



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

 Tampere University



**BOSCH**  
Invented for life

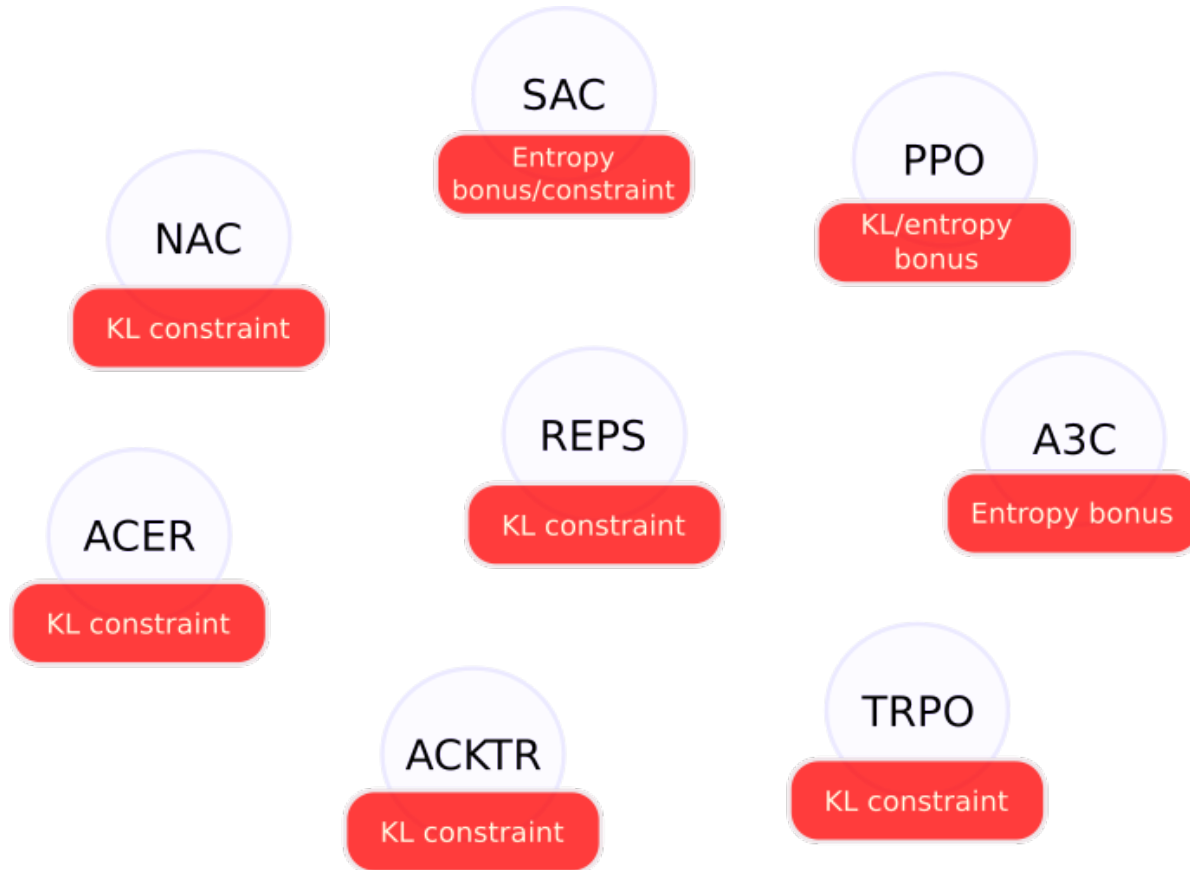


Max Planck Institute for  
**Intelligent Systems**

# Projections for Approximate Policy Iteration Algorithms

**Riad Akrou**, Joni Pajarinen, Gerhard Neumann, Jan Peters  
*IAS, TU Darmstadt, Germany*

# Entropy Regularization in RL



**Widespread** with actor-critic methods

# Hard vs Soft Constraints

- Soft constraint (bonus term)

$$\max_{\pi} \underbrace{J(\pi)}_{\text{Policy return}} + \alpha \overbrace{\mathcal{H}(\pi)}^{\text{Entropy reg.}}$$

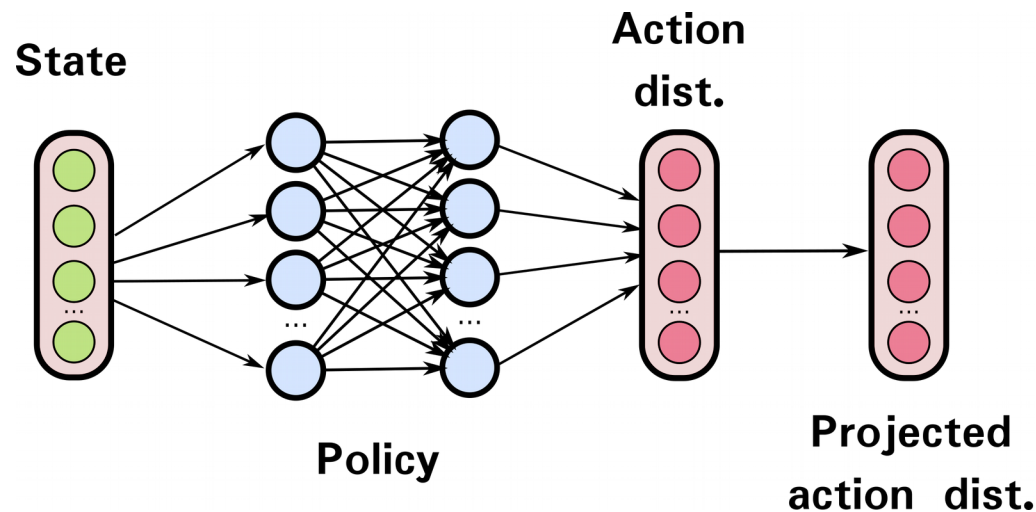
- Hard constraint

$$\max_{\pi} J(\pi) \text{ s.t. } \mathcal{H}(\pi) \geq \beta$$

- **Harder** to optimize, easier to interpret and tune

# Contributions

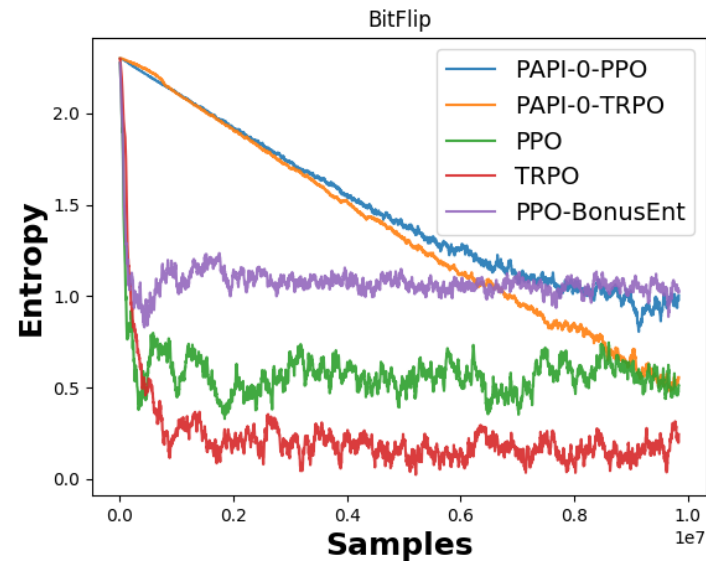
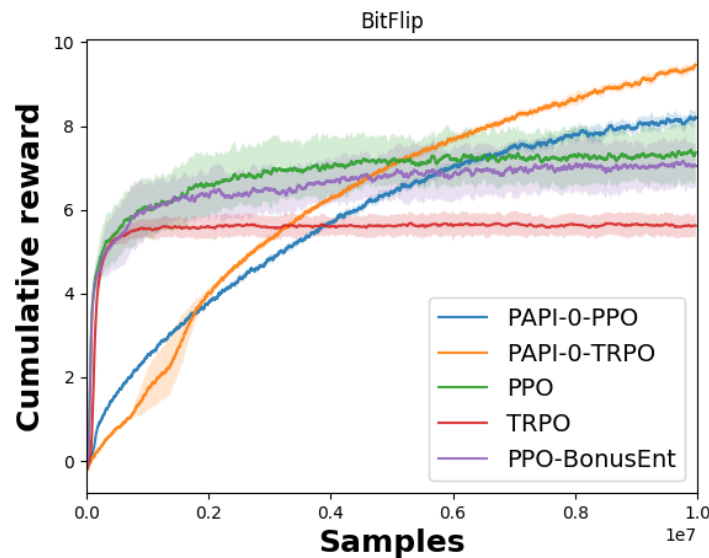
- Projections hard constraining Shannon **entropy** of **Gaussian** or **soft-max** policies



- Projections that outperform other **KL**-constrained optimizers used in deep RL

# Results

- Optimizing Objective  $\circ$  Projection( $\theta$ ) vs
  - Deep RL



- Projected gradient
- Direct policy search

# Results

- Optimizing Objective  $\circ$  Projection( $\theta$ ) vs
  - Deep RL



- Projected gradient
- Direct policy search