Stellenbosch University
Iyunivesithi
Universiteit

Safe, visualizable reinforcement learning for process control with a warm-started actor network based on PI-control

E.H. Bras, Prof. T.M. Louw, and Prof. S.M. Bradshaw
20068530@sun.ac.za, tmlouw@sun.ac.za, smb@sun.ac.za

Reinforcement Learning-based process control

- Data-driven, adaptive control through trial-and-error
- Roots in optimal control theory
- Assumptions regarding process dynamics removed
- Process control requires safety-aware approaches

Aim and Objectives

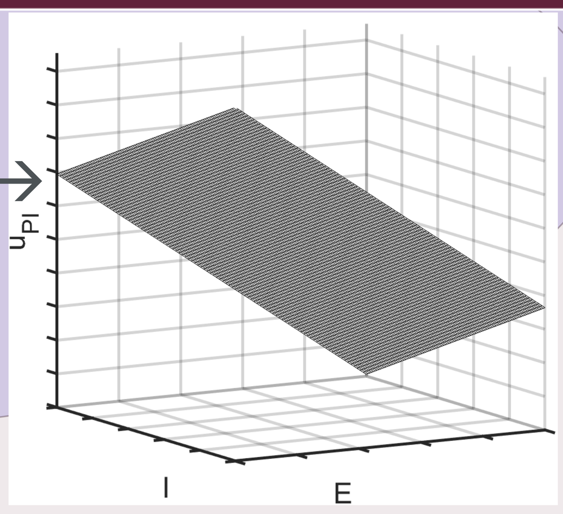
- ***Aim:***
- To support the adoption of model-free RL by practitioners
- ***Objectives:***
- Establish a synergy between RL and classical control
- Develop generally applicable tuning guidelines for typical process dynamics
- Extend application to clarify overlap between RL and model predictive control (MPC)

PI control

- Control law: $u_{PI} = K_c\left(E + \frac{1}{T_I} I\right) + I_b$

Key properties

- Well-established
- Safe
- u_{PI} is a plane acting on $E \times I$-space →

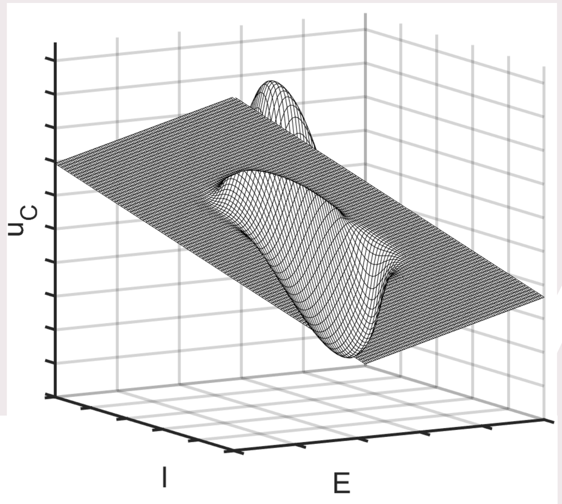


Training?

- $u_{RL}(E, I, \boldsymbol{\theta})$ and $h(E, I, \boldsymbol{\theta})$ adapted by adjusting parameters $\boldsymbol{\theta}$ incrementally

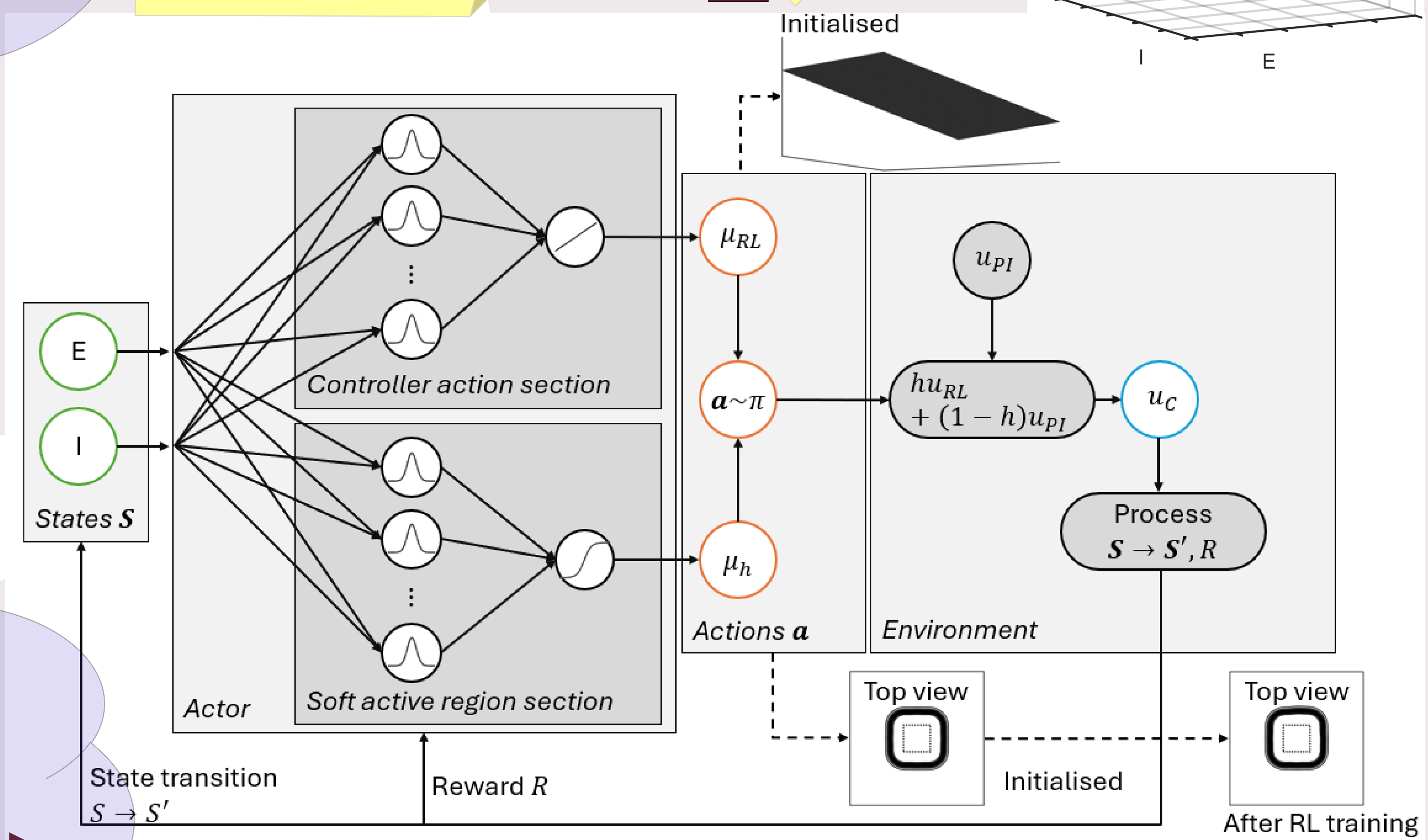
Training yields a non-linear control policy

RL agent Training



Novel, combined approach developed

Initialization of $\boldsymbol{\theta}$: Fit NN to PI plane before training RL agent



$\text{reward} = -(SP - CV)^2$

a → Process → CV

RL agent

RL

Icon made by Freepik from www.flaticon.com

$$u_C = hu_{RL} + (1-h)u_{PI}$$

Improvement + Safety

- What does the RL agent bring to the table?
- FOPTD process as example

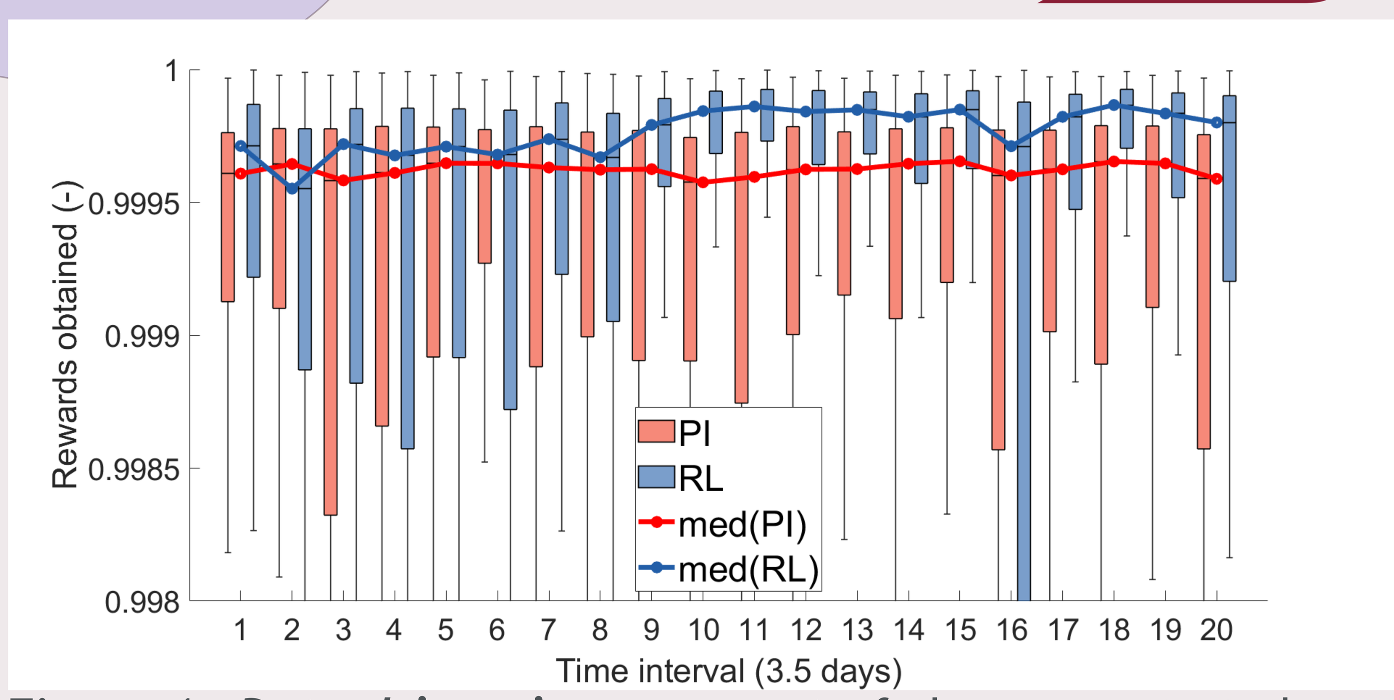


Figure 1: ***Data-driven improvement*** of the warm-started policy ***given observable states***

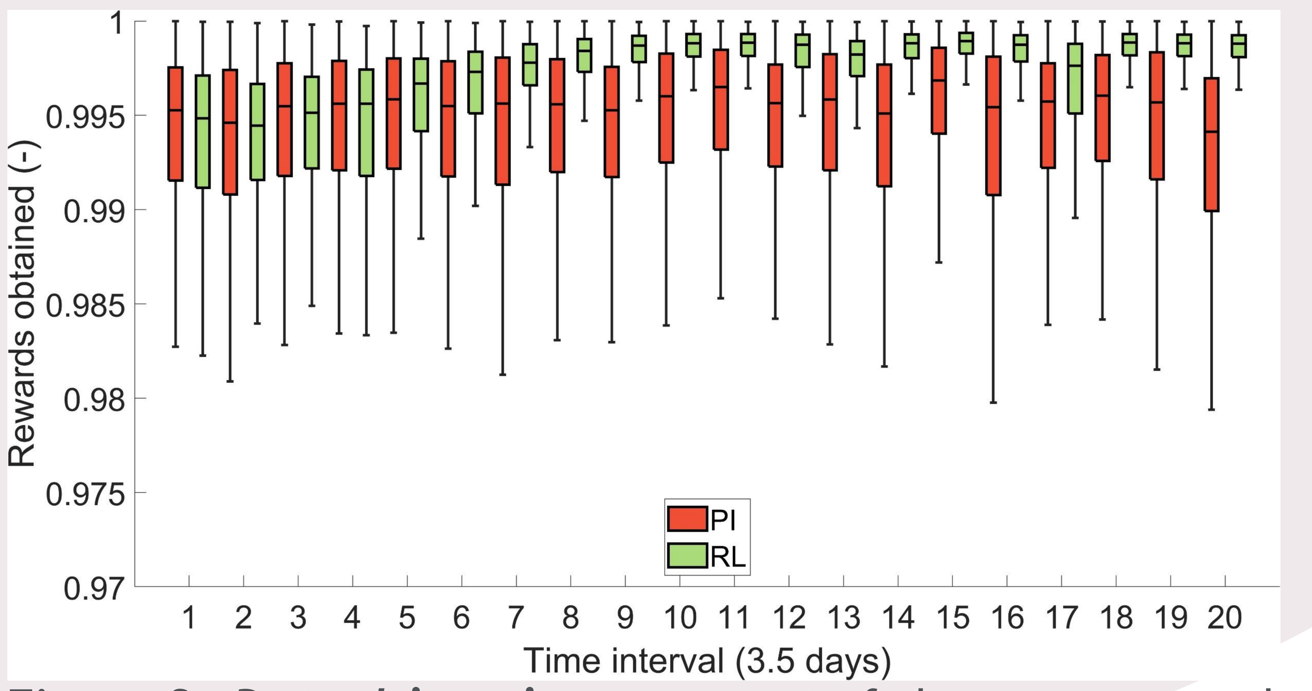


Figure 2: ***Data-driven improvement*** of the warm-started policy ***given partially observable states*** (e.g., induced by stiction dynamics)

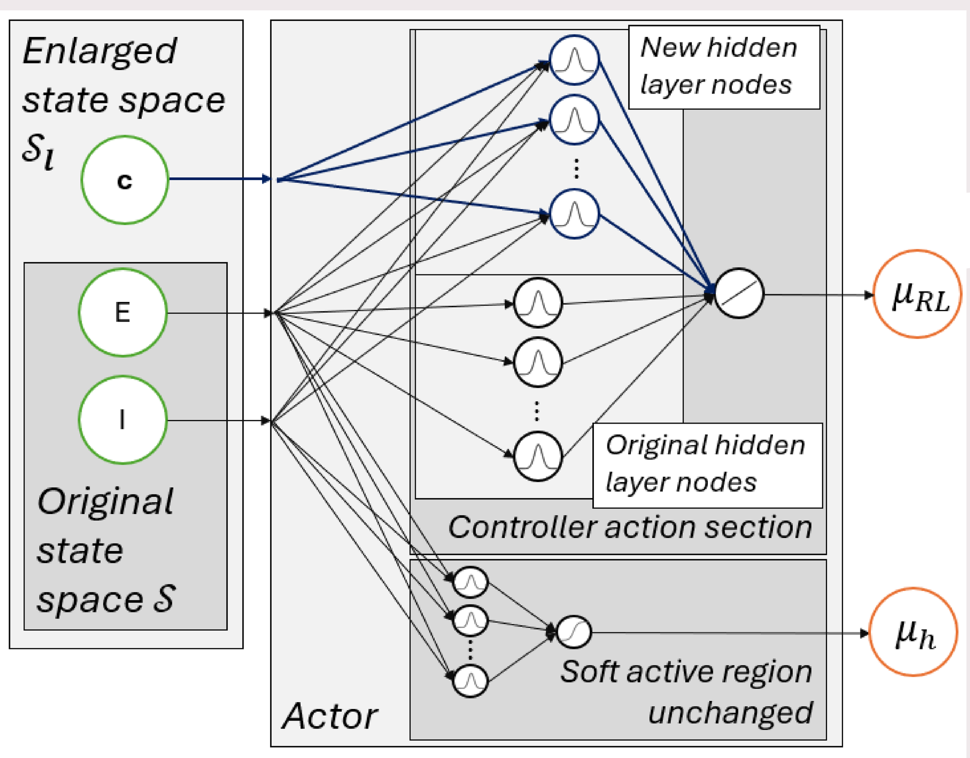


Figure 3: Opportunity to ***leverage feedforward measurements without assuming knowledge about feedforward dynamics***.

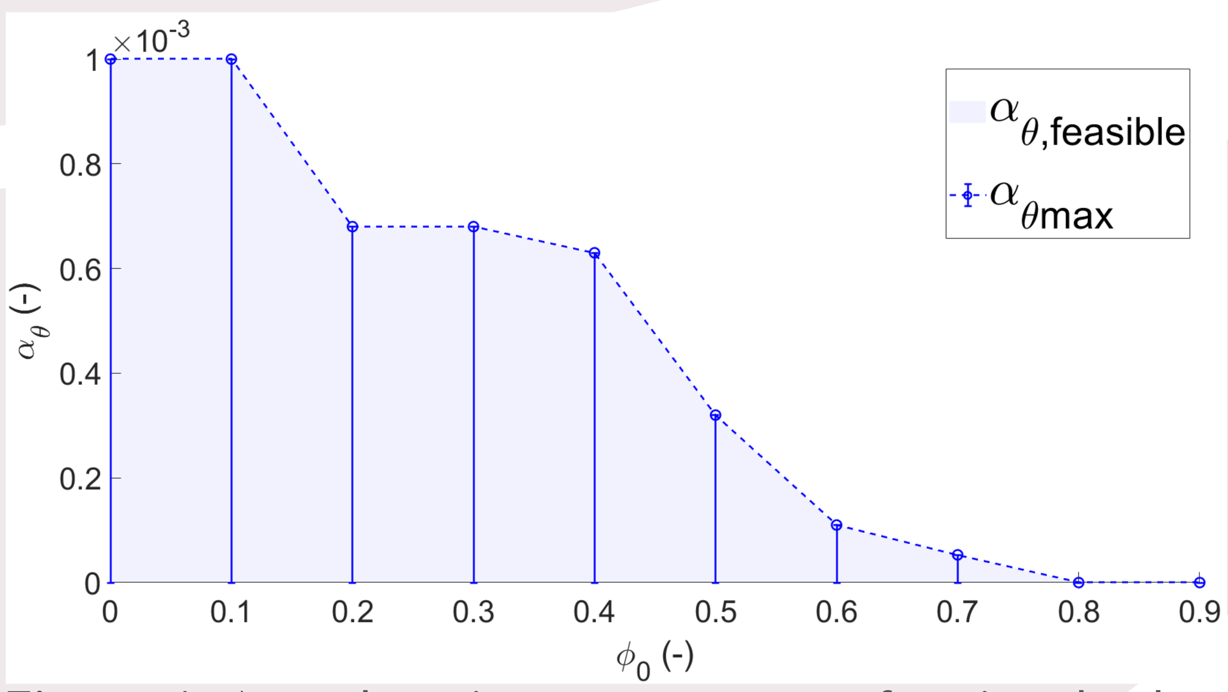


Figure 4: Actor learning rate α_θ versus fraction dead time ϕ_0. ***RL agent tuning*** defined in terms of a set of insensitive hyperparameters which ***displays robustness to a range of process dynamics***.

Future work

- α_θ guidelines for challenging process dynamics:
  - Underdamped
  - Inverse response
- Relationship of continuing online RL to MPC

Postgraduate Symposium 2024

Chemical Engineering