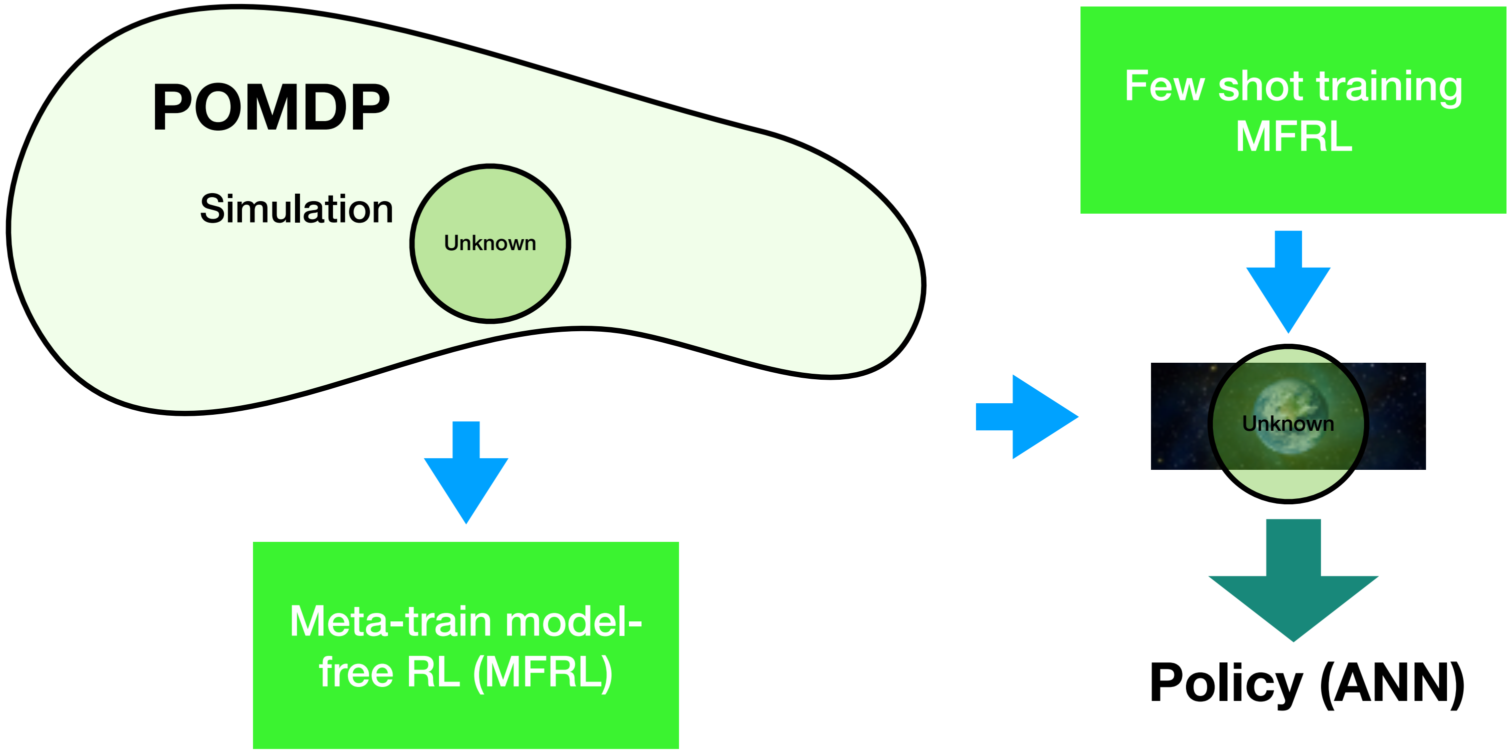
Our set-up

- We store the prior knowledge in a policy
- TRPO used for meta optimization
- Policy gradient with GAE (Schulmann 2015) as RL algorithm - fast and stable

Meta RL (in accelerator control)



- Possible scenarios:
    - Inaccurate simulation → Prepare agent for real training in reliable and fast way
    - Non-stationarity → Environment changes regularly, fast, stable retraining
    - Several similar computational demanding problems → Common pre-training
    - …