formulation

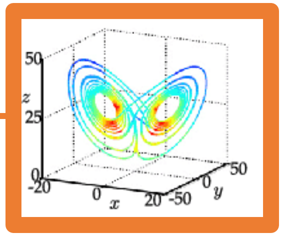


02

Methods

03

Differential RL and Results



04

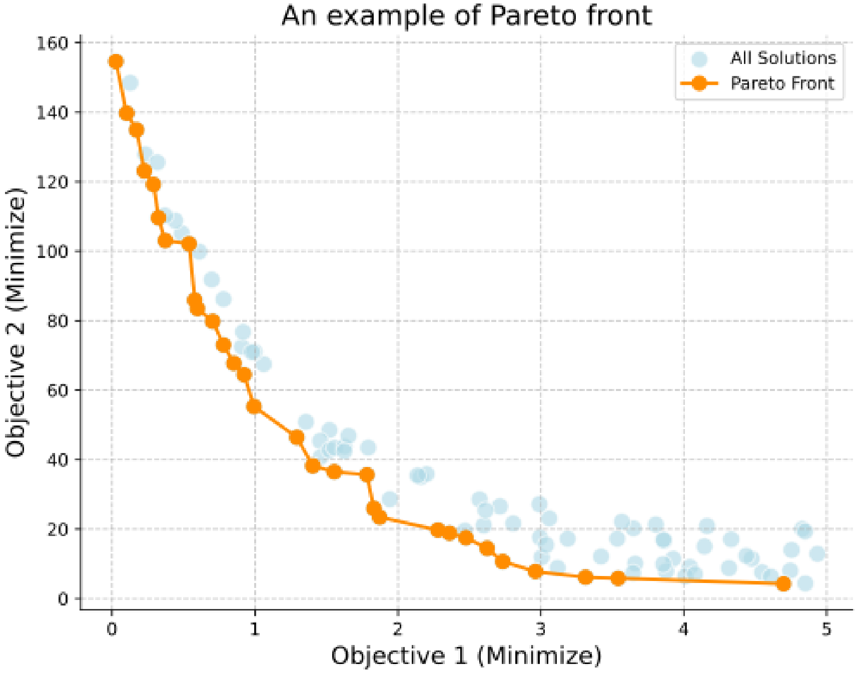
Explainable RL with SINDy and Results

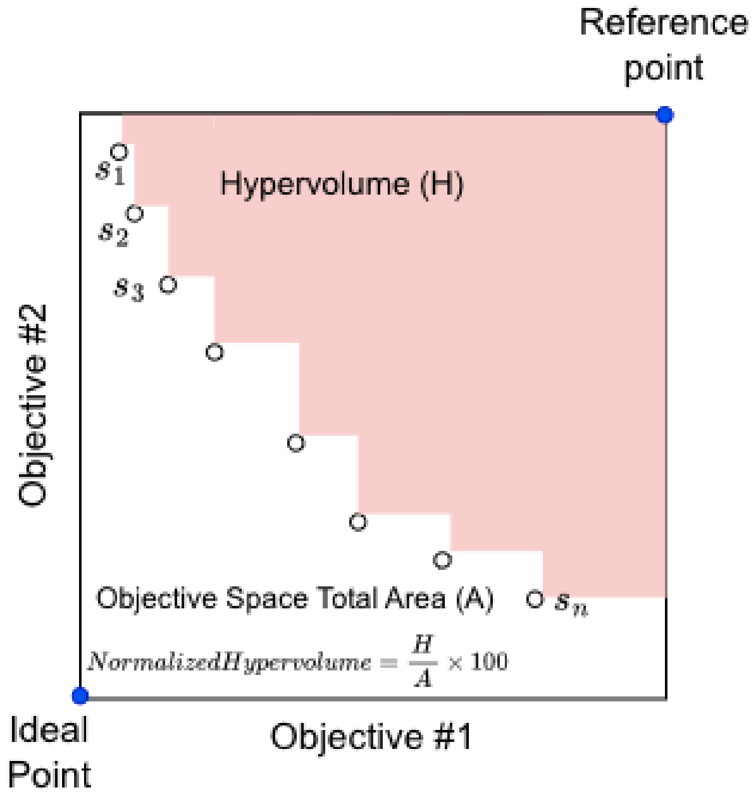
05

Summary and References

MULTI-OBJECTIVE OPTIMIZATION: BALANCING THE TRADE-OFFS

- Simultaneously optimizing two or more objectives (usually conflicting)
- Find set of solutions where no solution dominates another; Pareto front
- In Multi objective constrained problems – avoid any solutions that violates the constraints

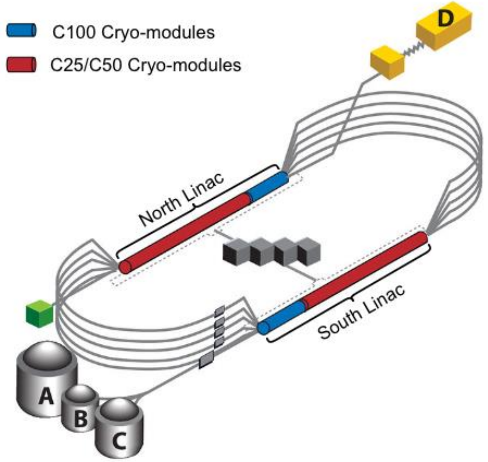




Jefferson Lab

MULTI-OBJECTIVE OPTIMIZATION FOR CEBAF

- Continuous Electron Beam Accelerator Facility uses super conducting RF cavities
- The facility uses total 25 cryo-modules with roughly 200 cavities in each linac
- Challenge: Distribution of gradients to simultaneously optimize both heat load on cryo-system and number of trips while maintaining required beam energy



Heat load of cavity i $$H_i = \frac{G_i^2 l_i}{\omega_i Q_i(G_i)}$$

G_i is gradient setting on cavity i
l_i is the length of cavity i
ω_i is the shunt impedance of cavity i
$Q_i(G_i)$ is the quality curve of cavity i

$$E_{linac} = \sum_{i=1}^{n} G_i l_i$$

$$a_i \leq G_i \leq b_i$$

Individual cavity gradient is bounded

Multi-Objective Optimization Problem

$$\text{Minimize} \quad H(\mathbf{G}) = \sum_{i=1}^{N_c} H_i(G_i), \; T(\mathbf{G}) = \sum_{i=1}^{N_c} T_i(G_i)$$

$$\text{Subject to } |E_{linac} - \sum_{i=1}^{N_c} G_i l_i| < \delta E,$$

$$a_i \leq G_i \leq b_i.$$

Fast Shut Down (FSD) Trips

Trip rate of cavity i $$T_i = e^{A + B_i(G_i - F_i)}$$

A, B, F_i are coefficients from historical data
G_i is gradient setting on cavity i

SIMULATION / ENVIRONMENT

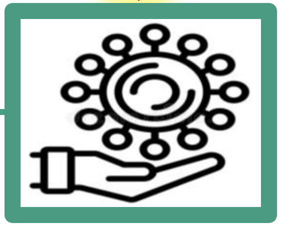
- Simulated environment is built in python using the ideal equations
- Devised four environments to gradually increase the problem complexity
- Gymnasium standards for plug and play
- TensorFlow implementation for auto-gradient calculation and back-propagation

Env	Number of Cavities	$E_{linac} \pm \delta E$ (MeV)	Hypervolume	
			Ref (H, T)	Ideal (H, T)
8D	8	20.08 ± 0.40	(22.4, 0.05)	(20.9, 0.015)
16D	16	50.00 ± 0.60	(100.0, 0.40)	(88.0, 0.015)
32D	32	120.00 ± 0.80	(290.0, 0.40)	(262.0, 0.010)
North linac	200	1050.00 ± 2.00	(2530.0, 6.00)	(2380.0, 1.000)

OUTLINE

01

Background and Problem Formulation

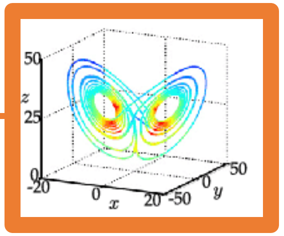


02

Methods

03

Differential RL and Results



04

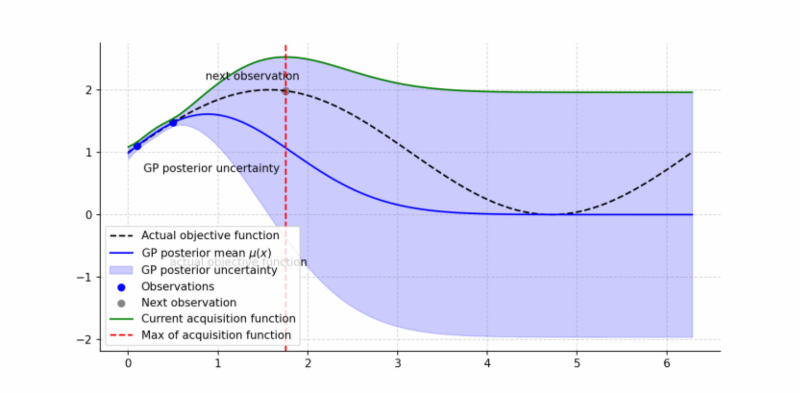
Explainable RL with SINDy and Results

05

Summary and References

TRADITIONAL (OTHER) METHODS

Multi-Objective Bayesian Optimization (MOBO)

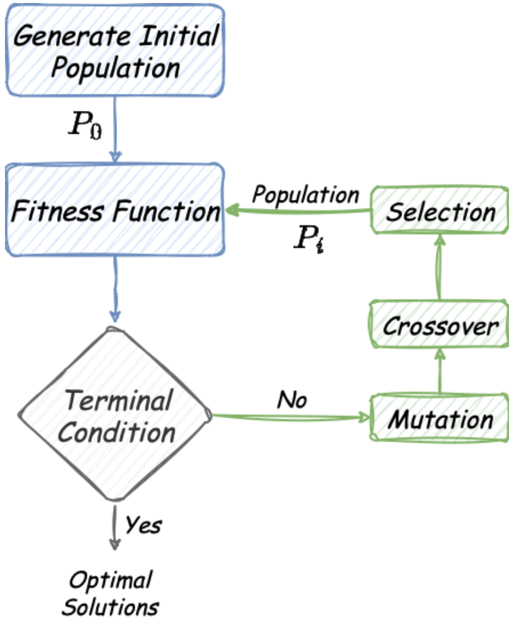


- BO is sample efficient but slow -- in traditional form O(n^3)
- Scaling can be improved with correlated kernels, physics informed kernels or BNN instead of GP
- More suitable with expensive simulations

MOBO acquisition function
expected hypervolume improvement (EHVI)

Constraints modeled as separate GP and used in acquisition function

Multi-Objective Genetic Algorithm (MOGA)



- Mimics natural selection and evolution
- Works with a group of potential solutions – individuals in a population
- Requires running fitness function (simulation) for each individual in population at each step!
- Can not leverage gradient propagation through simulations.

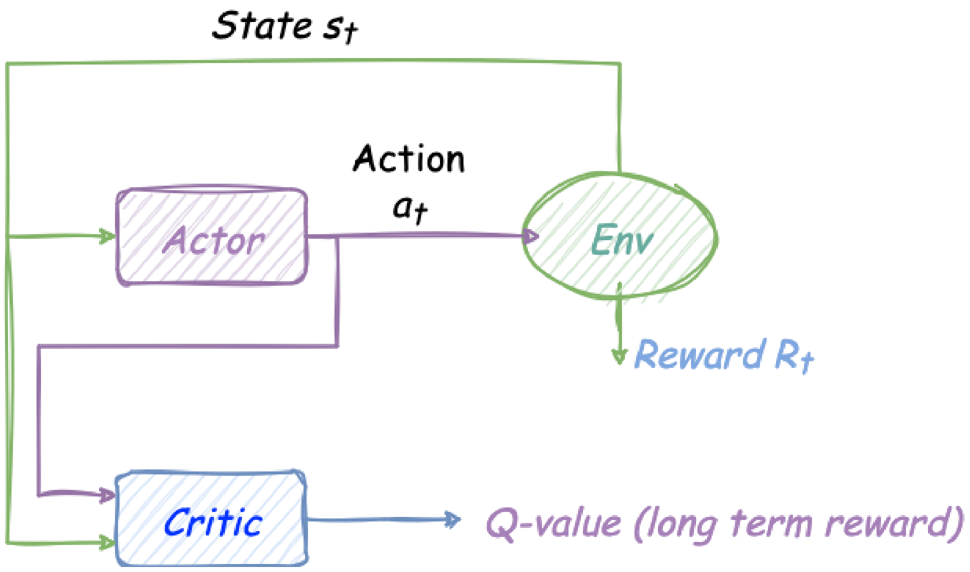
MOGA fitness function

$$[H(G), T(G)]$$

Remove any solution that violate constraints in fitness

Jefferson Lab

REINFORCEMENT LEARNING



- RL is a controls algorithm by design
- It iteratively interacts with the environment to learn
- Balances between exploration and exploitation to learn global optimal
- SOCT: Scientific Optimization and Control Toolkit; a robust RL toolkit designed with a focus on scientific applications (actively maintained; addition of new recent algorithms)[3]

- In its traditional form deep RL is not sample efficient
- Requires many interactions with the simulation
- More suitable with fast and less expensive simulations

MORL Reward Function

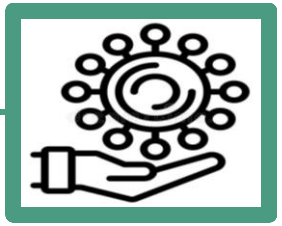
$$R_h = -1 \times \sum_{i=0}^{N} H_i + P \quad and \quad R_t = -1 \times \sum_{i=0}^{N} T_i + P$$

$$P = \begin{cases} -5 \times |E - E_{min}| & \text{if } E < E_{min} \\ -5 \times |E - E_{max}| & \text{if } E > E_{max} \\ 0 & \text{Otherwise} \end{cases}$$

OUTLINE

01

Background and Problem Formulation

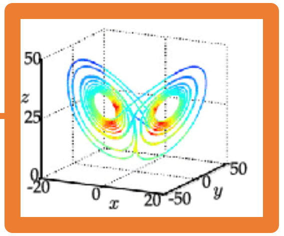


02

Methods

03

Differential RL and Results



04

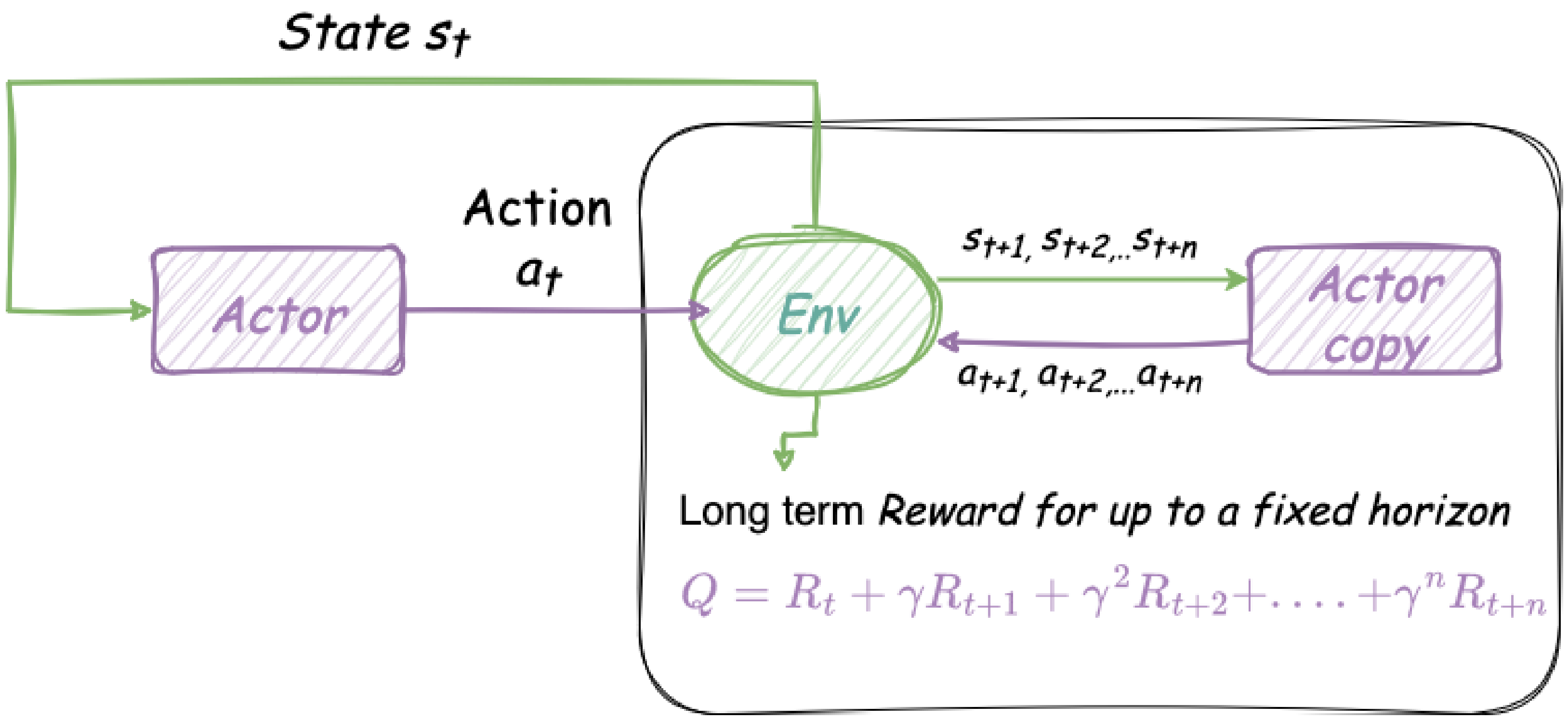
Explainabl
e RL with
SINDy and
Results

05

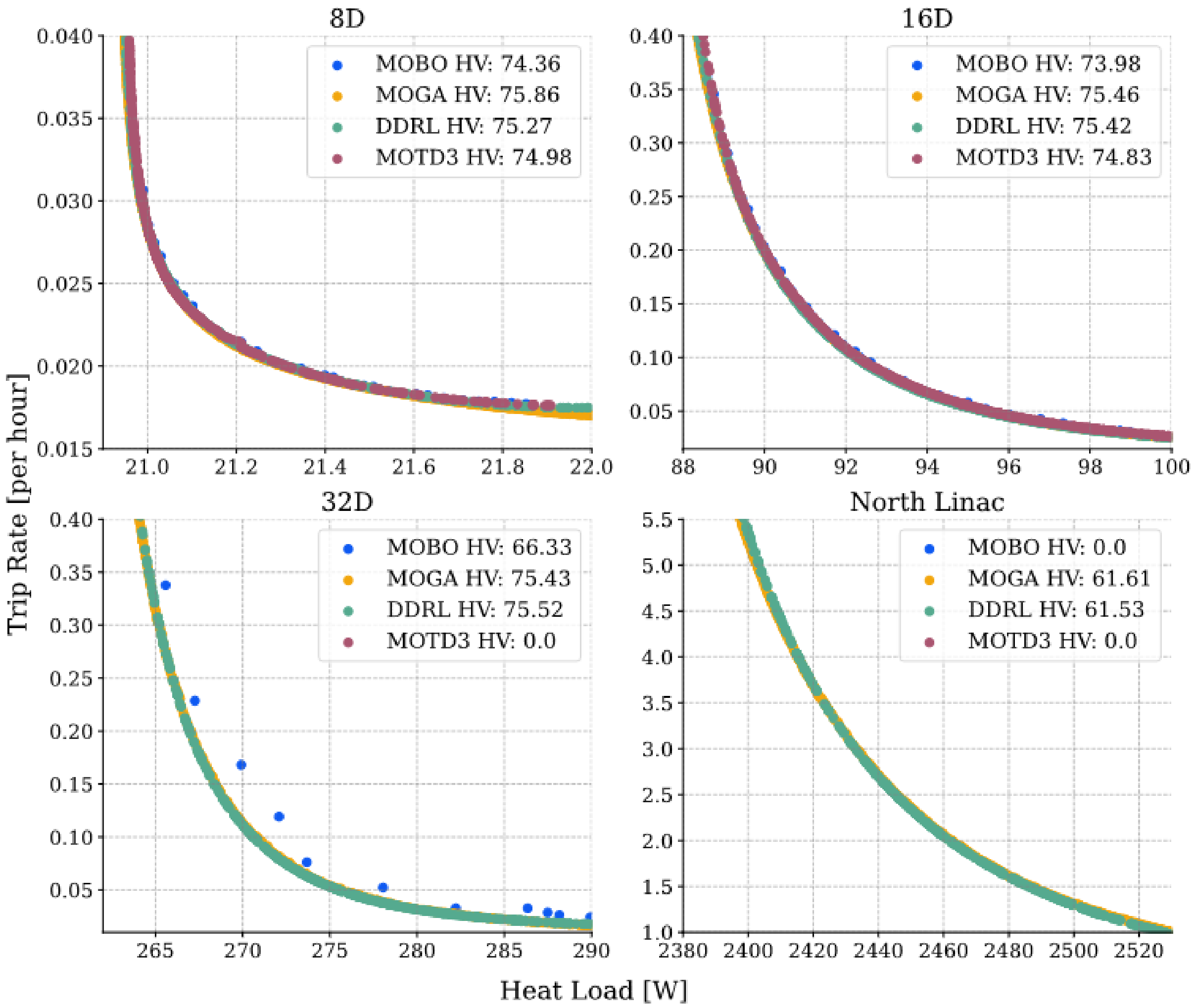
Summary and References

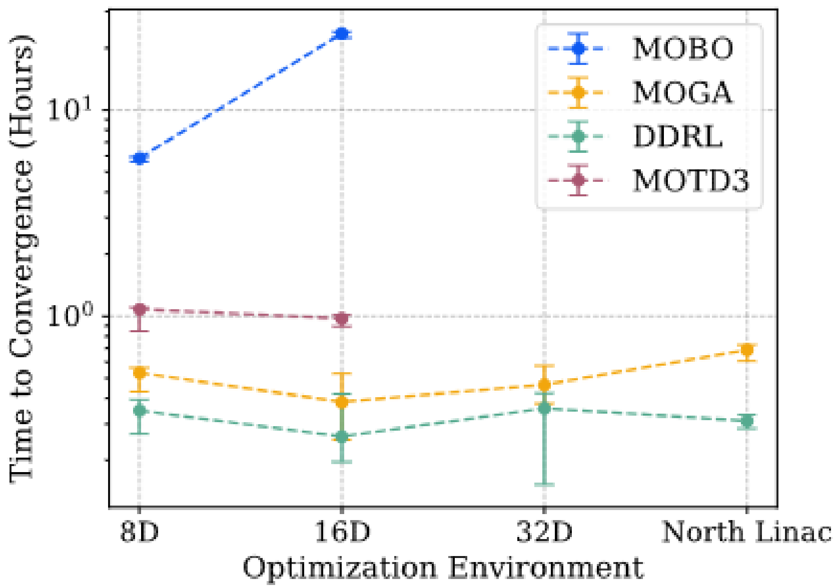
DEEP DIFFERENTIAL RL

- Deep Differential Reinforcement Learning (DDRL) requires **differentiable environment**
- Learns quickly by leveraging back-propagation via environment



RESULTS



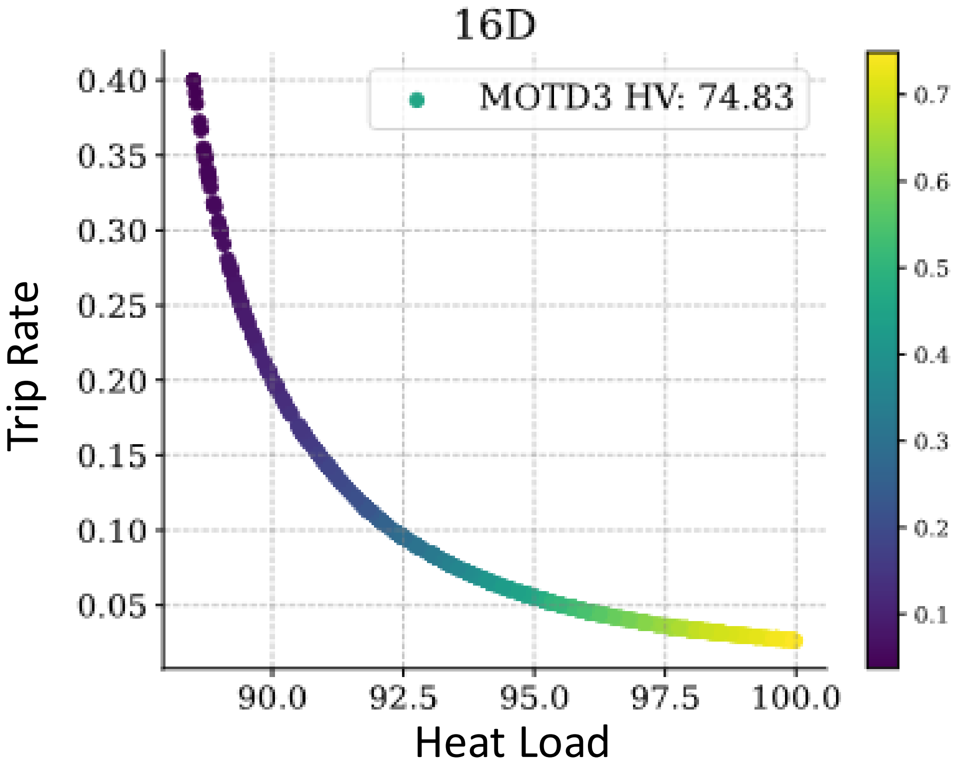


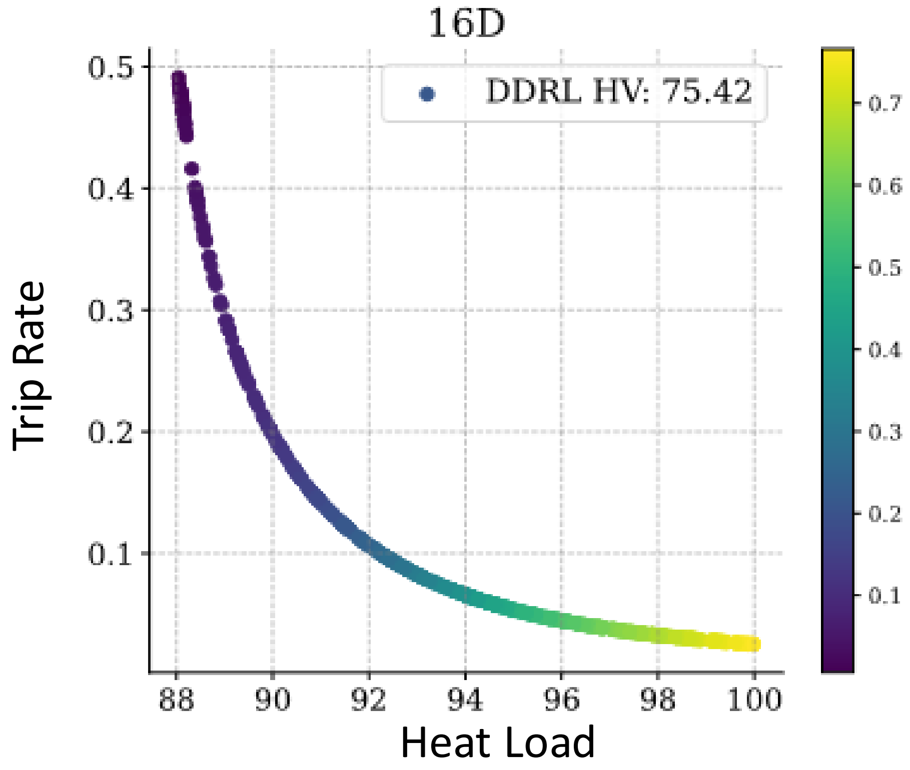
- DDRL and MOGA are faster to converge in time
  - Takes significantly more simulation runs
  - Best with faster (and differential) simulations

- MOBO is best in being sample efficient on low dimensional problems or with expensive simulations

RESULTS

- Conditional input in RL provides control over solutions in pareto front
- Tunable parameters for operators to choose optimal point on a Pareto front given current requirements



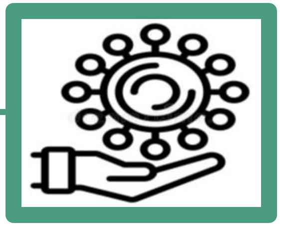


- DDRL and MOGA are faster to converge in time → Best with faster (and differential) simulations
- MOBO is best in being sample efficient on low dimensional problems or with expensive simulations

OUTLINE

01

Background and Problem Formulation

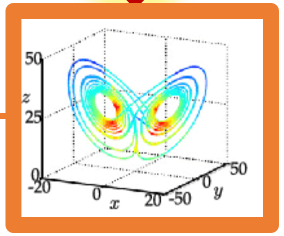


02

Methods

03

Differential RL and Results



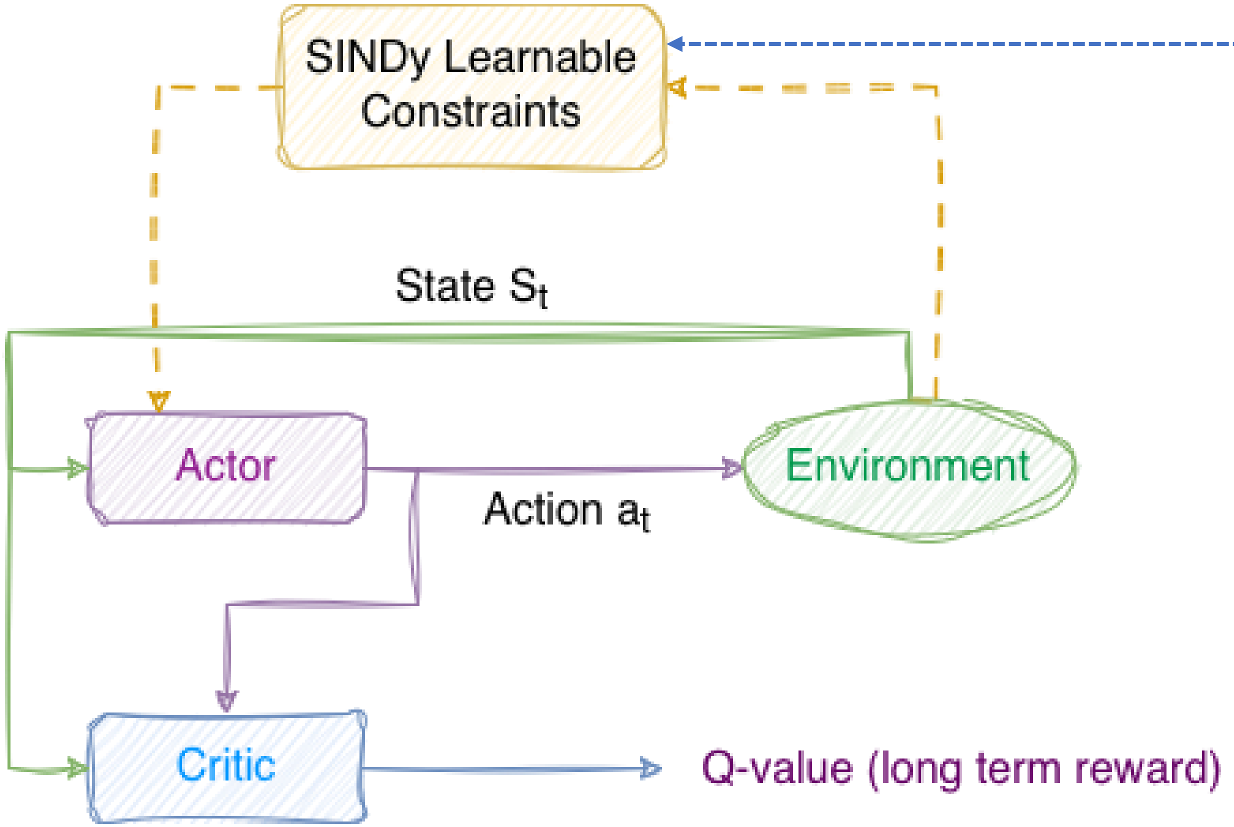
04

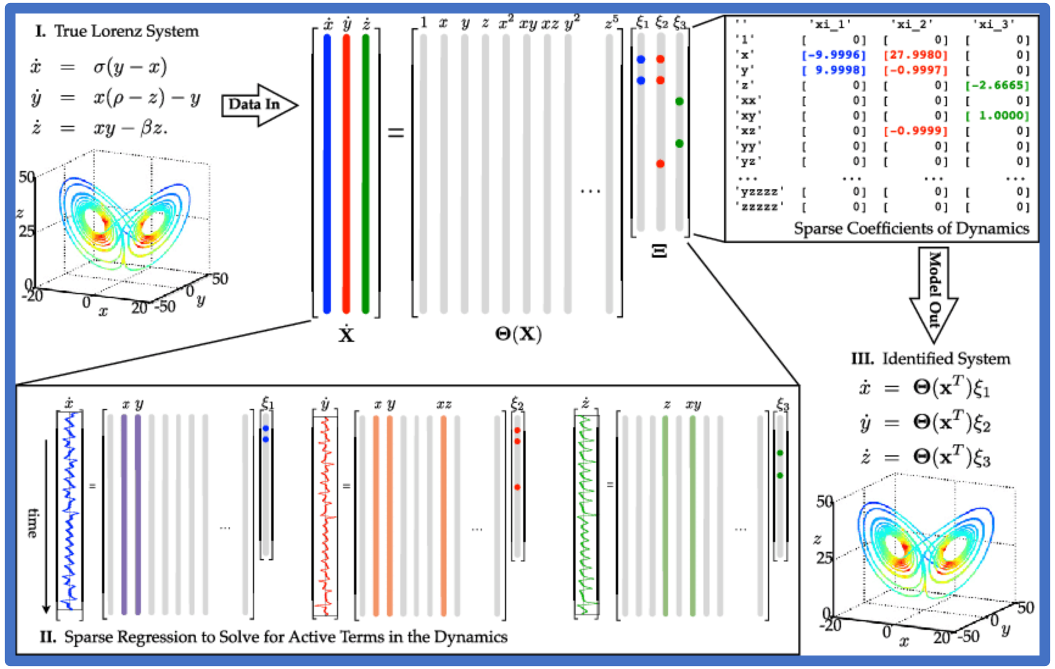
Explainable RL with SINDy and Results

05

Summary and References

EXPLAINABLE RL





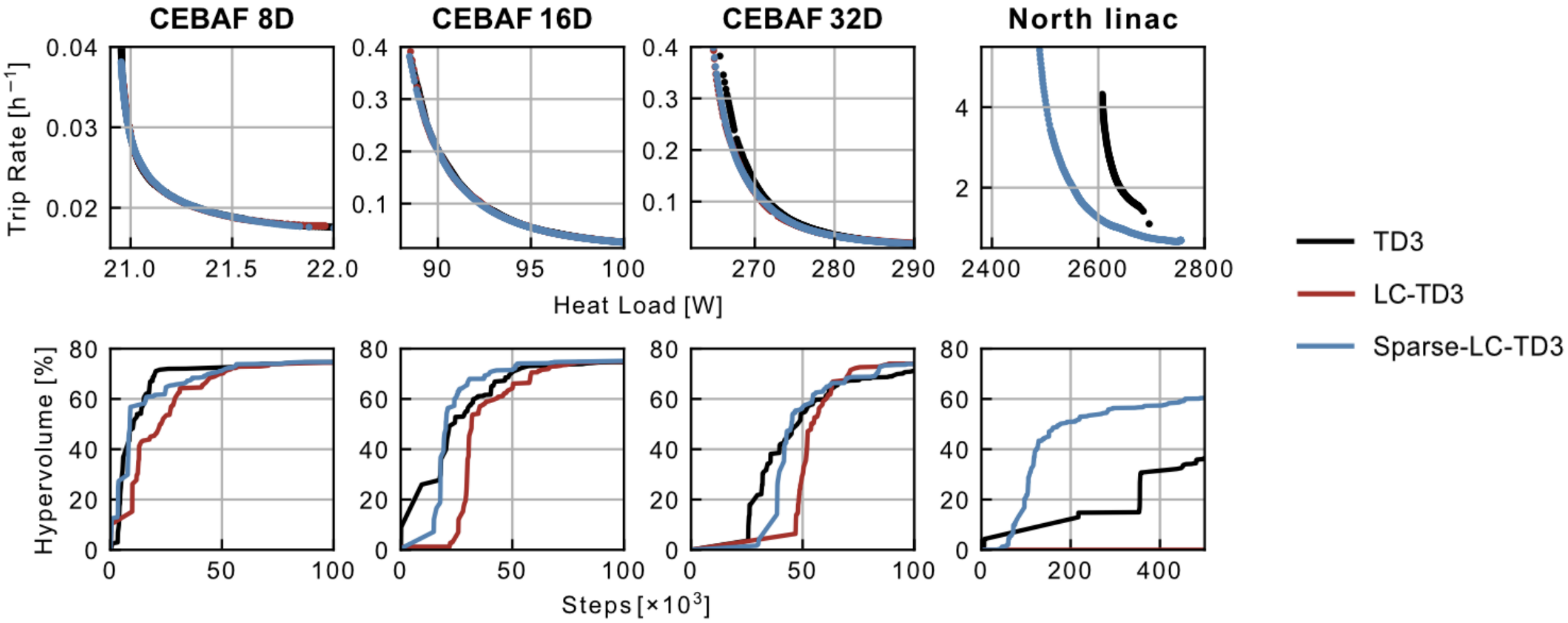
S.L. Brunton, J.L. Proctor, & J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. U.S.A. 113 (15) 3932-3937, https://doi.org/10.1073/pnas.1517384113 (2016).

Minimize $\mathbb{E}\left[\left(O_\xi(s,a) - o\right)^2\right]$ — Learn how actions affect **energy gain**

Minimize $\mathbb{E}\left[C\left(O_\xi(s,\pi_\phi(s))\right)\right]$ — Use surrogate to estimate **energy target**

NN VS SINDY LEARNABLE CONSTRAINTS (LC)

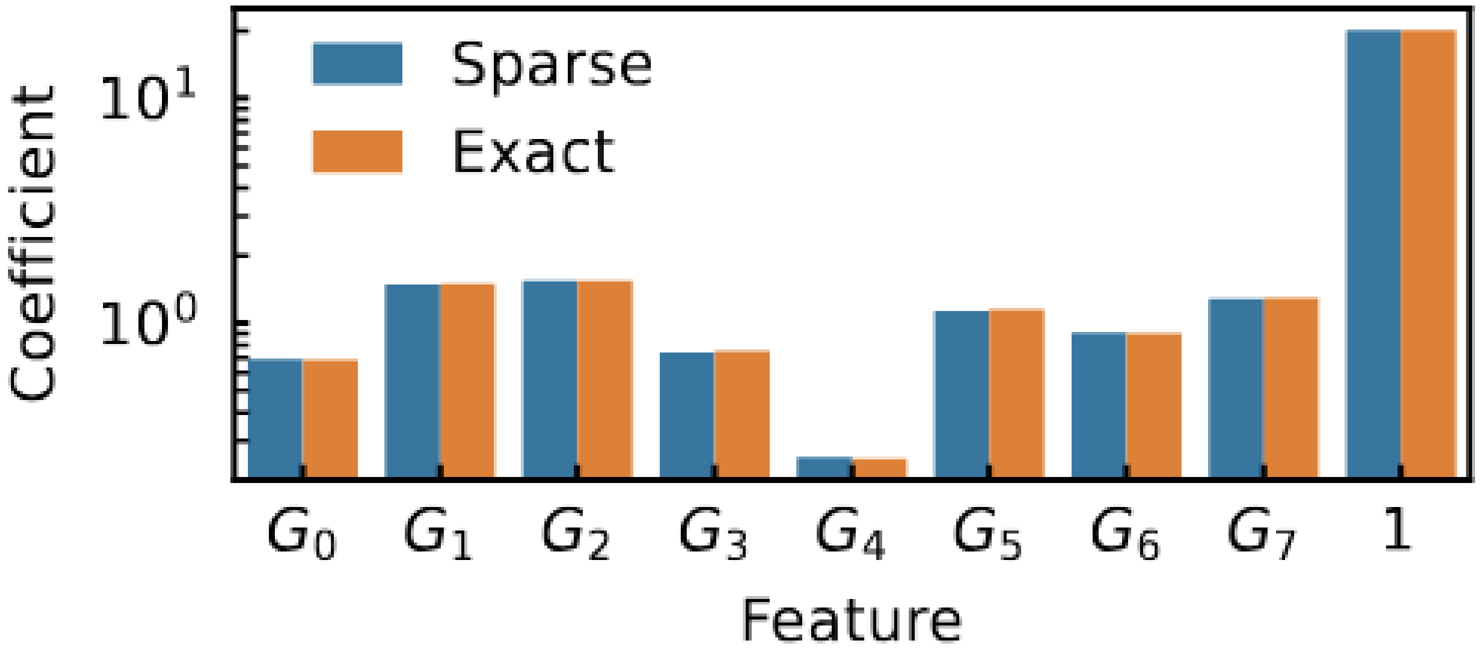
- Sparse LC-TD3 achieves superior performance on all problems



Jefferson Lab

IMPROVING EXPLAINABILITY

- By defining SINDy function libraries we can explore specific contribution
- We considered up to 3rd order polynomial
- SINDy properly identified the dominant functions and coefficients values
- Sparse dictionary model builds surrogate equation for physical observables that can be used for explainability and post verification



Jefferson Lab

SUMMARY

- DDRL outperform other algorithms – faster and better convergence
- On smaller scale problems all the reached similar solution, but MOBO is slowest even being sample efficient (attributed to fast simulation)
- Slow and Expensive Simulations → MOBO; Fast Simulations → RL
- MOBO, MOTD3 did not converge on higher dimensional problems
    - high dimensionality and combination of local and global hard-constraints
- Sparse dictionary equations accurately capture energy landscape for multi-objective problems
- Sparse LC-TD3 improves traditional TD3 performance
- Grey-box RL approach combines predictive power with operator interpretability

REFERENCES

1) *Rajput, K., Schram, M., Edelen, A., Colen, J., Kasparian, A., Roussel, R., Carpenter, A., Zhang, H., and Bennesch, J. Harnessing the power of gradient-based simulations for multi-objective optimization in particle accelerators. Machine Learning: Science and Technology (2025) doi: https://doi.org/10.1088/2632-2153/adc221*

2) *Colen, J., Schram, M., Rajput, K., and Kasparian, A. Explainable physics-based constraints on reinforcement learning for accelerator controls, 2025 - Under review at Machine Learning: Science and Technology Arxiv: https://arxiv.org/abs/2502.20247*

3) Scientific Optimization Control Toolkit (SOCT): https://github.com/JeffersonLab/SciOptControlToolkit

Thanks to all the co-authors: Jonathan Colen, Malachi Schram, Auralee Edelen, Armen Kasparian, Ryan Roussel, Adam Carpenter, He Zhang, and Jay Benesch.

Questions?