



Resource-Conditioned Reinforcement Learning for Calorimeter Design

Sara Zoccheddu | Supervisors: Shah Rukh Qasim, Nicola Serra

30/03/2026

Problem



Complex Design Space

High-dimensional space of discrete and continuous choices with strict constraints on material, cost, and geometry.



Evolving Constraints

Traditional methods optimize for fixed constraints, requiring repeated re-tuning for each new scenario.



Need for Flexibility

A flexible approach is needed – one that delivers optimal designs across a range of possibilities.

Why Reinforcement Learning?

Differentiable & Bayesian Methods

- Prone to local minima in structured parameter spaces
- Require a predefined, fixed detector model
- Cannot vary the number of components
- Struggle with discrete + combinatorial choices
- Bayesian Optimization: poor scaling in high dimensions

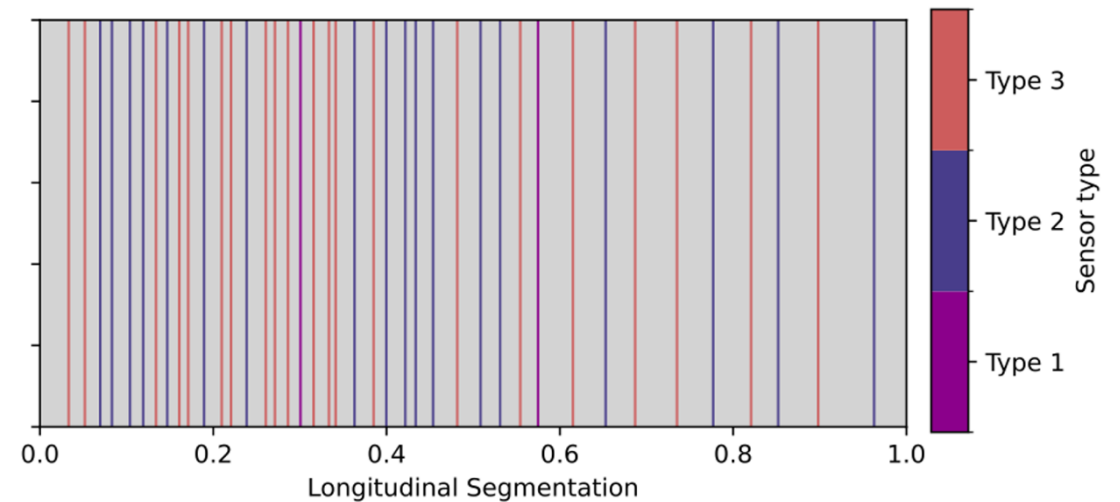
Reinforcement Learning

- Exploration helps escape local optima
- No predefined model needed — builds from scratch
- Variable number of layers / components
- Naturally handles mixed discrete + continuous actions

Calorimeter

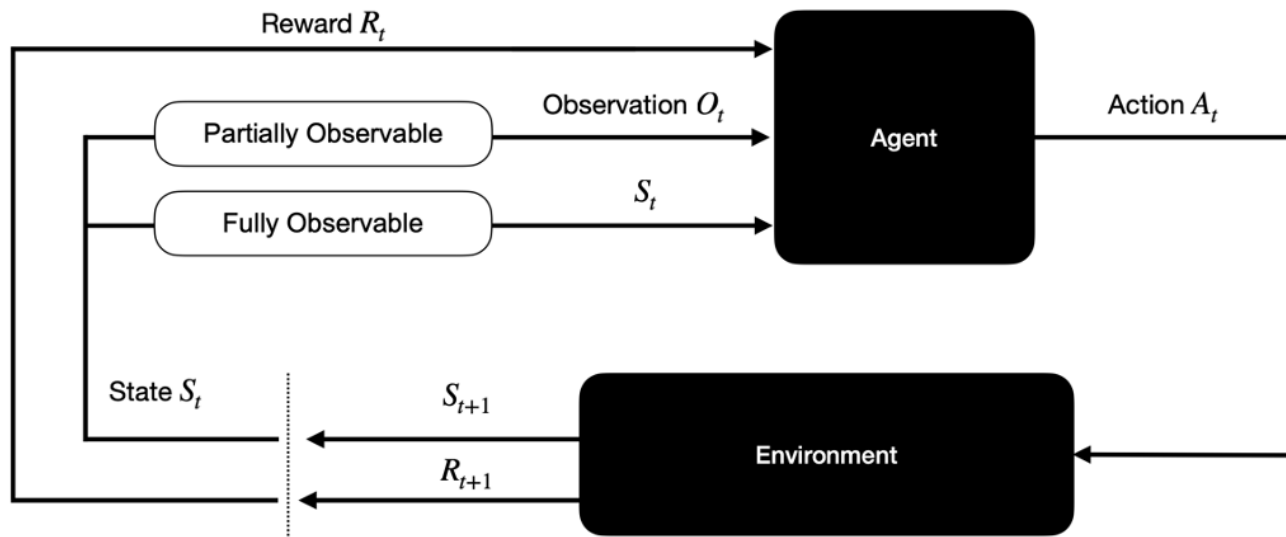
A calorimeter measures a particle's energy by recording it through sensors. Different sensors have different costs and performance characteristics.

- Each vertical line is an active sensor layer at a specific depth.
- Particle showers pass through layers, depositing energy signals.
- Tighter spacing or higher-performance layers = better resolution but higher cost
- 3 different thicknesses (types) available in this design



[2] Physics Instrument Design with Reinforcement Learning (2025), by Shah Rukh Qasim, Patrick Owen, and Nicola Serra.

Reinforcement Learning Setup



Agent-Environment Loop

Algorithm

Proximal Policy Optimization (PPO)

Observation & Action Space

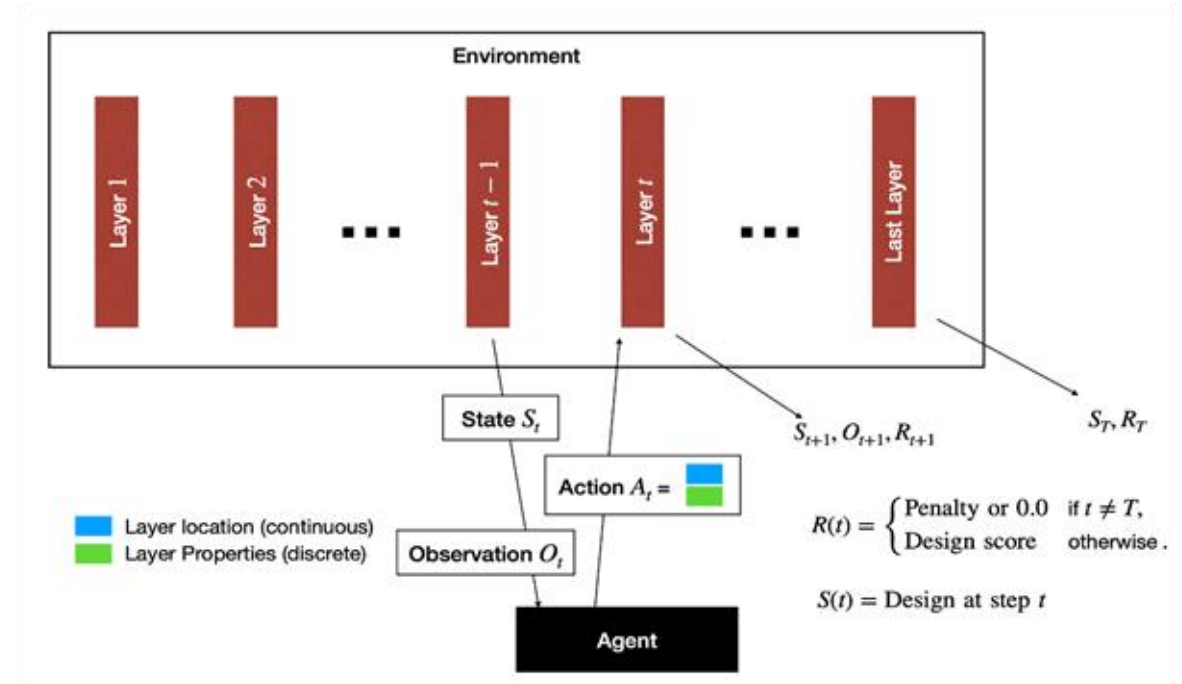
Observation Space

- Last sensor position

Action Space

Mixed continuous + discrete action:

- Continuous: gap to the next active layer
- Discrete: sensor type (out of 3 available thicknesses)



Reward

$$\text{Score} = -\max(0, \Sigma_{\text{em}50} - 8) - \max(0, \Sigma_{\text{em}100} - 5) - \max(0, \Sigma_{\text{had}50} - 25) - \max(0, \Sigma_{\text{had}100} - 18)$$

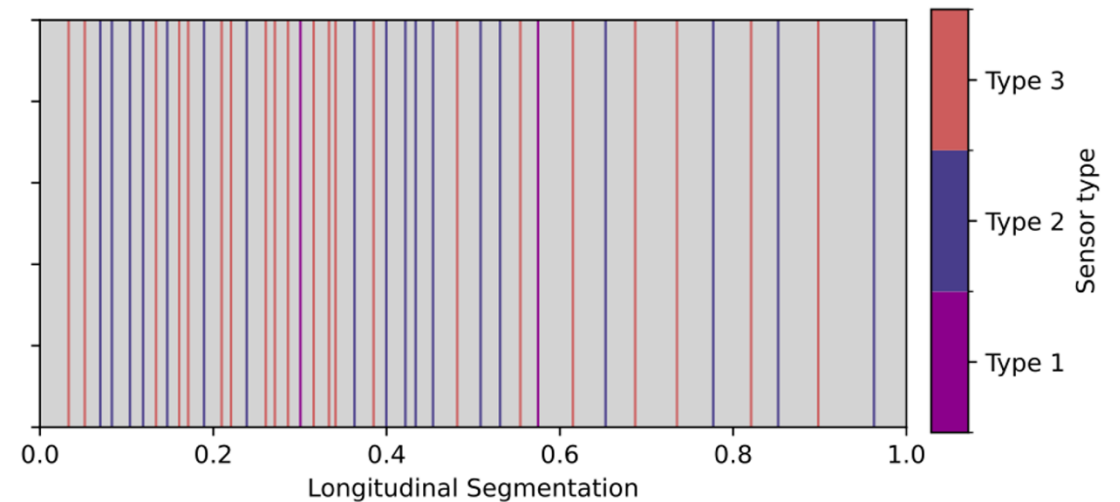
[2] Physics Instrument Design with Reinforcement Learning (2025), by Shah Rukh Qasim, Patrick Owen, and Nicola Serra.

The Challenge

What if we want to explore with different budgets?

Re-training a new agent for each budget level is **expensive** and impractical.

We need a **single agent** that can produce optimized designs for any budget.



Proposed Solution

Budget-conditioned RL: the RL agent receives the budget as part of its observation.



Single Training Run

One agent learns to generate optimized designs across the full spectrum of budgets.



Family of Designs

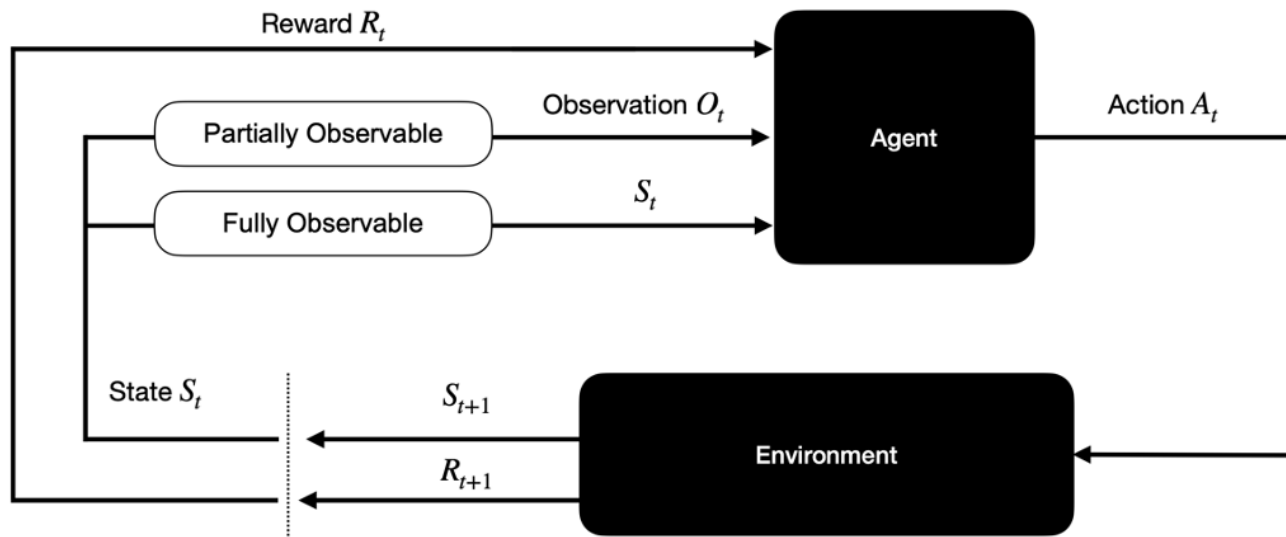
Produces a set of configurations matched to different constraint levels.



Trade-off Visibility

Directly exposes performance–cost trade-offs (e.g., energy resolution vs. material usage).

Reinforcement Learning Setup



Agent–Environment Loop

Algorithm

Proximal Policy Optimization (PPO)

Budget Conditioning

At each episode, a scalar budget is drawn uniformly from a predefined interval. The normalized budget is appended to the observation vector.

Observation & Action Space

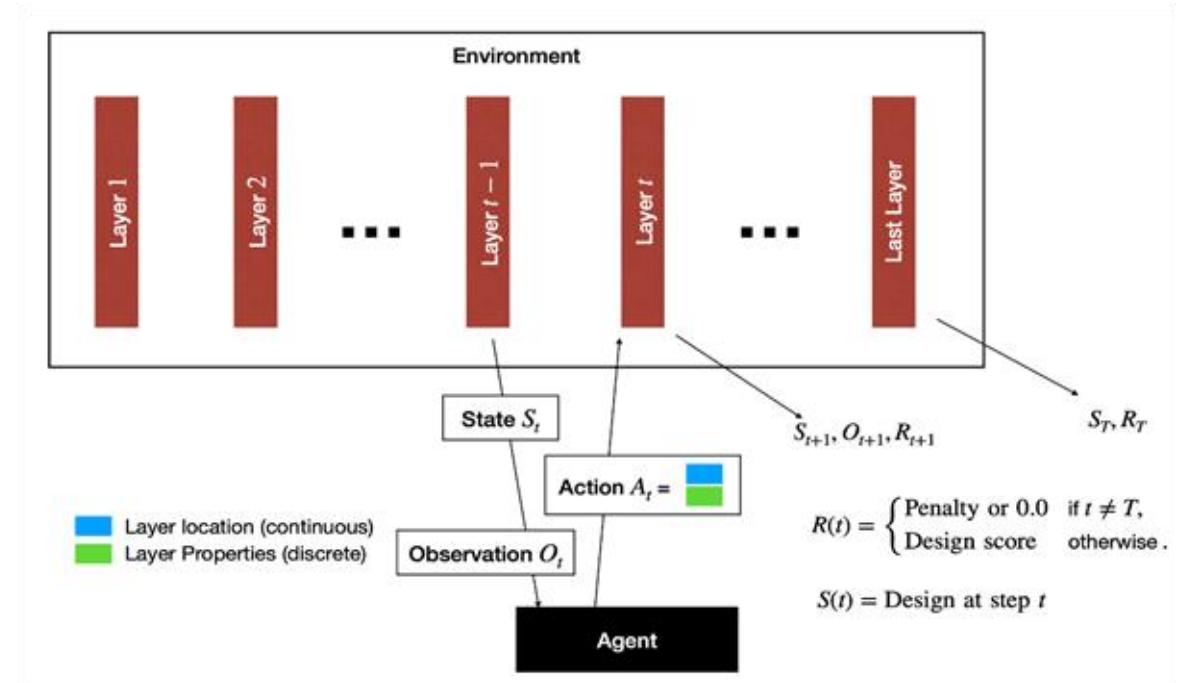
Observation Space

- Last sensor position
- **How much budget has been used**
- **Total budget in the episode (conditioning variable)**

Action Space

Mixed continuous + discrete action:

- Continuous: gap to the next active layer
- Discrete: sensor type (out of 3 available thicknesses)

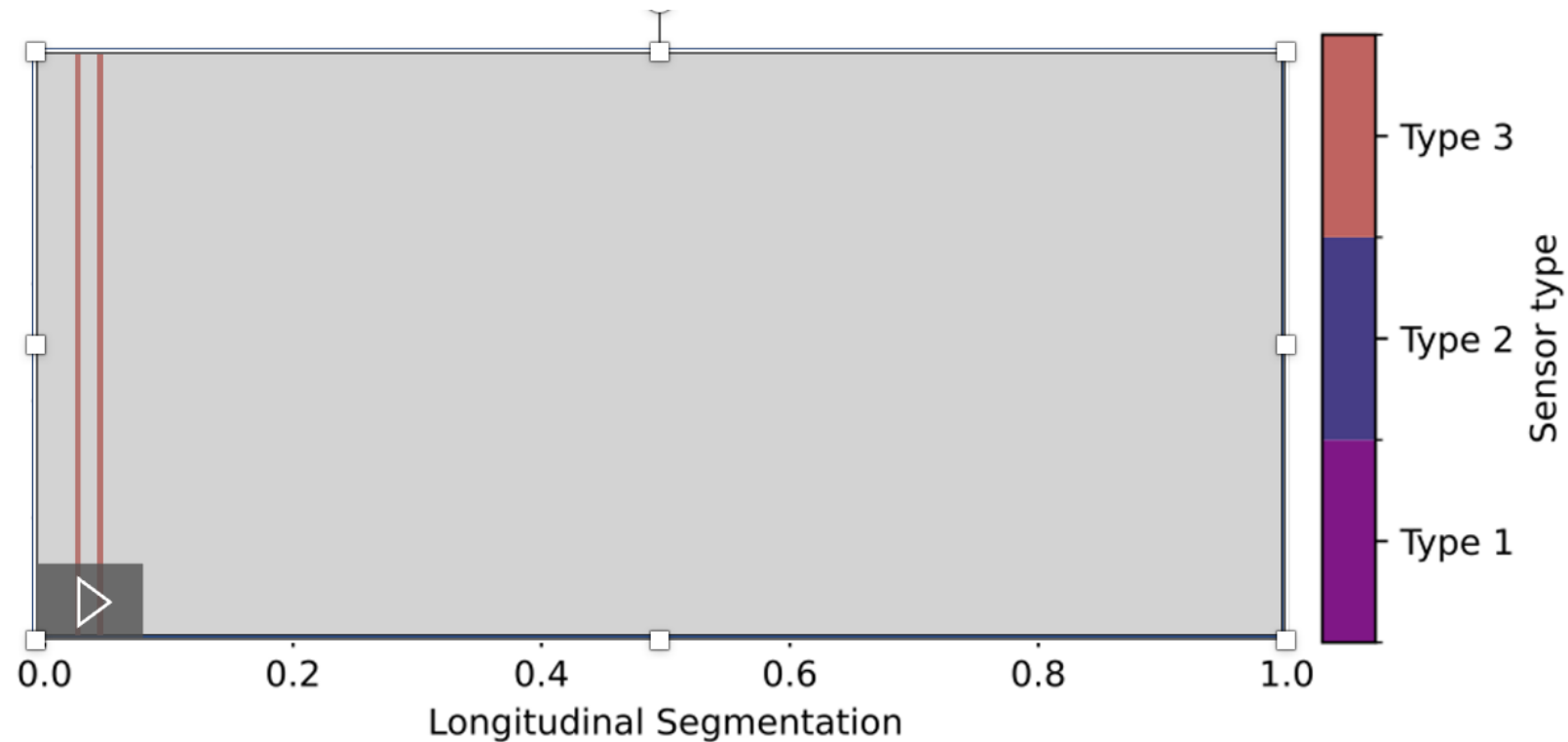


Reward

$$\text{Score} = -\max(0, \Sigma_{\text{em}50} - 8) - \max(0, \Sigma_{\text{em}100} - 5) - \max(0, \Sigma_{\text{had}50} - 25) - \max(0, \Sigma_{\text{had}100} - 18)$$

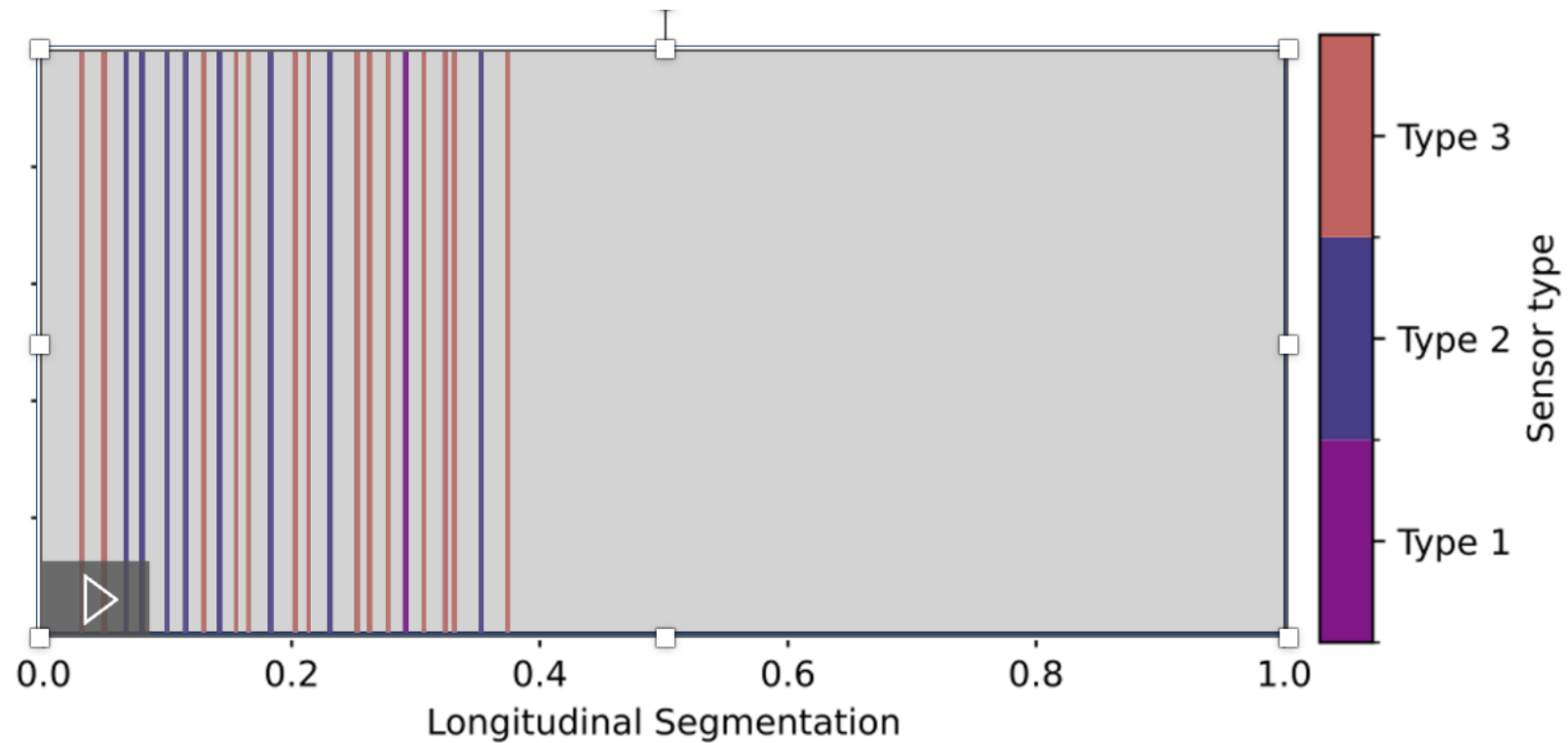
Example of Design:

Starting from an empty design, each new line represents the agent placing the next sensor layer.

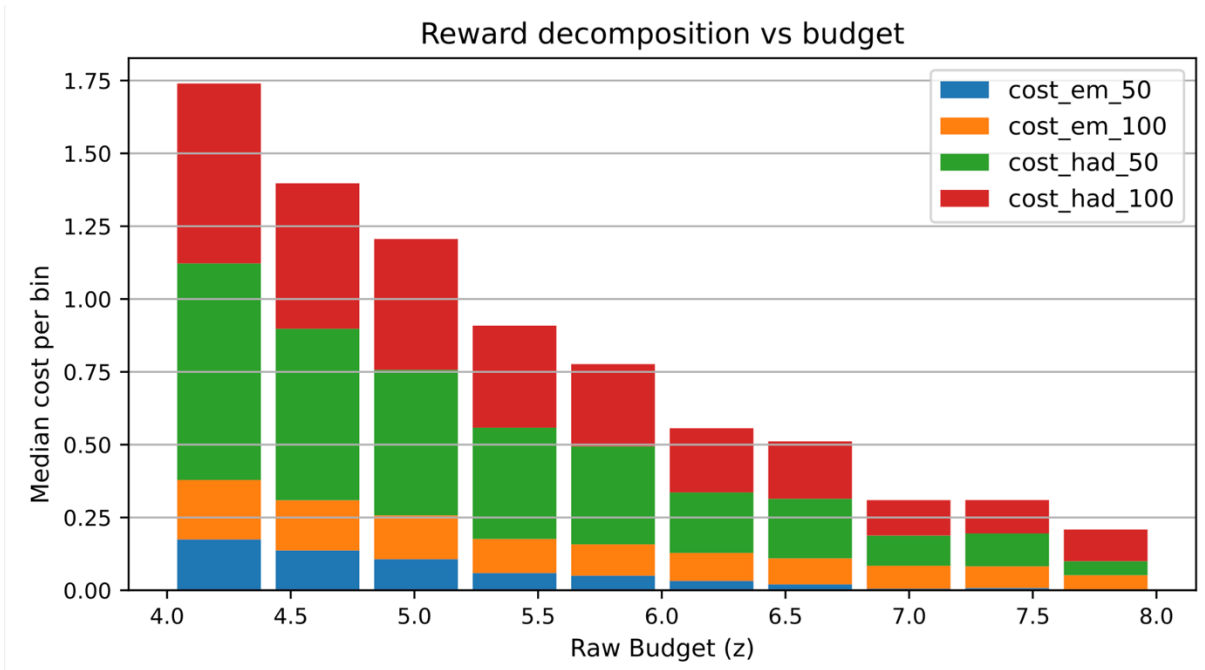
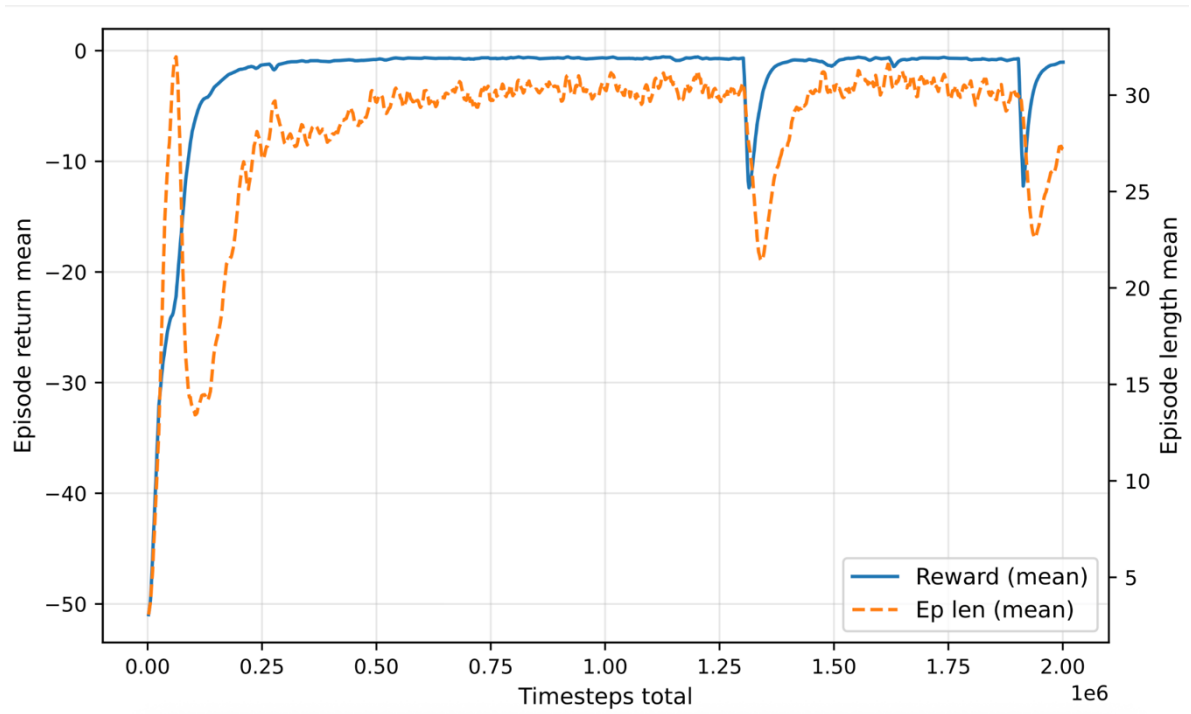


Example of Design:

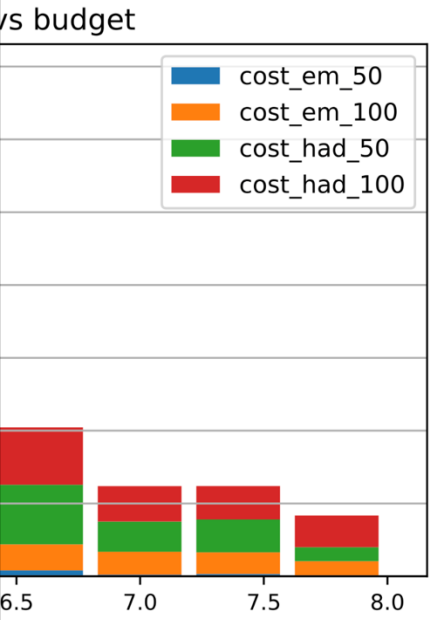
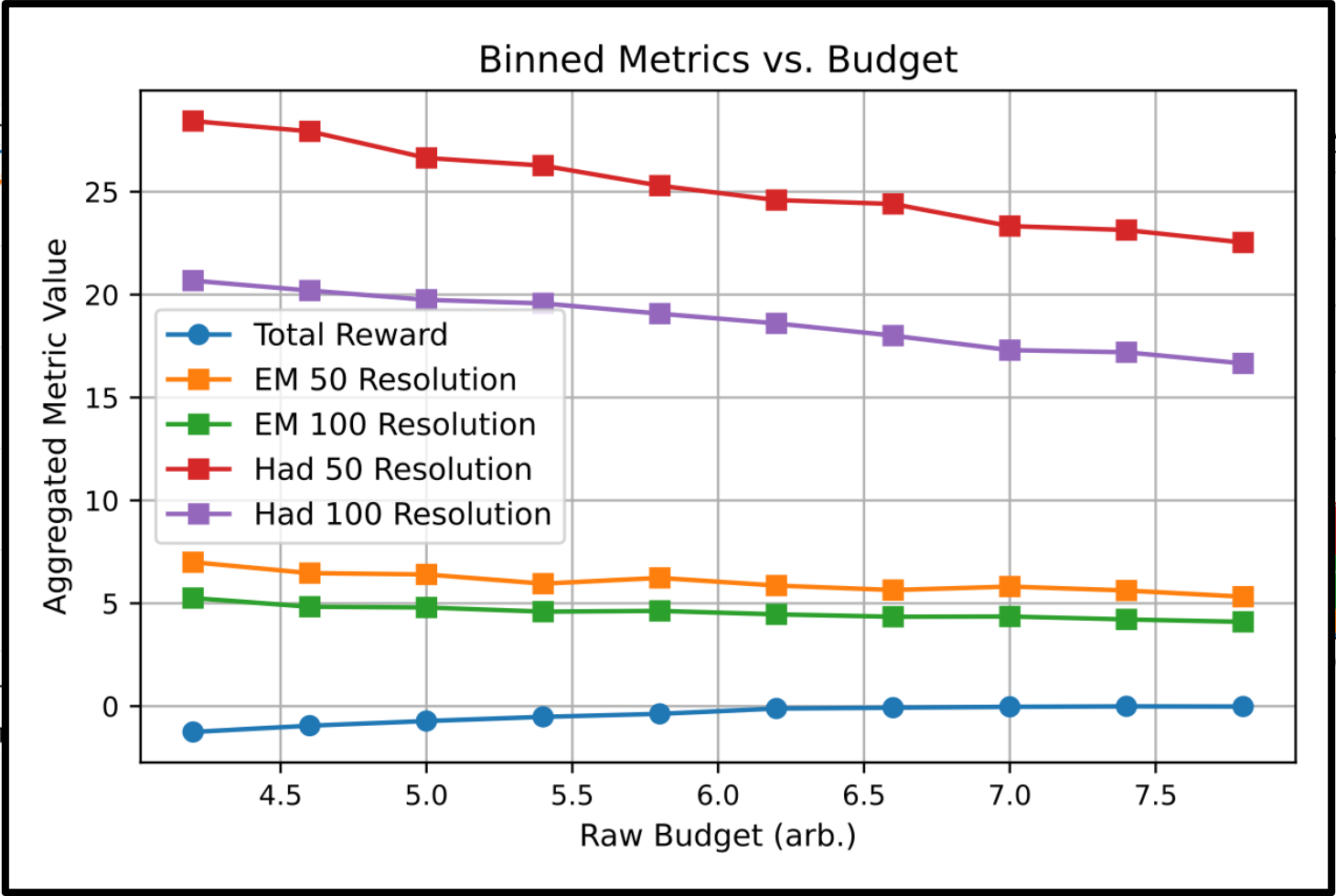
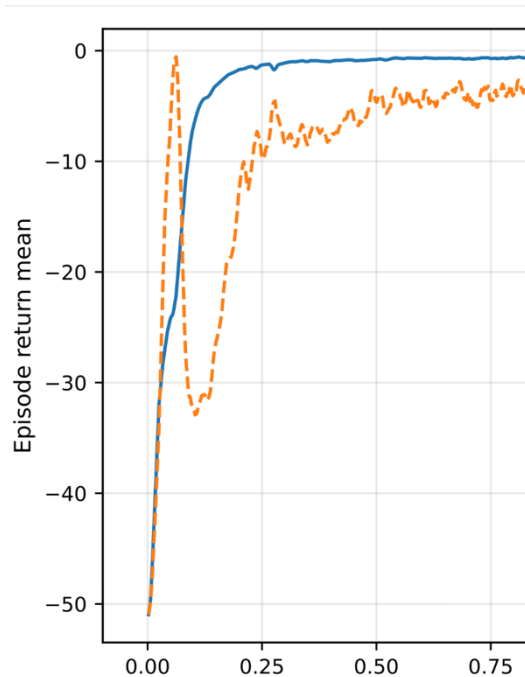
Starting from an empty design, each new line represents the agent placing the next sensor layer.



Training Results



Training Results



Conclusions



Family of Designs Policy

By conditioning on the available budget, our RL framework generates optimized calorimeter configurations across the entire range of resource constraints.



Trade-off Exploration

Allows decision-makers to consider trade-offs and make informed decisions under evolving budget scenarios.



Future Work

Extend conditional formulation to additional degrees of freedom: material, detector size, spatial envelopes, task-specific operating points.



Thank You

Questions?